Proper Estimation of Cross-Section Gravity Models: A Note

Peter Egger*

April 4, 2001

Abstract

This paper suggests to follow a two-way Hausman-Taylor panel approach when estimating cross-section gravity models. In contrast to OLS and (usually) also the random effects model, this framework provides consistent estimates. In contrast to the fixed effects approach, it allows to estimate parameters of variables such as GDP or GDP per capita, which vary only in a single dimension. The Hausman-Taylor model is therefore the most appropriate estimator of a cross-section gravity model.

Key words: Gravity model; Panel econometrics

JEL classification: C33; F14; F15

*Austrian Institute of Economic Research, Arsenal, Objekt 20, P.O. Box 91, A-1030 Vienna, Austria. Phone: +43-1-798-2601-475; E-mail: Peter.Egger@wifo.ac.at.

1 Introduction

The gravity model of bilateral trade is one of the most deeply analyzed theoretical concepts in international trade (see Anderson, 1979; Bergstrand, 1985, 1989, 1990; Helpman & Krugman, 1985; etc.). Additionally, its empirical application has produced an enormous amount of research on the effects of economic integration. More recently, deeper insights in the proper econometric analysis have been provided and have demonstrated that panel econometric methods are able to obtain better interpretable and more consistent estimates than their OLS counterparts. So far, this literature has concentrated on bilateral trade data of three main dimensions: exporter country, importer country and time (compare Mátyás, 1997; Egger, 2000). Although this is useful given the nature of typical aggregate trade data sets, it does not directly refer to the empirical tradition, where cross-section analysis is the most common practice (see Oguledo & MacPhee, 1994, for an overview; and Nilsson, 2000, for a recent application).

This paper provides insights, that the proper analysis of cross-section bilateral trade flows corresponds to a two-way panel analysis with exporter and importer country effects. A two-way model in the tradition of Hausman & Taylor (1981) and Wyhowski (1994) is suggested to be the most appropriate one. This model has three advantages:

1. If exporter and importer effects are important, it is more consistent than OLS (otherwise it is equivalent).

2. If some of the explanatory variables are correlated with the error term, it is more consistent than the random effects model (otherwise it is

equivalent).

3. In contrast to the fixed effects model, it provides parameter estimates of variables, which vary only in a single dimension (such as GDP or GDP per capita).

I estimate a panel of OECD exports in 1997 and find that traditional OLS yields considerably biased parameter estimates especially for distance, common borders and importer GDP, whereas the Hausman-Taylor model is consistent.

2 The Empirical Model: A Two-Way Hausman-Taylor Approach

I estimate a specification, which - besides two additional trade preference variables - corresponds to the familiar gravity model as estimated by Hamilton & Winters (1992), Baldwin (1994), Nilsson (2000), and others:[1]

$$\begin{aligned} E_{ij} = \beta_o + \beta_1 Y_i + \beta_2 Y_j + \beta_3 (Y/P)_i + \beta_4 (Y/P)_j \\ + \beta_5 D_{ij} + \beta_6 B_{ij} + \beta_7 L_{ij} + \beta_8 V_{ij} + \beta_9 R_{ij} + u_{ij} \end{aligned} \quad (1)$$

With $u_{ij} = x_i + m_j + v_{ij}$ and all continuous variables already in logs. Subscript i refers to exporter countries and j to importer countries. E measures nominal bilateral exports, Y is nominal GDP, P is population, D is distance between capitals (or economic centers) in miles, B and L are dummy

[1] Baldwin (1994) estimates a random effects model on a data set with time variation. However, his approach is similar in terms of the explanatory variables in use.

variables, which are set to 1 if two countries face common borders (B) or a common language (L). Finally, V and R measure trade impediments in terms of bilateral viability of contracts (V) and rule of law (R), see below for a more detailed description of the sources and definitions of variables.

As easily can be seen, GDP (Y) and GDP per capita (Y/P) vary only in a single dimension. Usually, (1) is estimated by OLS, and both x_i and m_j are set to zero, which is a testable restriction. If one applies panel econometric methods (which is more proper in most circumstances) and treats x_i and m_j as fixed, both Y and Y/P are captured by the fixed effects. Implicitly, the fixed effects model (FEM) assumes that all explanatory variables are correlated with the unobserved effects, and it eliminates this correlation by the Within transformation. Instead, if x_i and m_j are treated as random, parameters for both Y and Y/P can be estimated. From this perspective, the random effects model (REM) seems preferable over the FEM. However, the REM relies on the crucial assumption that both x_i and m_j are i.i.d. with zero mean and variance σ_x^2 and σ_m^2, respectively. If there is correlation between the explanatory variables and the unobserved effects, this results in inconsistent estimates for both the variance components (σ_x^2 and σ_m^2) and the parameters, which can be tested by the familiar Hausman (1978) test. Hausman & Taylor (1981) present a model (hereafter HTM), which is able to remove the endogeneity problem by using the various dimensions of data variance in order to construct instruments exclusively from inside the model. For our purpose, we follow the lines of Wyhowski (1994), who generalizes the idea of Hausman & Taylor for the two-way model.[2]

[2] This model captures the various possibilities of endogeneity of the unobserved effects.

For simplicity, assume that all variables, which vary in both dimensions (i and j), are not correlated with the unobserved effects (x_i and m_j). Cornwell et al. (1994) label such variables as *doubly exogenous*. In contrast, assume Y_i and $(Y/P)_i$ are correlated with x_i and Y_j and $(Y/P)_j$ are correlated with m_j only. According to Cornwell et al. (1994) I refer to these variables as *singly exogenous* ones.

The data are first ordered by exporters and then by importers and I denote the set of *doubly exogenous* (ij-variant) variables by $\mathbf{X}$. For our purpose, we need the following matrices: $J_x = (1/x)(\iota_x \iota_x')$ with x as the number of exporters, $J_m = (1/m)(\iota_m \iota_m')$ with m as the number of importers. $Q_4 = (1/xm) \times (\iota_{xm} \iota_{xm}')$ where xm is the number of observations, $Q_2 = (I_x \otimes J_m) - Q_4$, $Q_3 = (J_x \otimes I_m) - Q_4$, $Q_1 = I_{xm} - Q_2 - Q_3 - Q_4$. In words, premultiplying the data by Q_1 derives the familiar Within transformation, which wipes out all effects varying only in a single dimension. OLS on this model corresponds to the FEM. The corresponding error vector is v_{ij} with variance $\sigma_v^2 = \lambda_1$. Premultiplying the data by Q_2 or Q_3 yields the two Between transformations for the exporter and the importer dimension. Usually, OLS on the Between-transformed model gives errors $\mu_{x.}$ and $\mu_{m.}$ with variance $\lambda_2 = x\sigma_x^2 + \sigma_v^2$. and $\lambda_3 = m\sigma_m^2 + \sigma_v^2$. Then, one also has $\lambda_4 = x\sigma_x^2 + m\sigma_m^2 + \sigma_v^2$. With these estimates at hand, one can transform the model according to Fuller & Battese

Therefore, it is surprising that it is not implemented in econometric packages so far. The FEM and the REM form boundary cases of the HTM in terms of the assumption about correlation between regressors and residuals. The FEM assumes correlation of all regressors and the REM assumes no correlation at all (see Hausman & Taylor, 1981, or Baltagi, 1995).

(1973, 1974) in order to come up with typical elements for all variables:

$$y^*_{ij} = y_{ij} - \theta_1 \overline{y}_{i.} - \theta_2 \overline{y}_{.j} + \theta_3 \overline{y}_{..} \tag{2}$$

where $\theta_1 = 1 - (\sigma_v / \lambda_2^{1/2})$, $\theta_2 = 1 - (\sigma_v / \lambda_3^{1/2})$, $\theta_3 = \theta_1 + \theta_2 + (\sigma_v / \lambda_4^{1/2}) - 1$, and dots and bars refer to variable means. OLS on the transformed data is equivalent to GLS and obtains the REM parameter estimates. However, in case of correlation of (some of) the explanatory variables with the unobserved effects, the Between regressions are affected yielding inconsistent estimates of σ_x^2, σ_m^2, θ_1, θ_2, and θ_3. This can be avoided in the following way:

1. Estimate the Within model (i.e. FEM) and obtain Within residuals in the spirit of Amemiya (1971): $u_x = Q_2(E_{ij} - \beta_0 - \mathbf{X}\beta_k^{Within})$ and $u_m = Q_3(E_{ij} - \beta_0 - \mathbf{X}\beta_k^{Within})$.
2. Regressing u_x onto Y_i and $(Y/P)_i$ with the instruments $P_x = [Q_2\mathbf{X}, Q_2Y_j, Q_2(Y/P)_j]$ gives residuals with (consistent) variance λ_2.
3. Regressing u_m onto Y_j and $(Y/P)_j$ with the instruments $P_m = [Q_3\mathbf{X}, Q_3Y_i, Q_3(Y/P)_i]$ gives residuals with (consistent) variance λ_3.

The model can then be transformed as described in (2). Finally, 2SLS on the transformed model with the set of instruments $P = [P_v, P_x, P_m]$ with $P_v = Q_1\mathbf{X}$ yields the two-way HTM, which is consistent, if the instruments are proper (this can be tested by an over-identification test according to Hausman & Taylor, 1981) and it is more efficient than the Within model, as long as there are enough instruments for the two Between equations at hand. Amemiya & MaCurdy (1986) and Breusch et al. (1989) present expanded sets of instruments, which are also discussed in Wyhowski (1994) for the two-way model. Due to the asymptotic difficulties associated and their reduced appropriateness for unbalanced data, I apply the traditional set as suggested in Wyhowski (1994). Additionally and similar to the REM, the HTM provides parameter estimates of both the only exporter-variant and the only importer-variant variables such as GDP or GDP per capita.[3]

3 Data and Estimation Results

I estimate the above model for bilateral exports of OECD countries to OECD and non-OECD countries in 1997.[4] Nominal exports in USD are from OECD (Monthly Statistics of International Trade), IMF (Direction of Foreign Trade), and the Vienna Institute of Comparative Economic Studies (hereafter WIIW).

[3]Interestingly and as pointed out in Egger (2001), the most appropriate estimator for gravity panels which vary in the time dimension too, is a Hausman-Taylor AR(1) model. In this case, one should include interaction effects between exporter and importer effects. Then, the model boils down to a two-way framework with time effects and bilateral effects. The latter comprehensively account for all bilateral time-invariant influences. Therefore, the FEM not any longer provides parameter estimates of variables like distance or common borders. In contrast, GDP or GDP per capita do not share this problem, since they also vary in the time dimension. If one treats time effects as fixed (which is reasonable), this requests only a one-way HTM, which renders the analysis less complex as compared to the cross-section case, although this seems counter-intuitive from a first glance.

[4]**Exporting countries (OECD):** Belgium-Luxembourg, Denmark, Germany, Finland, France, Greece, UK, Ireland, Italy, Netherlands, Austria, Portugal, Sweden, Spain, Hungary, Poland, Czech Republic, Australia, Iceland, Japan, Canada, New Zealand, Norway, Switzerland, Turkey, USA, Mexico, South Korea. **Importing countries:** OECD countries as mentioned, Bulgaria, Romania, Slovakia, Slovenia, Ukraine, Algeria, Egypt, Morocco, South Africa, Argentina, Brazil, Chile, Panama, Venezuela, China, Hong Kong, India, Indonesia, Malaysia, Philippines, Singapur, Thailand, Estonia, Lithuania, Latvia.

GDP in nominal USD and population are from OECD (Economic Outlook and National Accounts Volume 1), IMF (International Financial Statistics), and WIIW. Distance is calculated as in Egger (2000), and viability of contracts and rule of law are country ratings published by the Economic Freedom Network (Economic Freedom of the World; see also the References). These two variables represent information on trade impediments, since a lower rating for viability of contracts implies that trade costs are higher. The same holds true for rule of law, which measures how likely contract breakers can be prosecuted. We expect a positive sign for the impact of each of the two variables. In sum, we come up with 1443 observations in the regression analysis.

$>$ Table 1 $<$

Table 1 presents the regression results for three different models on (1). OLS is the traditional concept applied, which omits the presence of x_i and m_j in the error term. FEM is the corresponding Within model. As easily can be seen from the F-tests on exporter and importer fixed effects, the restriction of OLS is significantly rejected. The Hausman-test indicates that there is indeed correlation of (some of) the explanatory variables with the unobserved country effects rendering the REM estimates inconsistent. Finally, the HTM is estimated assuming that only Y an Y/P are correlated with the unobserved country effects. The appropriateness of this assumption can be justified by both an intuitive and an econometric argument. Intuitively, exporter (importer) effects measure an exporter's (importer's) unobserved propensity to export, which - besides other factors - is likely to be a function of size and relative factor endowments. Econometrically, this can be tested by the over-identification test according to Hausman & Taylor (1981). The corresponding test statistic is not significant and underpins the appropriateness of our choice. Table 1 also reports details on the estimation of the variance components. First and as expected, the Between variance components (σ_x^2, σ_m^2) are larger than the Within component (σ_v^2). Second, the importer component is considerably larger than the exporter component, which is not surprising, since the variation among OECD countries should be smaller than that among OECD-plus-non-OECD countries. Table 1 reports average values for each θ. We have to bear in mind that bilateral trade panels are unbalanced, which has to be accounted for in the estimation of the variance components. Accordingly, θ_1 (θ_2) is a vector of length xm, which has exporter specific (importer specific) entries according to the available number of partner countries and we should write: θ_{1i}, θ_{2j}, and θ_{3ij} (see Baltagi, 1995, for more details on unbalanced panel data). Of course, the relatively high values for θ tell us already, that we will end up with parameter estimates for $\mathbf{X}$, which are relatively close to the FEM estimates.[5] This is also confirmed by the results, where only V_{ij} differs from the FEM in a pronounced way.[6] Finally, the comparison with OLS demonstrates that it seems worth investigating in the HTM. The parameter estimate for importer GDP doubles, the impact of distance is one third higher in absolute value, the impact of common borders is half as important as suggested by OLS, and the sign of

[5] Note that $\theta_1 = \theta_2 = \theta_3 = 0$ corresponds to the OLS model, whereas $\theta_1 = \theta_2 = \theta_3 = 1$ is the Within estimator (FEM).

[6] However, the difference is not large enough in order to reject the appropriateness of HTM as indicated by the over-identification test.

viability of contracts changes to a negative one. Additionally, the HTM is far more successful in predicting bilateral trade than OLS as indicated by the root mean squared error.

4 Conclusions

The present paper suggests to estimate cross-section gravity models by a two-way Hausman & Taylor (1981) model. This model has several advantages with respect to the traditionally applied OLS, but also the fixed effects and the random effects model. If exporter and importer country effects are important, it allows to consistently estimate parameters in contrast to OLS. In the presence of correlation between the explanatory variables and the unobserved country effects, it obtains consistent estimates in contrast to the random effects model. Finally and in contrast to the fixed effects approach, it yields estimates for variables, which vary either only in the exporter or the importer dimension such as GDP or GDP per capita. An application on bilateral OECD exports in 1997 demonstrates that OLS parameters depart considerably from the consistent Hausman-Taylor model estimates. Especially, the parameters of distance, common borders, and GDP are affected.

References

Amemiya, T. (1971), The Estimation of the Variances in a Variance-Components Model, *International Economic Review* 12, pp. 1-13.

Amemiya, T., and T.E. MaCurdy (1986), Instrumental-Variable Estimation of an Error-Components Model, *Econometrica* 54, pp. 869-880.

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

Baldwin, R. (1994), *Towards and Integrated Europe* (London: Centre for Economic Policy Research).

Baltagi, B.H. (1995), *Econometric Analysis of Panel Data* (Wiley, Chichester).

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 Factor Proportions Theory in International Trade, *Review of Economics and Statistics* 71, pp. 143-153.

Bergstrand, J. H. (1990), The Heckscher-Ohlin-Samuelson Model, The Linder Hypothesis and the Determinants of Bilateral Intra-Industry Trade, *Economic Journal* 100, pp. 1216-1229.

Breusch, T.S., G.E. Mizon, and P. Schmidt (1989), Efficient Estimation Using Panel Data, *Econometrica* 57, 695-700.

Economic Freedom Network: Free the World (http:\\www.freetheworld.com).

Egger, P. (2000), A Note on the Proper Econometric Specification of the Gravity Equation, *Economics Letters* 66, pp. 25-31.

Egger, P. (2001), An Econometric View on the Estimation of Gravity Models and the Calculation of Trade Potentials, WIFO-Working Paper 141/2001 (http://www.wifo.ac.at/Peter.Egger/rec_working.htm).

Fuller, W.A., and G.E. Battese (1973), Transformations for Estimation of Linear Models with Nested Error Structure, *Journal of the American Statistical Association* 68, pp. 626-632.

Fuller, W.A., and G.E. Battese (1974), Estimation of Linear Models with Cross-Error Structure, *Journal of Econometrics* 2, pp. 67-78.

Cornwell, C., P. Schmidt, and D.J. Wyhowski (1992), Simultaneous Equations and Panel Data, *Journal of Econometrics* 51, pp. 151-181.

Hamilton, C.B., and A.L. Winters (1992), Opening up International Trade with Eastern Europe, *Economic Policy* 14, pp. 77-116.

Hausman, J.A. (1978), Specification Tests in Econometrics, *Econometrica*

46, pp. 1251-1271.

Hausman, J.A., and W.E. Taylor (1981), Panel Data and Unobservable Individual Effects, *Econometrica* 49, pp. 1377-1398.

Helpman, E., and P.R. Krugman (1985), *Market Structure and Foreign Trade* (Cambridge, Mass.: MIT Press).

Oguledo, V.I., and C.R. MacPhee (1994), Gravity Models: A Reformulation and an Application to Discriminatory Trade Arrangements, *Applied Economics* 26, pp. 107-120.

Wyhowski, D.J. (1994), Estimation of a Panel Data Model in the Presence of Correlation Between Regressors and a Two-Way Error Component, *Econometric Theory* 10, pp. 130-139.

Table 1: Gravity Model Regression Results for Bilateral OECD Exports in 1997

Explanatory Variables:	OLS	2-way Fixed Effects (FEM)	2-way Hausman-Taylor (HTM)
Exporter GDP: Y_i	0.974 ***)	-	1.116 ***)
	0.021	-	0.311
Importer GDP: Y_j	0.778 ***)	-	1.284 ***)
	0.021	-	0.380
Exporter GDP per Capita: $(Y/P)_i$	0.340 ***)	-	0.517
	0.045	-	0.410
Importer GDP per Capita: $(Y/P)_j$	0.130 ***)	-	-0.424
	0.029	-	0.566
Distance: D_{ij}	-0.845 ***)	-1.282 ***)	-1.277 ***)
	0.030	0.041	0.040
Common Border: B_{ij}	0.646 ***)	0.321 **)	0.334 ***)
	0.155	0.128	0.125
Common Language: L_{ij}	0.293 *)	-0.082	-0.087
	0.177	0.151	0.148
Bilateral Viability of Contracts: V_{ij}	1.226 ***)	-10.218 ***)	-1.969
	0.307	3.871	2.485
Bilateral Rule of Law: R_{ij}	-0.489 ***)	1.366	0.498
	0.114	0.964	0.770
Observations	1443	1443	1443
Adj. R^2	0.792	0.875	0.845
Root Mean Squared Error	1.102	0.855	0.840
Exporter Effects: F(27, 1358)		183.590 ***)	
Importer Effects: F(53, 1358)		62.240 ***)	
Hausman Test: $\chi^2(6)$		15.420 **)	
σ_v^2		0.710	0.710
σ_x^2			1.295
σ_m^2			7.922
Average θ_{1i}			0.897
Average θ_{2j}			0.943
Average θ_{3ij}			0.890
Over-identification: $\chi^2(5)$			6.593

1) All continuous variables are in logs. Constant, exporter and importer effects are not reported in order to save space. ***) significant at 1%; **) significant at 5%; *) significant at 10%.