Regional Integration and Trade: Controlling for varying degrees of Heterogeneity in the Gravity Model

ABSTRACT

Using a panel dataset of bilateral export flows from 12 EU countries to 20 OECD trading partners over the 1992–2003 period, this paper examines whether the effect on trade of European regional integration, denoted by an EU dummy, holds across a representative number of specifications for two gravity models, one based on the traditional trade determinants, the other based on newer trade theories (NTT). For both gravity model specifications the coefficient of the EU dummy declines in magnitude and becomes insignificant as an increasing degree of country heterogeneity is admitted into the model. This suggests the fundamental importance of the econometric specification when evaluating trade policy effects.

JEL Classification: C33, F14, F15

Keywords: Gravity models, bilateral export flows, OECD countries, Panel data, Time and space interactions

I. INTRODUCTION

The empirical success of the gravity model in explaining various types of flow variables has led to its broad application in the international trade literature. In particular, the gravity model, based on Newtonian physics, has been used to measure the effect of borders on trade (McCallum 1995), to calculate potential trade volumes – initially for the central and eastern European (CEE) countries due to their formerly inward orientation (Wang and Winters 1991; and Baldwin 1994), as well as to estimate the effect of regional integration policies ranging from free trade agreements (FTAs) to complete economic integration.[1]

Studies of the trade effects of regional trade agreements (RTAs) most commonly include regional integration in Europe, not surprising, according to Greenaway and Milner (2002), because the European Union (EU) represents the deepest and most durable RTA worldwide and its succession of enlargements provide the basis for continual study. Typically, the effects of RTAs on trade focus on the enlargement process rather than the deepening of trade integration between the EU members. Cheng and Wall (2005), for example, use a balanced panel for four years (1982, 1987, 1992, and 1997) to consider the trade effects of four FTAs in addition to the effect of an expanding Europe. In comparing the results from the pooled OLS (POLS) and the fixed effects (FE) estimator, the positive coefficient of the dummy variable for the European trading bloc is significant only for the latter.

Opposing results are obtained by Bussière *et al.* (2005). Using a sample of 61 countries based on annual data from 1980 to 2003, the EU dummy coefficient is positive and significant using POLS but is negative and insignificant for the FE estimator. For a variant of the FE approach based on fixed exporter and fixed importer effects, the negative sign is even significant. In conducting a number of robustness checks on the

---

[1] The stages of economic integration from autarky to economic union broadly accord with the classification of Balassa (1976) in the following stages. First, a free trade area (FTA), such as the 1967 Association of Southeast Asian Nations (ASEAN) or the 1994 North American Free Trade Agreement (NAFTA), eliminates import tariffs and quotas between signatory countries. Second, a customs union (CU) extends a FTA by harmonising its external trade policy. An example of a CU is the 1957 European Economic Community (EEC). Third, a common market, such as the formation of the European Community (EC) in 1967, removes all barriers to the factors of production, including labour and capital. Finally, the European Union (EU), established under the 1992 Maastricht Treaty, represents the deepest form of economic integration where policies, most notably, monetary and fiscal policies, are formally coordinated.

results, the EU dummy remains insignificant for a subsample of years starting in 1993, but becomes positive and significant only when estimated for a subsample of OECD countries.

In brief, even a limited selection from the empirical literature illustrates how the sign and significance of trade policy effects can differ. The opposing results are usually explained in terms of whether the trade agreement delivered greater liberalisation and a reduction of intra-regional tariff and non-tariff barriers. The precise econometric specification, however, may also play an important role in terms of RTA effects on trade. This is because several approaches characterise the gravity model of trade: from its inception in the 1960s as an empirical model, it was traditionally estimated as a cross-sectional or pooled regression, sometimes for a series of cross-sections or for data averaged over several years,[2] and more recently using panel estimators.

Many of the empirical findings, however, are likely to suffer from bias arising from omitted variables. In the context of cross-sectional regressions, efforts to counter this bias have typically taken the form of augmenting the standard gravity model with relevant explanatory variables in line with theoretical underpinnings. With the additional dimensions of panel datasets, solving the omitted variable bias problem has re-emerged in the form of how to correctly control for heterogeneity across countries. Recently, a number of specifications – typically parsimonious in (time-varying) economic variables and abundant in fixed effects – have been presented as the correct econometric specification of the gravity model, with each claiming that all previous specifications are restricted versions of the general model. An econometrically mis-specified model can lead to biased estimates and an incorrect inference regarding the RTA effect.

Using a panel dataset of bilateral export flows from 12 EU countries to 20 OECD trading partners for the years 1992–2003 within a gravity framework, the effect of trade policy is evaluated by comparing the coefficient estimate of the EU dummy variable from the most general model with the results from several restricted FE models. The FE variants include the triple-indexed specification which has two sets of country-specific effects, one set for the exporting country and another for the importing country; the

[2] OLS year-by-year estimations are performed when the evolution of trade over a specific time period is of interest, data averages are used to reduce the effects of business cycles or shocks of short-term duration.

standard FE model in which bilateral trade is based on factors that are unique to each country-pair; a model which combines the elements from both the triple-indexed version and the standard FE specification; and a generalised gravity model wherein country-time interactions are also used to explain bilateral trade flows. The pooled OLS estimates, which ignore the dimensions of panel datasets, are used as a comparison with earlier studies. The various gravity specifications are estimated for a model based on the traditional determinants of bilateral trade in addition to one which follows the NTT determinants.

In comparing the results for the two models, the coefficient estimates vary quite considerably across the specifications for the traditional model in contrast to the more stable estimates for the NTT model, indicating that the consequences of heterogeneity are less benign for the traditional model. For both models, however, the EU dummy coefficient declines in magnitude and becomes insignificant as an increasing degree of country heterogeneity is allowed. The findings point to the importance of the econometric specification and estimation method when evaluating trade policy effects.

The layout of this paper is as follows. Section II sets out the relevant literature concerning the nature of omitted variable bias depending on whether the gravity model is estimated as a cross-sectional OLS regression or using a panel estimator. Section III presents the econometric specifications and data used to evaluate trade policy and outlines the restrictions placed on the gravity model parameters when nested versions of the general model are estimated. Section IV discusses the empirical results and Section V concludes.

II. THE RELEVANT LITERATURE

Developed independently by Tinbergen (1962) and Pöyhönen (1963), the gravity model of international trade has since enjoyed popularity mainly due to its empirical success in explaining bilateral trade patterns and its versatility in application. In its basic form, the standard gravity model explains bilateral trade flows by the economic size of two countries and the distance between them. As pointed out by Harrigan (2001), the standard gravity equation allowed no role for comparative advantage: neither relative technology levels nor relative endowments enter the equation.

In the augmented version of the gravity model, per capita income levels for both the exporting and the importing countries are included as additional regressors. The dependence of trade on per capita GDP stems from the Linder hypothesis. Linder (1961) proposed a demand-based theory which explains trade in terms of the similarity of demand characteristics between trading partners. Building on Linder's hypothesis, Gruber and Vernon (1970) append the absolute difference between the two countries' per capita incomes to the standard gravity equation as a way of capturing differences in consumption patterns. By identifying separate roles for GDP and per capita GDP, the augmented model is used to capture demand generated by non-homothetic preferences in the importing country and factor endowments in the exporting country (Bergstrand 1989).[3]

In short, the geographical distribution of a given country's trade flows with partner countries is traditionally determined by three sets of factors: export supply, as captured by GDP and per capita GDP of the exporting country; import demand, as given by GDP and GDP per head of the importing country; and other factors that influence trade either negatively, such as transport costs, typically proxied by distance, or positively, for example, bilateral policy agreements that reduce cross-border barriers to trade.

Anderson (1979) uses the Armington assumption that goods are differentiated by country of origin so that consumer preferences are based on the assumption that all goods are traded, implying that national income is the sum of traded goods output. In equilibrium, this amounts to the sum of home and foreign demand for the unique good that the country produces (Harrigan 2001).

For the NTT model, Helpman and Krugman (1985) use a monopolistic competition model to derive a (frictionless) model for bilateral trade in which consumers' tastes for variety are proportional to the number of varieties produced in equilibrium, which, in turn, depends on country size. The role of country size similarities is further

[3] Following Linnemann (1966), the augmented model could equivalently be specified in terms of GDP and the population of both countries.

emphasised by Helpman (1987). In essence, the volume of trade between two countries is proportional to the relative size of the two countries if both countries are specialised in their outputs, tastes are identical and homothetic, and trade is free, meaning that the prices of goods worldwide are identical. The size dispersion index shows how countries with differing relative size tend to engage in a lesser degree of bilateral trade. Conversely, bilateral trade tends to be higher between countries with similar relative size. Following the NTT literature, the gravity model can be expressed in terms of the overall size of the bilateral country-pairs, an index of similarity in size, in addition to the relative income per capita variable as a way to capture similarity of factor endowments.

Despite the attributes of the gravity model – its relative simplicity, parsimony and high explanatory power feature among its advantages – criticism of the model has been twin-pronged: initially, for its lack of theoretical foundations and more recently, because of a lack of attention to its econometric properties, without which the accuracy of the estimates may be questionable. More generally, the possible mis-specification of the gravity equation due to omitted variables relates to the estimation method. Indeed, Greenaway and Milner (2002) acknowledge that the estimation method is likely to be an important issue in terms of interpreting the gravity model coefficients.

In a cross-sectional context, potential bias has been alleviated by the inclusion in the model of relevant explanatory variables aligned with theoretical underpinnings. Drawing on the Armington assumption, Anderson (1979) was the first to derive a gravity model which accounted for transport costs as well as national tariffs in each country, both of which are expected to increase with distance. That trade costs differ depending on location was further elaborated by Bergstrand (1985). The derived gravity model, in which products are differentiated nationally by monopolistic competition, explicitly includes price indexes. Hence, prices should feature among the explanatory variables in the gravity model of trade.

Not satisfied with the ability of simple price indexes to account for all those factors which impede bilateral trade,[4] Anderson and van Wincoop (2003) seek to identify trade costs – often not directly observable – that give rise to international differences in

[4] For example, the chosen price index entails a degree of arbitrary selection while its inclusion in a model does not guarantee that the omitted variable bias problem is eliminated.

prices. They develop a gravity model which includes a multilateral resistance term, that is, a proxy for the bilateral trade barriers of two countries relative to their average trade barriers with all other trading partners. In effect, they theoretically justify the inclusion of a remoteness (or relative distance) variable in the gravity model, where remoteness is defined in terms of trade costs rather than by geographical location.

Yet, cross-sectional and POLS regressions do not cater for the dimensions of a panel dataset. More particular, an estimator that does not allow for (unobserved) heterogeneous trading relations may still incur biased estimates. In a panel context, the constrained minimisation approach used by Anderson and van Wincoop (2003) is abandoned in favour of the fixed effects (FE) estimator. This is equivalent to the least squares dummy variable (LSDV) estimator which introduces a dummy variable for each cross-sectional observation to allow for the effects of unobserved price indexes. Indeed, Feenstra (2003) advocates a preference for a gravity specification with fixed effects to capture the mis-specified factors because its benefits in terms of consistency and computational simplicity outweigh the relatively small loss in efficiency incurred.

Nevertheless, a problem arises in terms of the exact way the unobserved fixed effects (FE) should be specified. A number of variants of the FE approach – typically parsimonious in (time-varying) economic variables and abundant in fixed effects – have been claimed as the correct econometric specification of the gravity model. Among the first to incorporate country-pair effects into a trade model, Hummels and Levinsohn (1995) argue that even if the underlying theoretical model is correct, the model might not fit the data in every year for every country-pair. Given that border trade, seasonal trade, cultural ties, and trade restrictions vary across country-pairs, they advance an explanation for bilateral trade based on factors that are unique to each country-pair. Specifically, they assert that these factors can be accurately modelled as a country-pair fixed effect, $\gamma\omega_{ij}$. Essentially, the standard FE specification avoids biased parameter estimates arising from the omission of time-invariant bilateral variables:

$$Exp_{ij}^{t} = \alpha_0 + \gamma\omega_{ij} + \beta' x_{ij}^{t} + \varepsilon_{ij}^{t} \qquad (1)$$

where Exp_{ij}^t is exports from country i to country j at time t; x_{ij}^t denotes a $k \times 1$ vector of explanatory variables that vary over time; and ε_{ij}^t is the random error. The intercept is comprised of α_0 and the country-pair fixed effects, $\gamma\omega_{ij}$. In a two-way FE model, the intercept additionally includes time-specific effects, θ^t, in equation (1) to control for common shocks affecting all countries in the sample. Mátyás (1997) instead proposes a triple-indexed specification of the gravity model with two sets of country-specific effects based on fixed effects for both the exporting country, γ_i, and the importing country, ω_j:

$$Exp_{ij}^t = \alpha_0 + \theta^t + \gamma_i + \omega_j + \beta' x_{ij}^t + \varepsilon_{ij}^t \tag{2}$$

Egger and Pfaffermayr (2003) amalgamate the specific effects from both models, referring to the country-specific effects of the triple-indexed specification as the main effects and the country-pair effects from the standard FE model as the time-invariant exporter-importer bilateral interaction effects:[5]

$$Exp_{ij}^t = \alpha_0 + \theta^t + \gamma_i + \omega_j + \gamma\omega_{ij} + \beta' x_{ij}^t + \varepsilon_{ij}^t \tag{3}$$

Baltagi, *et al* (2003) argue that as much heterogeneity as possible must be controlled for in order to obtain reliable parameter estimates. This is because a model that does not span the whole vector space of possible treatments in explaining variations in bilateral trade is potentially mis-specified. They utilise a further dimension of the panel dataset by interacting the country-specific effects for both the exporter and the importer countries with time-specific effects. These interaction effects are included to capture country-specific, time-varying effects such as a country's business cycle; its cultural, political, or institutional characteristics; as well as unobserved factor endowment variables. Baier and Bergstrand (2007) also incorporate country-time effects in a gravity model of trade to account for the variation of multilateral price terms in a panel context. In short, the recommended specification consists of a generalised gravity model wherein a full interaction effects design is used to explain bilateral trade flows, namely three sets of main effects (time dummies, exporter-specific effects, and importer-specific effects) and

[5] The interaction of two variables implies the product of their effects. Apart from the intercept, a model that includes country-pair dummies yields identical coefficient estimates to one with exporter-importer interactions.

three sets of interactions (exporter-importer interactions, exporter-time interactions, and importer-time interactions):

$$Exp_{ij}^{t} = \alpha_0 + \theta^t + \gamma_i + \omega_j + \gamma\omega_{ij} + \gamma\theta_i^t + \omega\theta_j^t + \beta' x_{ij}^t + \varepsilon_{ij}^t \tag{4}$$

While efforts to control for unobserved heterogeneity entail an emphasis on the econometric properties of the gravity model, the various FE specifications have generally not been used to examine trade policy effects.[6] Although focusing mainly on the POLS and FE estimates, an important exception is Cheng and Wall (2005), who evaluate RTA effects on trade by comparing the coefficient estimates of five RTAs using an (asymmetric)[7] FE model against POLS as well as a number of nested versions of their benchmark model – symmetric FE, differenced FE, and the triple-indexed model. They illustrate that ignoring unobserved heterogeneity translates into biased estimates of bilateral trade relations, but while they control for country-pair heterogeneity, they do not cater for the additional dimensions of the panel given in the most general model.

Given that an econometrically mis-specified model may lead to incorrect inferences regarding the gravity model coefficient estimates, this paper examines whether the various approaches used in the literature have implications in terms of the trade effect of European regional integration, as denoted by an EU dummy. Previous studies tend to estimate several RTA effects on trade using a large sample of countries, but this may understate the effect of EU expansion simply because in a large sample that includes emerging economies, the rise in trade levels for these countries have tended to outpace any increases in trade by the more mature EU countries. Such trade patterns are not captured by time dummies, since their inclusion in a gravity model will control only for common shocks which affect all countries in the sample. Therefore, an OECD-based sample is used to evaluate trade policy. In short, the sign and significance of the EU dummy coefficient is compared across a representative number of specifications which allow for varying degrees of heterogeneity for two OECD-based gravity models, one

[6] Mátyás (1997) does not explicitly include regional dummies in the model but he suggests the presence of a significant trade bloc effect if both the exporter-specific and importer-specific effects are large for most of the countries within a trading bloc relative to the countries outside the bloc.

[7] The asymmetric FE specification, $\gamma\omega_{ij} \neq \omega\gamma_{ji}$, yields almost identical coefficient estimates of the RTA effects on trade as the symmetric FE specification, $\gamma\omega_{ij} = \omega\gamma_{ji}$.

based on the traditional trade determinants, the other based on newer trade theories (NTT).

III. MODEL SPECIFICATION AND DATA

Following the traditional trade literature, the econometric specification of the gravity model of bilateral exports in its most general form is:

$$Exp_{ij}^{t} = \alpha_0 + \theta^t + \gamma_i + \omega_j + \gamma\omega_{ij} + \gamma\theta_i^t + \omega\theta_j^t + \beta_1 GDP_i^t + \beta_2 GDP_j^t + \beta_3 GDPpc_i^t + \beta_4 GDPpc_j^t + \beta_5 EU_{ij}^t + \varepsilon_{ij}^t \qquad (5)$$

where Exp_{ij}^{t} are the bilateral export flows from 12 EU countries to 20 OECD partner countries, expressed in US dollars at constant 2000 prices; GDP_i^t and GDP_j^t are expressed in constant 2000 US dollars and denote the economic size of the exporting and the importing countries respectively; $GDPpc_i^t$ and $GDPpc_j^t$, expressed in constant 2000 US dollars, are the respective countries' per capita income levels. The equation includes the full set of main effects and interactions. All non-dummy variables are estimated in logarithms.

Following the NTT literature, equation (5) can be amended as follows:

$$Exp_{ij}^{t} = \alpha_0 + \theta^t + \gamma_i + \omega_j + \gamma\omega_{ij} + \gamma\theta_i^t + \omega\theta_j^t + \lambda_1 TGDP_{ij}^t + \lambda_2 sGDP_{ij}^t + \lambda_3 DGDPpc_{ij}^t + \lambda_4 EU_{ij}^t + \mu_{ij}^t \qquad (6)$$

where total GDP is the sum of GDP for both countries, $TGDP_{ij}^t = \ln(GDP_i^t + GDP_j^t)$, as a measure of the overall country incomes; the similarity index for each country-pair is derived from the two countries' shares of GDP, given by $sGDP_{ij}^t = \ln\{1 - [GDP_i^t / (GDP_i^t + GDP_j^t)]^2 - [GDP_j^t / (GDP_i^t + GDP_j^t)]^2\}$; and the absolute difference in GDP per capita as a measure of relative factor endowments between two trading partners is $DGDPpc_{ij}^{t} = \left|\ln GDPpc_{i}^{t} - \ln GDPpc_{j}^{t}\right|$. The remaining model variables are as before.[8]

The reference group of countries in the panel comprise bilateral export flows from 12 EU countries[9] to 20 OECD trading partners[10] over the period 1992–2003, with Belgium and Luxembourg treated as a single country. The data sources for both the traditional and NTT models are as follows. Nominal export flow data, denominated in US dollars, are from the International Monetary Fund's (IMF) Direction of Trade Statistics and are expressed in real terms based on US producer prices (2000 = 100), sourced from the IMF's *International Financial Statistics*. Data on Gross Domestic Product (GDP) and GDP per capita at constant 2000 US dollars are sourced from the World Bank's *World Development Indicators*. The binary-coded EU dummy variable takes the value of 1 when both countries of each country-pair observation are members of the EU, otherwise it is zero. The designated values hold for member countries throughout the sample period; for Austria, Finland, and Sweden, values of unity are assigned only after gaining official membership of the EU in 1995.

The FE variants of the gravity model place restrictions on the generalised FE equations given in (5) and (6). The specification by Egger and Pfaffermayr (2003) imposes the restriction that the exporter-time and importer-time interactions equal zero, $\gamma\theta_{i}^{t} = \omega\theta_{j}^{t} = 0$. Further restrictions are added by the triple-indexed specification; since it is a special case of the combined model it also adds country-pair restrictions on the intercept, $\gamma\omega_{ij} = 0$. If the space and time dimensions of the panel are ignored, then no degree of heterogeneity is allowed in the model and therefore: $\theta^{t} = \gamma_{i} = \omega_{j} = \gamma\omega_{ij} = \gamma\theta_{i}^{t} = \omega\theta_{j}^{t} = 0$.

[8] In estimating the core gravity model variables in the NTT model, a measure of shipping costs, given by the difference between the value of exports free on board (fob) by the exporting country and the value of imports cost, insurance and freight (cif) by the importing country is also included as a time-varying alternative to the geographic distance between the economic centres of the exporting and the importing countries.

[9] Austria, Belgium-Luxembourg, Denmark, Finland, France, Germany, Italy, the Netherlands, Spain, Sweden, Switzerland, and the United Kingdom.

[10] Austria, Belgium-Luxembourg, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Korea, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom, and the United States.

Traditional estimation of the gravity model by POLS (for all available years) ignores the dimensions of the panel dataset and in general explicitly includes several time-invariant variables:

$$Exp_{ij}^{t} = \alpha_0 + \eta_1 GDP_i^t + \eta_2 GDP_j^t + \eta_3 GDPpc_i^t + \eta_4 GDPpc_j^t$$
$$+ \eta_5 DIST_{ij} + \eta_6 ADJ_{ij} + \eta_7 LANG_{ij} + \eta_8 EU_{ij}^t + \nu_{ij}^t \quad (7)$$

where the additional variables in the traditional model consist of the geographic distance, measured in kilometres, between the capital cities of the exporting and importing countries, $DIST_{ij}$; and two dummy variables reflecting adjoining land borders, ADJ_{ij}, and a common language, $LANG_{ij}$, as a proxy for cultural and historical links between trading partners, all of which are sourced from the CEPII.[11] Equation (7) will be without the time superscripts if estimated as a cross-sectional (for a single year) OLS regression.

The time-invariant variables included in the gravity model (equation 7) – distance, adjacency of national borders, and a shared language – cannot be separately estimated by the standard and 2-way FE variants since only the within-group variation (over time) is used in forming the estimator (Hsiao 2003).[12] Indeed, Cheng and Wall (2005) advocate using the FE model since bilateral trade determinants such as historical, cultural, ethnic, political, and geographic factors are often difficult to observe and quantify. Furthermore, they assert that the FE estimator eliminates the need to include the geographic distance variable in the regression, a desirable aspect of the FE approach because of the shortcomings of distance as a measure of transport and information costs.[13] In capturing the effects of all omitted variables that are unit-specific but remain constant over time, the FE model effectively subsumes the time-invariant variables into the country-pair fixed effects.

[11] Le Centre d'Etudes Prospectives et d'Informations Internationales, available at http://www.cepii.org.

[12] If the coefficient estimates of the time-invariant variables are of interest, they can be obtained either by instrumental variables (IV) estimation or by using a three-stage regression (Plümper and Troeger 2007).

[13] The geographic distance between capital cities implicitly assumes that overland transport costs incur comparable charges as overseas transport costs whereas in reality the cost of cross-border trade will vary depending on the transport mode. In addition, the straight-line distance assumes only one economic centre per country but in fact a large country may have several economic centres.

IV. EMPIRICAL RESULTS

Tables 1 and 2 present the results for the traditional and NTT specifications of the gravity model respectively. Four variants of the FE approach are estimated: the triple-indexed specification proposed by Mátyás (1997), the standard FE model used by Hummels and Levinsohn (1995), a model which combines the elements from both the triple-indexed version and the standard FE model (Egger and Pfaffermayr 2003) and finally the generalised gravity model suggested by Baltagi *et al.* (2003), which is least likely to suffer from any bias due to omitted variables. The pooled OLS estimates, which ignore the dimensions of panel datasets, are used as a comparison with earlier studies.

The performance of both the traditional and NTT models in terms of goodness-of-fit is highly satisfactory with the independent variables explaining a high proportion of the variance of the dependent variable, ranging from 84% to 99% as additional degrees of heterogeneity are allowed in the model, but the empirical success of the gravity equations is an insufficient condition to justify their use; the econometric properties of the models are also worthy of attention.

Specifically, all main effects and interactions are significant at conventional levels, as indicated by the F-tests, with the exception of the time-specific effects for the full effects specification in Table 2. In line with the prescriptions of Baltagi *et al.* (2003), a gravity model which excludes one or more significant interaction effects risks omission bias and inconsistency of the regression coefficients. Therefore the unrestricted version of the FE approach is preferred,[14] as given by the full effects design, which allows for unobserved effects along several dimensions of the panel.

The joint significance of the interaction terms is also supported by the root mean squared error (RMSE) for both the traditional and NTT models, which in general declines as an increasing degree of heterogeneity is admitted into the model. For the traditional model, the RESET null of proper functional form is not rejected for the full effects design in contrast to POLS and the nested versions of the general model, implying that the full effects design is recommended. This is not true for the NTT model, according to which

[14] Note that the Hausman test rejects the appropriateness of the random effects (RE) estimator in favour of the FE specification.

the RESET test rejects all specifications except for the standard FE model.[15] In short, with the exception of the RESET test for the NTT model, the most general model is supported econometrically. In noting that the traditional and NTT models are non-nested expressions used to explain the same phenomenon, the validity of the models is tested using the *J* test (Davidson and MacKinnon 1981), according to which, the traditional model is chosen.[16]

Regarding the parameter estimates, all coefficients are correctly signed and, in general, lie within the expected range. Looking more closely at the size-related coefficient estimates, the positive and significant effect for all estimated models indicate that economic size matters for trade. The traditional trade determinants, however, are unusually high for the triple-indexed specification. Although this is not borne out for the NTT determinants related to size, it suggests that a specification using country-pair fixed effects is required to get unbiased coefficient estimates. These findings concur with the empirical results of Egger and Pfaffermayr (2003); when exporter-importer interactions are added to the triple-indexed specification, the combined model reduces to a conventional two-way FE model with time-specific effects and bilateral country-pair effects only, effectively yielding identical coefficient estimates for both models. Similarly, the main effects are rendered redundant in the full effects specification by Baltagi *et al.* (2003) in terms of the coefficient estimates, although their significance suggests their inclusion in the model is supported econometrically.

The empirical results can also be given an economic interpretation. By capturing the unobserved time-invariant variables, the country-pair effects are better aligned to the bilateral nature of trade whereas the exporter and the importer effects are already largely accounted for in a time-varying model based on GDP and GDP per capita. Thus, while the inclusion of the exporter and the importer effects are warranted econometrically in both the combined model and the full effects model, it is in fact the country-pair effects

[15] The fact that the most general model does not pass the RESET test is due to the time-varying measure of distance. Measured in this way, shipping costs carry statistical problems; since there cannot be e logarithm of a negative number, more than half of the observations are lost. When the shipping costs variable is excluded, the general model passes the RESET test, although many of the model coefficients become insignificant.

[16] This result is obtained when the shipping costs variable is excluded from the NTT model, otherwise, neither model can reject the other implying that the result is inconclusive.

that explain most of the variation of the dependent variable (Egger and Pfaffermayr 2003).

For the unrestricted FE specification in Table 1, the GDP coefficients are radically different to the restricted FE variants and are more in keeping with the POLS estimates. On the other hand, the size-related variables – total GDP and similarity of GDP – in Table 2 are much more stable across the specifications indicating that heterogeneity has more benign consequences in the NTT model. The magnitudes of the size variables in the unrestricted model are also similar to those obtained by Baltagi *et al.* (2003).

Regarding the relative endowment variables, the signs of the exporting and the importing GDP per head coefficients differ depending on the estimator used, while the former is dropped in the full effects design.[17] For the NTT model, the consistency in sign across the specifications supports Linder's hypothesis suggesting that similarity of relative factor endowments will increase trade between the OECD countries. The magnitude ranges from a high of unity (the triple-indexed model) to a low of near-zero (the combined model). The more reasonable full effects estimate is also in line with the result obtained by Baltagi *et al.* (2003).

While POLS generates the coefficient estimates for the time-invariant variables, it incurs the problem of omitted variable bias because of its failure to account for heterogeneous trading relations. Regarding the measure of transport costs in the NTT model, the time-invariant distance coefficient is much higher for POLS than the time-varying shipping costs used in the unrestricted model. The latter is consistent with Grossman (1998) who renounces a high trade-to-distance elasticity in a world of modest transport costs. In conjecturing that shipping costs are on average about 5% of the value of traded goods, the distance elasticity should be no more than –0.03.

Yet, despite the fact that the time-varying measure of transport costs is more in line with its expected trade-impeding effect, statistical differences in compiling trade data across countries means that the geographic distance measure continues to be widely used. Furthermore, distance has been used in gravity models of trade not only as a proxy for transport costs, but also as a broader measure of information costs. The fact that greater distance dampens trade might also be due to psychological distance: economic agents

[17] Collinearity is not an uncommon feature of generalised models.

generally find it more convenient to import from neighbouring countries, often for reasons related to greater availability of information or because of cultural and historical factors. The possibility that distance captures more complex phenomena was previously highlighted by Rauch (1999), who argues that greater distance is associated with greater information and search costs. This goes some way towards explaining the relatively high distance elasticity – usually ranging between –0.8 and –1.5 – that is typically found in the literature.

Of primary concern is the size and significance of the EU dummy coefficient as a way to evaluate trade policy across the different specifications used in the literature. Its potential trade-enhancing effect on trade is confirmed by the positive coefficient sign for the EU dummy across all specifications. Its magnitude is highest for the POLS and triple-indexed estimates, halves in size for the standard FE model, further decreases in size for the combined model and dwindles away to insignificance as an increasing degree of country and time heterogeneity is admitted into the model. Both the traditional and NTT models are consistent regarding the declining magnitude and loss of significance of the EU dummy coefficient.

The small magnitude of the EU dummy coefficient and its insignificance for the unrestricted model is not in keeping with the importance of trade liberalisation within the EU. According to the Single Market Review Series (European Commission 1996) the Internal Market programme brought about the removal of a number of obstacles to trade through mutual recognition including substantial progress towards dismantling technical barriers to trade, the liberalisation of public procurement and the development of simplified internal customs and fiscal controls. In short, the objective of the un-curtailed movement of goods between member states required the dismantling of trade barriers, with consequential beneficial effects on the volume of intra-EU trade

Yet, the empirical literature tends to find a limited effect of European regional integration on trade. Bussière *et al.* (2005) offer the following explanations. First, if most EU member countries have already joined the EU before the start period of the sample, then the EU dummy captures only the effect of EU expansion by those countries that entered during the sample period. In this sample, EU expansion consists of three countries, namely Austria, Finland, and Sweden, which became EU members in 1995.

Second, previous trade agreements between several EU members often coupled with historically close trading links means that EU entry may not have spurred further trade integration beyond existing high levels. For example, Austrian accession may have had little effect on trade owing to its strong trade links traditionally shared with Germany. In other cases, trade expansion may well have been anticipated before actual EU entry, for example, Portugal and Spain experienced marked trade increases before official EU membership.

The explanations for the general insignificance and small magnitude of the EU effect on trade is upheld for most specifications used by Bussière *et al.* (2005), including for the fixed effects model, the random effects (RE) model and a dynamic OLS specification for a large sample. For a sub-sample of OECD countries, however, they obtain a positive and significant coefficient estimate for the EU dummy using the two-way FE estimator. This result concurs with the findings of this paper in so far as a positive and significant effect is also obtained for an OECD-based sample using the same estimator. The results of this paper additionally show that the effect of regional integration, estimated by the nested versions of the most general FE model, is lower than previously found, as indicated by the POLS estimates. This is consistent with recent findings. For example, by allowing a time trend to differ across country-pairs, Bun and Klaassen (2007) conclude that the Euro effect is smaller than is generally perceived. Finally, the results of this paper also show that once the gravity model controls for the various dimensions of the panel as given by the full effects design, the effect of European regional expansion becomes insignificant. Consistent for both the traditional and the NTT models, this finding points to the importance of the econometric properties of the model when evaluating trade policy effects.

V. SUMMARY AND CONCLUSIONS

With the rising number of regional agreements, especially since the early 1990s, the effect of trade policy on trade flows has received much attention. The gravity model of international trade typically forms the basis for the empirical estimations wherein dummy variables are used to capture the expected positive effects of bilateral trade agreements between signatory countries. The popularity of the gravity model stems from its empirical success in explaining trade patterns coupled with its versatility in application. The empirical results, however, may not be reliable. This is because various estimation strategies have been undertaken to estimate the gravity model, many of which potentially suffer from bias due to the exclusion of variables – either inadvertently or because of difficulties in observing and quantifying relevant variables. Ascribing the gravity model with theoretical foundations in a cross-sectional context and controlling for heterogeneity in a panel context comprises the main strands in the literature to counter the problem of omitted variable bias.

Using a panel dataset of bilateral export flows from 12 EU countries to 20 OECD trading partners over the years 1992–2003, the coefficient estimates for two gravity models are compared – one based on the traditional trade determinants, the other based on newer trade theories (NTT) – to evaluate the effect of European regional integration on trade. Specifically, several variants of the FE model are estimated to allow for differing degrees of space and time heterogeneity: the triple-indexed specification, the standard FE model, a model which combines the elements from both versions, and a generalised gravity model, which is least likely to suffer from bias. The gravity model estimated by POLS is used to compare the results with earlier studies.

The results indicate that the coefficient estimates are plausible in sign and significance across all specifications. Focusing on the econometric properties, however, the preferred model constitutes the full effects design which allows for the greatest degree of heterogeneity. Of most interest is the effect of European regional integration on trade. The positive and significant coefficient of the EU dummy variable by POLS declines in magnitude as an increasing degree of heterogeneity is allowed in the model and becomes insignificant for the full effects model. This result is consistent for both the traditional and NTT models. These findings emphasise the need to correctly control for heterogeneity in a panel setting, otherwise the coefficient estimates and standard errors are potentially biased. Hence, an analysis of the econometric properties – not simply its empirical success – should justify the exact specification of the gravity model. The findings point to the importance of the econometric specification and estimation method when evaluating trade policy effects. The implications for EU trade policy depend on the specification of the model, making the effect of integration very difficult to quantify.

REFERENCES

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

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

Baier, S. L., and Bergstrand, J. H. (2007), Do Free Trade Agreements actually Increase Members' international Trade?, *Journal of International Economics*, 71, pp.72-95.

Balassa, B. (1976), Types of Economic Integration, in F. Machlup (ed.) *Economic Integration: Worldwide, Regional, Sectoral*. London: Macmillan.

Baldwin, R. E. (1994), *Towards an Integrated Europe*, London: CEPR.

Baltagi, B. H., Egger, P. and Pfaffermayr, M. (2003), A Generalized Design for Bilateral Trade Flows Models, *Economics Letters*, 80, pp. 391-397.

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

Bergstrand, J. H. (1989), The Generalized Gravity Equation, Monopolistic Competition, and the Factor-Proportions Theory in International, *Review of Economics and Statistics*, 71, pp. 143-153.

Bun, M. J. and Klaassen, F. J. (2007), The Euro Effect is not as Large as Commonly Thought, *Oxford Bulletin of Economics and Statistics*, 69, pp. 473-496.

Bussière, M., Fidrmuc, J. and Schnatz, B. (2005), Trade Integration of Central and Eastern European Countries: Lessons from a Gravity Model, ECB Working Paper Series, No. 545.

Cheng, I. and Wall, H. J. (2005), Controlling for Heterogeneity in Gravity Models of Trade and Integration, *Federal Reserve Bank of St. Louis Review*, 87, pp. 49-63.

Davidson, R. and MacKinnon, J. G. (1981), Several Tests for Model Specification in the Presence of Alternative Hypotheses, *Econometrica*, 49, pp. 781-793.

Egger, P. and Pfaffermayr, M. (2003), The Proper Panel Econometric Specification of the Gravity Equation: A three-way Model with Bilateral Interaction Effects, *Empirical Economics*, 28, pp. 571-580.

Feenstra, R. C. (2003), *Advanced International Trade: Theory and Evidence*, Princeton: Princeton University Press.

Greenaway, D. and Milner, C. (2002), Regionalism and Gravity, *Scottish Journal of Political Economy*, 49, pp. 574-585.

Grossman G. (1998), Comment. In Frankel, J. (ed.), *The Regionalization of the World Economy*, Chicago and London: The University of Chicago Press.

Gruber, W. H. and Vernon, R. (1970), The Technology Factor in a World Matrix. In Vernon, R. (ed.), *The Technology Factor in International Trade*. New York: Columbia University Press.

Harrigan, J. (2001), Specialization and the Volume of Trade: do the Data obey the Laws? NBER Working Paper, No. 8675.

Hausman, J. A. (1978), Specification Tests in Econometrics, *Econometrica*, 46, pp. 1251-1271.

Helpman, E. (1987), Imperfect Competition and International Trade: Evidence from Fourteen Industrialized Countries, *Journal of the Japanese and International Economies*, 1, pp. 62-81.

Helpman, E. and Krugman, P. R. (1985), *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*, Cambridge: MIT Press.

Hsiao, C. (2003), *Analysis of Panel Data*, Cambridge: Cambridge University Press.

Hummels, D. and Levinsohn, J. (1995), Monopolistic Competition and International Trade: Reconsidering the Evidence, *Quarterly Journal of Economics*, 110, pp. 799-836.

Linder, S. B. (1961), *An Essay on Trade and Transformation*, New York: John Wiley and Sons.

Linnemann, H. (1966), *An Econometric Study of International Trade Flows*, Amsterdam: North-Holland.

Mátyás, L., (1997), Proper Econometric Specification of the Gravity Model, *The World Economy*, 20, pp. 363-368.

McCallum, J. (1995), National Borders Matter: Canada-U.S. Regional Trade Patterns, *American Economic Review*, 85, pp. 615-623.

Plümper, T. and Troeger, V. E. (2007), Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects, *Political Analysis*, 15, pp. 124-139.

Pöyhönen, P. (1963), A Tentative Model for the Volume of Trade between Countries, *Weltwirtschaftliches Archiv*, 90, pp. 93-100.

Ramsey, J. B. (1969), Tests for Specification Errors in Classical Linear Least Squares Regression Analysis, *Journal of the Royal Statistical Society B*, 32, pp. 350-371.

Rauch, J. E. (1999), Networks versus Markets in International Trade, *Journal of International Economics*, 48, pp. 7-35.

Tinbergen, J. (1962), *Shaping the World Economy*, New York: Twentieth Century Fund.

Wang, Z. K. and Winters, L. A. (1991), The Trading Potential of Eastern Europe. CEPR Discussion Paper, No. 610.

Wooldridge, J. M. (2002), *Econometric Analysis of Cross-Section and Panel Data*, Cambridge: MIT Press.

White, H. (1980), A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, 48, pp. 817-838.

Table 1: Gravity Model of Traditional Determinants of Export Flows

Regressors	POLS[a]	Triple-indexed[a]	Standard FE[a]	Combined[a]	Full effects[a]
	Restricted Models				Unrestricted Model
GDP_i^t	0.76** (78.03)	3.11* (1.72)	2.31** (5.40)	2.56** (5.92)	0.97** (47.38)
GDP_j^t	0.80** (91.44)	3.18** (3.26)	2.21** (6.57)	1.91** (5.65)	0.52** (28.49)
$GDPpc_i^t$	0.17** (3.88)	–3.07* (–1.63)	–2.49** (–5.22)	–2.11** (–4.45)	–
$GDPpc_j^t$	–0.37** (–13.64)	–2.37** (–2.10)	–1.39** (–3.65)	–0.90** (–2.37)	0.98** (7.05)
$DIST_{ij}$	–0.83** (–55.56)	–	–	–	–
ADJ_{ij}	0.48** (14.38)	–	–	–	–
$LANG_{ij}$	0.18** (5.97)	–	–	–	–
EU_{ij}^t	0.30** (12.48)	0.34** (6.91)	0.17** (11.70)	0.09** (6.27)	0.03 (1.60)
INPT	–12.18** (–19.34)	–93.14** (–2.80)	–61.20** (–8.58)	–68.46** (–7.98)	–29.25** (–17.99)
Nr of obs	2709	2709	2709	2709	2709
R^2	0.893	0.841	0.989	0.991	0.996
RMSE	0.460	0.564	0.150	0.137	0.103
RESET[b]	17.76**	27.36**	4.83**	8.68**	2.37
Hausman[c]	–	–	114.18**	–	–
Wald tests for the main effects and interactions:					
Time	–	2.09**	–	52.86**	3.26**
Source	–	100.61**	–	84.59**	70.51**
Host	–	181.43**	–	1305.02**	1081.13**
Source–Host	–	–	325.73**	552.43**	325.73**
Source–Time	–	–	–	–	8.19**
Host–Time	–	–	–	–	6.84**

[a] The reported test-statistics in parentheses are heteroskedasticity robust (White 1980).
[b] Using powers of the predicted dependent variable (Ramsey 1969).
[c] Based on the difference between the FE and the RE estimators (Hausman 1978).
** denotes significance at the 5% level; * denotes significance at the 10% level.

Table 2: Gravity Model of New Trade Theory Determinants of Export Flows

Regressors	Restricted Models POLS[a]	Restricted Models Triple-indexed[a]	Restricted Models Standard FE[a]	Restricted Models Combined[a]	Unrestricted Model Full effects[a]
$TGDP_{ij}^{t}$	1.50** (110.00)	1.68** (3.74)	1.33** (30.53)	2.18** (15.21)	1.89** (5.80)
$sGDP_{ij}^{t}$	0.81** (42.78)	0.93** (4.12)	0.73** (7.86)	0.91** (10.23)	1.22** (3.37)
$DGDPpc_{ij}^{t}$	–0.04 (–1.11)	–1.02** (–9.19)	–0.04 (–0.47)	–0.03 (–0.33)	–0.47** (–2.25)
$DIST_{ij}$	–0.74** (–53.53)	–	–	–	–
ADJ_{ij}	0.54** (17.44)	–	–	–	–
$LANG_{ij}$	0.19** (6.32)	–	–	–	–
$SHIP_{ij}^{t}$	–	–	–	–	–0.02** (–3.99)
EU_{ij}^{t}	0.40** (17.84)	0.22** (4.39)	0.13** (9.37)	0.06** (4.56)	0.01 (0.31)
$INPT$	–13.65** (–38.80)	–25.37** (–2.11)	–15.96** (–13.04)	–39.01** (–9.99)	–30.44** (–3.57)
Nr of obs	2709	2709	2709	2709	1265
R^2	0.885	0.853	0.989	0.991	0.997
RMSE	0.477	0.543	0.154	0.138	0.105
RESET[b]	8.77**	27.01**	0.73	10.53**	6.09**
Hausman[c]	–	–	47.20**	–	–
Wald tests for the main effects and interactions:					
Time	–	2.89**	–	59.57**	0.65
Source	–	101.45**	–	129.29**	39.25**
Host	–	191.89**	–	1074.40**	604.81**
Source–Host	–	–	606.26**	450.73**	378.79**
Source–Time	–	–	–	–	2.94**
Host–Time	–	–	–	–	9.49**

[a] The reported test-statistics in parentheses are heteroskedasticity robust (White 1980).
[b] Using powers of the predicted dependent variable (Ramsey 1969).
[c] Based on the difference between the FE and the RE estimators (Hausman 1978).
** denotes significance at the 5% level; * denotes significance at the 10% level.