4.1 Introduction

Panel data provides information on the behavior of individual units, both across the individual units and over time, hence they have both cross-sectional and time series dimensions. Panel data enable us to control for heterogeneity among said individual units and gives more information about the data in terms of variability and allows for greater degrees of freedom as well as enhance greater efficiency. In panel data sets we would ordinarily assume that whereas for a given individual unit there is correlation over time, however across units there is less collinearity among the different individual units. Panel data also allows us to account for aggregation biases as well as enables the researcher to examine dynamics in relation to adjustments.

Owing to the numerous benefits of using panel-data it is not surprising that it is popular in economic literature. However, its popularity is thwarted by the fact that its limitations act as stumbling blocks in carrying out analysis. The process of design and data collection in itself is a cumbersome process and usually the time series dimension is of short-panel since it is more difficult to obtain data on one individual over a very long period of time. And even if said data were obtained there might arise distortion of measurement errors, selectivity problems and cross-section dependence acting as hindrance to rightful inference. However despite these drawbacks, the importance of panel data cannot be overemphasized.

Empirically, the use of conventional cross-section estimation is criticized since it is not able to deal with bilateral heterogeneity. Much recent empirical studies have therefore emphasized the importance of explicitly allowing for the presence of time-specific effects in order to capture business cycle effects or deal with globalization issues. This brings us to the idea of panel data methods.

This project uses panel data to specify the "gravity model" of international trade and to estimate the effects of various pertinent variables to the volume of trade. It is well known that international trade flows can be well described by a 'gravity equation' in which bilateral trade flows are a log-linear function of the incomes of and distance between trading partners. Indeed, the gravity equation is one of the greater success stories in empirical economics. Modelling and predicting foreign trade flows has long been an important task in international economics. One of the most fruitful ways to formalize this has been through the use of gravity type models.

One of the inevitable issues we face in conducting said analysis is that we do not know whether the unobserved time-invariant variables are correlated with the time-varying variables. In such conditions, the fixed effect model is suggestively most appropriate since it is almost always consistent no matter whether the true DGP is fixed or random. However with fixed effect models, all time-invariant variables get dropped.

Cheng and Wall (2002) argues that fixed-effects models are more appropriate than other panel data models. They argue that the major drawback of the fixed-effect model, that is not being able to estimate coefficients on time-invariant variables can be dealt with by estimating the regression of the individual-specific effects on the individual-specific variables by OLS. This however might not be a realistic approach since it ignores the potential correlation between the repressors and the unobserved heterogenic individual effects which may lead to biased estimates.

In contrast, Serlenga and Shin (2007) employed the Hausman-Taylor instrumental variable technique in which they capture certain degrees of cross-sectional dependence through heterogeneous time-specific factors to avoid the bias of uncorrected estimates. Their results indicate that their approach fits the data reasonably well and their estimation yields sensible results in comparison to the traditional approaches.

In this project, I thoroughly analyze the impact of the different variables on the volume of trade through the different models and draws conclusions on which is most appropriate to carry out the relevant inferences.

4.1.1 The Gravity model in detail

The gravity model is a commonly used model for estimating the impact of a wide range of policy issues in international trade in a number of related topics such as regional trading groups, currency unions and various trade distortions. In the context of international trade the gravity model has been applied since 1940s to explain the determinants of a diverse range of flows. It states that the size of trade flows between different countries is a factor of supply conditions in the point of origin, demand conditions in the point of destination and whether the forces related to the trade flows act as impediments or act as instigators between the two countries.

Despite its popularity in empirical use, until the seminal paper by Anderson (1979) the gravity model was criticized heavily on the basis that it lacked a theoretical foundation. However, afterwards it became clearer that the theoretical grounds for the gravity model can be obtained by the classical trade models such as the Ricardian Model or the Hecksher Ohlin Model or even the new trade theory models such as Increasing Returns to Scale.

The gravity model does no derive its strength from its ability to test the validity of these three trade models against each other, but rather its capacity to reconcile theory and empirics in international trade by incorporating most of the empirical phenomena.

In recent decades this model has gained popularity as well as seen a lot of controversy in terms of its econometric specification. Many argue that using traditional cross-sectional specification may be inefficient since it might not be able to capture the issues heterogeneity which comes with bilateral trade flows, thus a panel-based specification may be preferred. A panel-based approach may be able to capture heterogeneity issues by incorporating country-pair 'individual effects' and then test the assumption that unobserved individual effects are correlated with all the explanatory variables.

4.2 Method and Data

We argue that the proper specification of a panel gravity model should include main (exporter, importer, and time) as well as time invariant exporter-by-importer (bilateral) interaction effects. The econometric representation of the gravity model I will use in this project takes the form of a triple-indexed model which is given below:

$$\boldsymbol{lnY_{ijt} = \alpha_i + \gamma_j + \lambda_t + \beta_1 lnX_{ijt} + \beta_2 lnX_{it} + \beta_3 lnX_{jt} + \beta_4 lnZ_{ij} + U_{ijt}}$$

For i, j = 1,……….,N i≠j t= 1,……, T

Where:

Y_{ijt}= volume of trade from home country I to target country j at time t

X_{ijt}= explanatory variables with variations in all three dimentions

X_{it}, X_{jt}= explanatory variables with variations in i or j and t.

Z_{ij}= explanatory variables that do not vary over time but might vary in i and j

α_i, γ_j= local country and target country effect respectively that might be correlated with other explanatory variables

λ_t= time effect common to all cross sectional units

U_{ijt}= white noise disturbance term

The major source of data for this project is the Journal of Applied Econometrics Data Archive. This was the same dataset used by Serlenga and Shin (2007) whose paper was published in the Journal of Applied Econometrics.

There 91 observations of country pairs in Europe (European Union) who trade with each other. The country samples include all of the 15 EU countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxemburg, Netherlands, Portugal, Spain, Sweden and United Kingdom where Belgium and Luxemburg are treated the same thereby resulting in the 91 pairs.

The observations cover the period from 1960 to 2001, there are 6 time varying explanatory variables, 3 time invariant explanatory variables and 6 time specific common factors. We must note that time invariant explanatory variables have zero within variation and individual invariant explanatory variables have zero between variations.

Since it is a panel data the data organization is a little more intricate than what we have seen so far. The first 42 columns represent all time observations for the first country pair, the next 42 represents all time observations for the second country pair and so on until the 91st country pair of the whole sample. In order for easier understanding of the nature of the data, I provided the detailed description of the data in the below:

trade: Sum of logged exports and imports, bilateral trade flow

gdp: the sum of logged real gross domestic products

sim: a measure of similarity between two trading countries

rlf: a measure of relative factor endowments

rer: the logged bilateral real exchange rate

cee: a dummy equal to 1when both belong to European community

emu: a dummy equal to 1 when both adopt the common currency

dist: a geographical distance between the capital cities of the two countries

bor: a dummy equal to 1 when the trading partners share a boarder

lan: a dummy equal to 1 when both countries speak the same language

rert: a log of real exchange rates between European currencies and the US dollar

ftrade, fgdp, fsim, frlf and frer are the time specific common factors (individual means) of the variables **trade,gdp,sim,rlf** and **rer**.

4.2.1 Between and within variation in the data.

From the table below, we observe that the first two variables are just used to represent the cross-section dimensions of the data hence their descriptions are not meaningful. So our variables of interest are the real variables which are trade, the dependent variable and all other variables which are the explanatory variables.

Looking at the variable of interest, trade, we notice that there is more between variation which is from one country pair to the next, than within variation which is the variation of particular country pairs over time which tells us there is more variation in trade flows between one particular pair to the other, as compared to variation in the same pair over time. This pattern runs through the dataset since we observe that the between effects are greater than the within effects.

We notice that for Border, Language and Distance the within variation is zero, which means that the variables are time-invariant which intuitively makes sense because we cannot expect common borders, similarity in language or the distance between two countries to change over time.

4.3 Econometric estimations

Given the general econometric model given above, I will estimate the pooled OLS, the random and fixed effects specifications as well as the group mean specification.

4.3.1 Pooled OLS

Even though we are choosing panel-based estimation over cross sectional estimation we might not be able to get rid of heterogeneity by using a pooled regression model since it does not deal with issues of heterogeneity which might lead to model misspecification and possibly affect inference. The pooled model is the regular OLS estimation specifies constant coefficients. I run two versions of this model, one with the regular standard errors and the other with the robust standard errors. Below is the general form of the OLS:

$$Yit=\alpha+\beta Xit+uit$$

Where Yit is the dependent variable, Xit is the independent variable and α is the constant term and β's are coefficients of the explanatory variables.

The major problem with this model is the inability to distinguish between the different pairs of countries that constitute our variables of interest. This implies that by combining all the pairs of countries by pooling we deny the heterogeneity or individuality that may exist among them. So if heterogeneity differ across the individual units, the unobserved heterogeneity induces autocorrelation, therefore any latent effects left out of the model will carry across all periods. This model is oversimplified by assuming that the individual units are homogenous and this may result in autocorrelation in panel data analysis which is possibly why pooled regressions are not widely used in the literature.

In this regression, I will pool all the observations together and run the regression model, not recognizing the cross section and time series nature of the data. I am estimating this model because I would use the results when the Hausman Test I will conduct later in the paper indicates that the Random Effect model is more appropriate than the Fixed Effects Model out of the two individual-specific effects models. Then I could undertake the Breusch-Pagan Lagrange Multiplier test to determine whether the Pooled OLS model is more appropriate than the Random Effects model. If the test results concludes that the Pooled OLS is better, then I would conduct inference using that output.

4.3.1.1 The Group Means Estimator (Between Effects)

The pooled regression model can be estimated using sample means. The group mean estimator may be useful when the explanatory variables are measured with error, the group means estimator will average it out, if the error is random OLS will not be consistent.

4.3.2 Individual-specific effects models

Taking into account the limitations of the pooled OLS, an individual-specific effects model might be a preferred alternative since it can deal with the issue of heterogeneity. In this kind of model we assume that the unobserved heterogeneity across individuals are captured by αi.

This model has the form:

$$Yit = X'it\beta + (\alpha i + \varepsilon it)$$

If the individual-specific effects are correlated with the explanatory variables we use the fixed-effect model or else we use the random effects model.

4.3.2.1 Fixed Effect Model

The fixed effect model allows for heterogeneity or individuality among the different pairs of countries by having individual intercept values. Although the intercept may vary across the pairs, it does not vary across time thus making it time invariant. So, in the general model above, the fixed effect model has the αi possibly correlated with $X'it$ where $X'it$ could be endogenous. Hence this model is able to consistently estimate β for the time varying $X'it$.

Since the cultural, historical and political factors among the different countries are difficult to observe and measure, allowing each pair of countries to have their own dummy variables may lead to the inclusion of country-pair effects. While the dummy variables may be correlated with both the bilateral trade and the explanatory variables, the country pair effects allows for separation according to the direction of trade.

However, the time invariant issue of fixed effects does not allow for estimating coefficients on time invariant variables such as language dummies, distance or common border even though incorporating the effect of these can be quite important in certain contexts. In addition to dealing with unobserved heterogeneous individual effects we also need to deal with the correlation of those effects with both time invariant and time varying explanatory variables, so that no potential bias arises which will possibly affect inference.

I will therefore examine whether these time invariant variables have an effect on our model by dropping them from my estimation of the fixed effects model and running the regression again.

4.3.2.2 Random Effect Model

If the unobserved individual heterogeneity is assumed to be uncorrelated with the included variables then we may use the random effects model which essentially specifies that the explanatory variables have a common mean value for the intercept. What that means is that the individual specific-effects are distributed independently of the explanatory variables, hence it should be included in the error term. So, in the general model above, the random effect model assumes that αi is purely random and that $X'it$ must be exogenous. Hence this model is able to correct standard errors for auto correlated clustered errors and conduct predictions.

4.3.3 Presentation and discussion of results

Following economic theory GDP should have a positive relationship with trade flows. GDP is supposed to have a positive impact on the amount of trade that flows into a country. RLF measures the difference in terms of relative factor endowments between two countries. The larger the difference, the higher the volume of inter-industry trade will be, the opposite is true. In fact all the other explanatory variables are expected to have a positive impact on trade with the exception of distance. The effect of transportation costs proxied by geographical distance between capital cities. The further the distance between two countries the less likely they are to trade, all else constant. EMU is a situation where both countries adopt the same currency for transactions and it is expected to have a positive impact on trade flows. A single currency will reduce the transaction costs of trade within member countries. The common language dummy, LAN has a value equal to one when both countries speak the same official language and is meant to capture similarity in cultural and historical backgrounds between trading countries. Another consideration is the impact of bilateral real exchange rates, RER, which is defined as the price of the foreign currency per the home currency unit and is meant to capture the relative price effects. A depreciation of the home currency relative to the foreign currency (an increase in RER) is expected to lead to more export and less import for the home country. The effect of real exchange rates on total trade flow will be positive if the export component of the total trade is significantly larger than the import component, the opposite is true.

Table 4.1. Estimates from the five regression outputs

Variable	OLS	OLS_rob~t	Random_~s	Between	Fixed_E~s
gdp	1.54	1.54	1.54	1.49	1.54
	0.01	0.01	0.01	0.14	0.01
sim	0.85	0.85	0.85	1.31	0.84
	0.02	0.02	0.02	1.58	0.02
rlf	0.02	0.02	0.02	0.76	0.02
	0.01	0.01	0.01	0.16	0.01
rer	0.09	0.09	0.09	0.43	0.09
	0.00	0.00	0.00	0.06	0.00
cee	0.23	0.23	0.23	0.53	0.17
	0.02	0.02	0.02	0.12	0.03
emu	0.21	0.21	0.21	-0.13	0.21
	0.05	0.04	0.05	0.12	0.07
dist	-0.68	-0.68	-0.68	(omitted)	-0.70
	0.02	0.02	0.02		0.02
bor	0.53	0.53	0.53	(omitted)	0.54
	0.03	0.03	0.03		0.03
lan	0.25	0.25	0.25	(omitted)	0.26
	0.03	0.03	0.03		0.03
rert	0.18	0.18	0.18	-0.32	(omitted)
	0.02	0.02	0.02	0.08	
_cons	-10.93	-10.93	-10.93	-18.15	-9.94
	0.24	0.26	0.24	3.42	0.26
N	3822	3822	3822	3822	3822
r2	0.91	0.91		1.00	0.89
r2_a	0.91	0.91		1.00	0.89
rmse	0.58	0.58	0.58	0.06	0.57

legend: b/se

From the table above, observe that the signs of the coefficients in all the five models are the same, which indicates the relationship between the variables are the same across models. The estimates in the table above have been rounded to two decimal places which makes the estimates look exactly the same but I have presented the individual regression outputs in Appendix 3, so that I can explain certain variations in the results. From the results displayed in the appendix, the coefficients of the Fixed Effects Model and the Random Effects model are very similar, whereas the coefficients of the Pooled OLS is much smaller if we look at the GDP and SIM coefficients, where as it is larger if we look at the rer and emu coefficients. The distinction between the pooled and the individual-specific effects are expected, since by pooling the regression assumes that there is no relationship between the time-invariant variables and the time-varying variables and also robs the model of its heterogeneity by not treating it as the correctly specified panel-form. Looking further into the table, we notice that the coefficients of SIM and RLF are almost identical in the two models. This means that the similarity between countries and factors of endowment are exogenous to other variables. From Appendix 3 we find the probability value is less than 0.05 (in all the models estimated) which means all the coefficients of these models are not equal to zero (jointly significant).We also notice that all the explanatory variables are statistically significant variables in explaining the dependent variable, Trade. There is a positive association between all the independent variables and Trade.

Furthermore, I move on to consider the time invariant variables and drop them while estimating my fixed effect model to test whether they really have an effect on the model. I estimate all five models once more. The estimates are presented in the table below:

Table 4.2 Testing for time invariant variables

Variable	OLS	OLS_rob~t	R_E	F_E
gdp	1.54	1.54	1.54	1.72
	0.01	0.01	0.01	0.02
sim	0.85	0.85	0.85	0.99
	0.02	0.02	0.02	0.02
rlf	0.02	0.02	0.02	-0.10
	0.01	0.01	0.01	0.01
rer	0.09	0.09	0.09	0.15
	0.00	0.00	0.00	0.00
cee	0.23	0.23	0.23	0.55
	0.02	0.02	0.02	0.03
emu	0.21	0.21	0.21	0.25
	0.05	0.04	0.05	0.09
dist	-0.68	-0.68	-0.68	
	0.02	0.02	0.02	
bor	0.53	0.53	0.53	
	0.03	0.03	0.03	
lan	0.25	0.25	0.25	
	0.03	0.03	0.03	
rert	0.18	0.18	0.18	(omitted)
	0.02	0.02	0.02	
_cons	-10.93	-10.93	-10.93	-16.09
	0.24	0.26	0.24	0.23
N	3822	3822	3822	3822
r2	0.91	0.91		0.81
r2_a	0.91	0.91		0.81
rmse	0.58	0.58	0.58	0.74

legend: b/se

We can clearly observe from this table that the estimates for the fixed effect model has changed. We should however note that these are the full models and that I will present the basic model in the appendix. (See Appendix 4)

4.4.1 Hausman Test

The Hausman Test is a test that helps us to choose between the fixed effect and the random effects model. The Hausman test-statistics can be only calculated for the time-varying explanatory variables. The test is given by:

$$H=(\beta RE^{\hat{}}-\beta FE^{\hat{}})'(V(\beta RE^{\hat{}})-V(\beta FE^{\hat{}}))(\beta RE^{\hat{}}-\beta FE^{\hat{}})$$

The degrees of freedom are equal to the number of regressors for the time-varying explanatory variables. If I find a statistically significant p-value, then I will reject the null and use the fixed-effects model, otherwise I will use the random-effects model.

The test follows the following hypothesis:

Fixed effect model is consistent under Ho and Ha

Random effect inconsistent under Ha, efficient under Ho

Results from the Hausman Test is presented in the table below:

Table 4.3 Hausman Test

	Coefficients (b) fixed	(B) .	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
gdp	1.721463	1.543887	.1775764	.0091454
sim	.9908631	.8494035	.1414596	.0130911
rlf	-.1048688	.0249879	-.1298567	.0056637
rer	.1526907	.0906828	.0620079	.0018494
cee	.5457236	.233327	.3123966	.0215762
emu	.2546195	.2137038	.0409156	.0749959

```
                 b = consistent under Ho and Ha; obtained from xtreg
  B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test:  Ho:  difference in coefficients not systematic

              chi2(6) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =    17011.11
            Prob>chi2 =      0.0000
```

.

From the results above we find the probability value is less than 5% therefore we can reject the null hypothesis. This means that according to the Hausman Test, fixed-effects model is the most appropriate model to use in this analysis.

4.4.2 Mundlak's Approach

Since our errors are heteroskedastic and contain serial correlation, we can adopt the Mundlak (1978) to test whether we should use a random-effects or fixed-effects estimator. This is just to further confirm the results from the Hausman test.

The added benefit of the Mundlak approach, besides providing an alternative, is that I can use the robust estimator of the variance-covariance matrix.

In the model $Yit = X'it\beta + (\alpha i + \varepsilon it)$ the Mundlak approach determines if αi and εit are correlated, hence it suggests the specification: $\alpha i = X\bar{i}\,\theta + Ui$ $E(\alpha i\ |Xi\ ) = X\bar{i}\,\theta$

Here $X\bar{i}$ is the panel-level mean of $X'it$ and Ui is the time-invariant observable effect that is unrelated to the dependent variables. Hence, if I get θ=0, I can deduce that the covariates are unrelated which is what we are basically testing in the Wald Test. Incorporating the specification into the random effects model I get:

$$Yit = X'it\beta + X\bar{i}\,\theta + Ui + \varepsilon it \text{ ..... (1)}$$

$$E(Yit\ |Xit\ ) = X'it\beta + X\bar{i}\,\theta\text{..... (2)}$$

Here the first equation replaces αi with $X\bar{i}\,\theta + Ui$ and the second equation relies on assumption that the dependent variables and the unobservable heterogeneous impacts are mean independent. Thereby the approach preserve the specification of the random effects model but incorporates the correlation of the individual specific effects and the mean of the time varying variables. Hence we test $H0{:}\bar{\theta} = 0$ which indicates whether the fixed effect model or the random effect model should be used. If you reject that the coefficients are jointly zero, the test suggests that there is correlation between the time-invariant unobservables and your explanatory variables, meaning that the fixed-effects assumptions are satisfied. If you cannot reject the null that the generated explanatory variables are zero, there is evidence of no correlation between the time-invariant unobservable and your explanatory variables; that is, the random effects assumptions are satisfied. Results from this test is displayed in the table below:

Table 4.4: Results from Mundlak's Approach

```
. test fgdp fsim frlf frer

 ( 1)  fgdp = 0
 ( 2)  fsim = 0
 ( 3)  frlf = 0
 ( 4)  frer = 0

           chi2(  4) =12516.34
         Prob > chi2 =    0.0000
```

Since the p-value is less than 0.05, the test suggests that the time-invariant explanatory variables are indeed related to our explanatory variables and satisfies the fixed-effects assumptions. We reject the null hypothesis. This suggests that time-invariant unobservables are related to our explanatory variables and that the fixed-effects model is appropriate. Note that I used a robust estimator of the variance-covariance matrix. I could not do this using the Hausman test. This conforms to the result we obtained from the Hausman Test. (Find Appendix 6 for test for more variables)

4.5 Diagnostic tests for the fixed effect model

Since the results of the Hausman Test indicate that the Fixed Effect Model is the model we shall pursue in conducting our analysis, we can conduct some diagnostic tests to determine the strength of the Fixed Effects Model.

4.5.1 Modified Wald test for Groupwise Heteroscedasticity

Following Greene (2000), this test calculates a modified Wald Test statistics for group wise heteroscedasticity in the residuals of a fixed-effects regression model. The reason why residuals in panel data may deviate from homoscedasticity may be due to error variances specific to the cross-sectional unit. I undertake a test Modified Wald test for groupwise heteroscedasticity in fixed effect regression model under the following hypothesis:

Ho: sigma(i)^2 = sigma^2 for all i

Ha: not Ho

The results are displayed in the table below.

Table 4.5 Results from heteroscedasticity test

```
Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (42)  =      377.94
Prob>chi2 =       0.0000
```

Since the p-value is less than 5%, I reject the null and conclude that the residuals are heteroscedastic.

4.5.2 Pesaran Cross-Sectional Dependence Test for testing the Fixed Effect Model

A common assumption in panel-data models is that the error terms are independent across cross-sections. This test follows the methods shown in **Pesaran (2004)** and enables to detect whether there is serial correlation in the residual or not across entities. We can state the hypotheses as:

Ho: *There is no serial correlation*

Ha: *There exists serial correlation*

Below is the results from the test.

Pesaran's test of cross sectional independence = 70.674, Pr= 0.0000

Average absolute value of the off-diagonal elements = 0.431

The probability value is less than 5%, I reject the null hypothesis that this model contains no serial correlation. Hence, the model does contain serial correlation which means our results are less efficient, since serial correlation in panel data models biases the standard errors.

Even though the Hausman Test suggested that we use the fixed-effects model to conduct our panel-data analysis, we found that residuals from the fixed-effect model is heteroscedastic as well as serial correlation, making the results given by the model less efficient. However, the Mundlak approach, which can be used when errors are not homoscedastic and have serial correlation reaffirms that the fixed effects model should be chosen. Hence, we take into account the results from all the models and we observe the comparability of the results.

4.5.3 Breusch and Pagan Lagrangian multiplier test for random effects

Carrying out the test by the Random Effect Model (see Appendix 3), we find the probability value is less than 5%, which means the coefficients are not zero, hence the model is correctly specified. Additionally, in looking at the independent variables we can see that each of the variables are statistically significant in explaining the variation in trade.

Even though I have concluded on using estimates from the fixed effects model for my analysis, I go ahead to run a diagnostic test for the random effect model. I undertake a Breusch- Pagan test to see whether I should use random effect or pooled estimation.

Table 4.6 Estimates from the Breusch and Pagan Lagrangian multiplier test for random effects

```
Breusch and Pagan Lagrangian multiplier test for random effects

        trade[date,t] = Xb + u[date] + e[date,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                   trade |   3.607528       1.899349
                       e |   .3280544       .5727603
                       u |          0              0

        Test:   Var(u) = 0
                             chibar2(01) =     0.00
                          Prob > chibar2 =   1.0000
```

From the results above variance for u is 0 and the p value is 1 which means I cannot reject the null and therefore I conclude that the pooled regression will be better than the random effect model.

4.6 Conclusion

In this empirical project, I have used panel data from 91 country pairs who trade with each other and analyzed how the results could differ in different panel models. The data covers the period from 1960 to 2001. The results from my estimations indicate that the Random Effects Model are not very useful in reality because of the assumption that the unobserved individual heterogeneity is uncorrelated with the explanatory variables. This assumption is often violated. Both the Hausman test and the Mundlak approaches concludes that the Fixed-Effects model is the right way to go. Although diagnostics on the Fixed-Effects models indicate that residuals are heteroskedastic as well as the existence of issues of serial correlation, the Mundlak approach which takes these limitations into consideration yielded the same preference for the Fixed-Effects Model.

An extension of this project could be done in the future by analyzing Chang and Wall's approach. However, we must take notice of the fact that unless certain assumptions are met, their results can be biased. Given this limitation, their model may however be a better fit as compared to the random-effects model, since the Hausman test has indicated that the unobserved time invariant explanatory variables are indeed correlated with the time varying explanatory variables. In the interest of economic policy, it would be interesting to investigate the role of explanatory variables such as RLF and EMU in the gravity models of international trade over different time periods as well as the important role a factor like institutional quality could impact the inflow and returns from trade.

REFERENCES

Anderson, James E. (1979) “A Theoretical Foundation for the Gravity Equation”, American Economic Review, 69(1), pp. 106-116

Cheng I, Wall HJ. 2002. Controlling heterogeneity in gravity models of trade.Working Paper 1999-010C: Federal Reserve Bank of St. Louis.

Laura Serlenga and Yongcheol Shin (2007), Gravity Models of Intra-EU Trade: Application of the CCEP-HT Estimation in Heterogeneous, Journal of Applied Econometrics, Vol. 22, No. 2, pp. 361-381

Mundlak, Y. 1978: On the pooling of time series and cross section data, Econometrica 46:69-85.

Pesaran, M. H. (2004) “General diagnostic tests for cross section dependence in panels”, University of Cambridge, Faculty of Economics, Cambridge Working Papers in Economics No. 0435.

APPENDICES

APPENDIX 1 Summary Statistics (Between and Within Variation)

Variable		Mean	Std. Dev.	Min	Max	Observations	
id	overall	46	26.27129	1	91	N =	3822
	between		0	46	46	n =	42
	within		26.27129	1	91	T =	91
date	overall	7488	4427.75	0	14976	N =	3822
	between		4480.836	0	14976	n =	42
	within		0	7488	7488	T =	91
trade	overall	4.894827	1.899349	-1.036677	9.202658	N =	3822
	between		.8496973	3.272892	6.173741	n =	42
	within		1.703687	-.1863124	8.531353	T =	91
gdp	overall	13.4148	1.015532	10.48128	15.58845	N =	3822
	between		.3809567	12.63993	13.98507	n =	42
	within		.9431833	11.24849	15.10698	T =	91
sim	overall	-1.33562	.6574685	-3.586714	-.6931475	N =	3822
	between		.0408685	-1.392706	-1.24715	n =	42
	within		.656227	-3.543623	-.6360684	T =	91
rlf	overall	8.401505	1.23432	1.27593	10.20999	N =	3822
	between		.1649871	8.040432	8.642879	n =	42
	within		1.223506	1.276418	10.08854	T =	91
rer	overall	-.9609118	2.852824	-8.865199	5.623883	N =	3822
	between		.4003432	-1.535224	-.3322535	n =	42
	within		2.825262	-8.836762	5.599592	T =	91
cee	overall	.4358974	.4959387	0	1	N =	3822
	between		.3125141	.1098901	1	n =	42
	within		.38806	-.1684982	1.326007	T =	91
emu	overall	.0379383	.191072	0	1	N =	3822
	between		.1391537	0	.6043956	n =	42
	within		.1326681	-.5664574	.5434328	T =	91
dist	overall	7.102279	.6122743	5.153292	8.119398	N =	3822
	between		1.74e-15	7.102279	7.102279	n =	42
	within		.6122743	5.153292	8.119398	T =	91
bor	overall	.1538462	.3608484	0	1	N =	3822
	between		0	.1538462	.1538462	n =	42
	within		.3608484	0	1	T =	91
lan	overall	.1098901	.3127937	0	1	N =	3822
	between		1.40e-17	.1098901	.1098901	n =	42
	within		.3127937	0	1	T =	91
rert	overall	4.179426	.5647658	3.2835	4.8836	N =	3822
	between		.5715369	3.2835	4.8836	n =	42
	within		5.16e-15	4.179426	4.179426	T =	91
ftrade	overall	4.894824	.8396359	3.2729	6.1737	N =	3822
	between		.8497025	3.2729	6.1737	n =	42
	within		6.34e-15	4.894824	4.894824	T =	91
fgdp	overall	13.4148	.3764521	12.6399	13.9851	N =	3822
	between		.3809655	12.6399	13.9851	n =	42
	within		1.68e-14	13.4148	13.4148	T =	91
fsim	overall	-1.335614	.0403895	-1.3927	-1.2471	N =	3822
	between		.0408737	-1.3927	-1.2471	n =	42
	within		1.70e-15	-1.335614	-1.335614	T =	91
frlf	overall	8.401505	.1630376	8.0404	8.6429	N =	3822
	between		.1649923	8.0404	8.6429	n =	42
	within		1.04e-14	8.401505	8.401505	T =	91
frer	overall	-.9609095	.3955942	-1.5352	-.3323	N =	3822
	between		.400337	-1.5352	-.3323	n =	42
	within		1.35e-15	-.9609095	-.9609095	T =	91

APPENDIX 2 PEARSON CORRELATION CO-EFFICIENT

```
. //pearson rank correlation (APPENDIX 2)
. pwcorr trade gdp sim rlf rer cee emu dist bor lan rert
```

	trade	gdp	sim	rlf	rer	cee	emu
trade	1.0000						
gdp	0.8109	1.0000					
sim	-0.0643	-0.4477	1.0000				
rlf	-0.1851	-0.0760	-0.0674	1.0000			
rer	0.1879	-0.0688	0.0255	-0.1081	1.0000		
cee	0.5302	0.4314	-0.0274	0.0162	0.0554	1.0000	
emu	0.1489	0.1320	0.0017	0.0062	-0.0308	0.2259	1.0000
dist	-0.5731	-0.2612	-0.0408	0.3568	-0.3268	-0.2547	0.0017
bor	0.4534	0.2542	-0.0176	-0.3221	0.0451	0.2174	0.0406
lan	0.1111	-0.0707	0.0134	0.0019	0.4106	0.0910	-0.0435
rert	0.4014	0.3404	0.0505	0.1083	0.1253	0.4978	0.1167

	dist	bor	lan	rert
dist	1.0000			
bor	-0.5950	1.0000		
lan	-0.2704	0.0449	1.0000	
rert	0.0000	0.0000	-0.0000	1.0000

APPENDIX 3 RESULTS FROM THE FIVE REGRESSION ESTIMATIONS

```
. reg trade gdp sim rlf rer cee emu dist bor lan rert
```

Source	SS	df	MS
Model	12498.0819	10	1249.80819
Residual	1286.28181	3811	.337518187
Total	13784.3637	3821	3.60752779

Number of obs	= 3822
F(10, 3811)	= 3702.94
Prob > F	= 0.0000
R-squared	= 0.9067
Adj R-squared	= 0.9064
Root MSE	= .58096

trade	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gdp	1.543887	.0130767	118.06	0.000	1.518249	1.569525
sim	.8494035	.0171976	49.39	0.000	.815686	.883121
rlf	.0249879	.0084132	2.97	0.003	.008493	.0414828
rer	.0906828	.0039186	23.14	0.000	.0830001	.0983655
cee	.233327	.0244463	9.54	0.000	.1853978	.2812562
emu	.2137038	.0507735	4.21	0.000	.114158	.3132496
dist	-.6828109	.0226465	-30.15	0.000	-.7272113	-.6384105
bor	.5311423	.0337685	15.73	0.000	.4649363	.5973482
lan	.2487985	.0340738	7.30	0.000	.1819939	.3156031
rert	.1813633	.0209483	8.66	0.000	.1402924	.2224342
_cons	-10.93177	.2446889	-44.68	0.000	-11.4115	-10.45203

```
. reg trade gdp sim rlf rer cee emu dist bor lan rert, robust

Linear regression                                      Number of obs =    3822
                                                       F( 10,  3811) = 3366.38
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9067
                                                       Root MSE      =  .58096

------------------------------------------------------------------------------
             |               Robust
       trade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gdp |   1.543887   .0131458   117.44   0.000     1.518113     1.56966
         sim |   .8494035   .0167181    50.81   0.000     .8166262    .8821808
         rlf |   .0249879   .0079525     3.14   0.002     .0093963    .0405794
         rer |   .0906828   .0035705    25.40   0.000     .0836824    .0976831
         cee |    .233327    .022626    10.31   0.000     .1889668    .2776871
         emu |   .2137038   .0412096     5.19   0.000     .1329088    .2944988
        dist |  -.6828109   .0248209   -27.51   0.000    -.7314744   -.6341474
         bor |   .5311423   .0331714    16.01   0.000     .4661068    .5961777
         lan |   .2487985   .0333143     7.47   0.000     .1834829    .3141141
        rert |   .1813633   .0211936     8.56   0.000     .1398115    .2229152
       _cons |  -10.93177   .2584233   -42.30   0.000    -11.43843   -10.42511
------------------------------------------------------------------------------


. xtreg trade gdp sim rlf rer cee emu dist bor lan rert, re

Random-effects GLS regression                   Number of obs      =      3822
Group variable: date                            Number of groups   =        42

R-sq:  within  = 0.8882                         Obs per group: min =        91
       between = 0.9830                                        avg =      91.0
       overall = 0.9067                                        max =        91

                                                Wald chi2(10)      =  37029.36
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       trade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gdp |   1.543887   .0130767   118.06   0.000     1.518257    1.569516
         sim |   .8494035   .0171976    49.39   0.000     .8156968    .8831103
         rlf |   .0249879   .0084132     2.97   0.003     .0084982    .0414775
         rer |   .0906828   .0039186    23.14   0.000     .0830025     .098363
         cee |    .233327   .0244463     9.54   0.000      .185413    .2812409
         emu |   .2137038   .0507735     4.21   0.000     .1141897     .313218
        dist |  -.6828109   .0226465   -30.15   0.000    -.7271972   -.6384246
         bor |   .5311423   .0337685    15.73   0.000     .4649573    .5973272
         lan |   .2487985   .0340738     7.30   0.000     .1820151    .3155819
        rert |   .1813633   .0209483     8.66   0.000     .1403055    .2224212
       _cons |  -10.93177   .2446889   -44.68   0.000    -11.41135   -10.45219
-------------+----------------------------------------------------------------
     sigma_u |          0
     sigma_e |  .57276029
         rho |          0   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
. xtreg trade gdp sim rlf rer cee emu dist bor lan rert, be
note: dist omitted because of collinearity
note: bor omitted because of collinearity
note: lan omitted because of collinearity

Between regression (regression on group means)  Number of obs      =      3822
Group variable: date                            Number of groups   =        42

R-sq:  within  = 0.3633                         Obs per group: min =        91
       between = 0.9964                                        avg =      91.0
       overall = 0.4624                                        max =        91

                                                F(7,34)            =   1358.84
sd(u_i + avg(e_i.))=  .0556863                  Prob > F           =    0.0000

------------------------------------------------------------------------------
       trade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gdp |   1.485261   .1411022    10.53   0.000     1.198507    1.772015
         sim |   1.308857   1.580153     0.83   0.413      -1.9024    4.520114
         rlf |   .7585614   .1606698     4.72   0.000     .4320411    1.085082
         rer |   .4314732   .0584806     7.38   0.000     .3126263    .5503201
         cee |   .5312583   .1227943     4.33   0.000     .2817103    .7808064
         emu |  -.1291572   .1181197    -1.09   0.282    -.3692054     .110891
        dist |          0  (omitted)
         bor |          0  (omitted)
         lan |          0  (omitted)
        rert |  -.3157786   .0780031    -4.05   0.000    -.4743001   -.1572572
       _cons |  -18.14686   3.418517    -5.31   0.000    -25.09413    -11.1996
------------------------------------------------------------------------------

. xtreg trade gdp sim rlf rer cee emu dist bor lan rert, fe
note: rert omitted because of collinearity

Fixed-effects (within) regression               Number of obs      =      3822
Group variable: date                            Number of groups   =        42

R-sq:  within  = 0.8885                         Obs per group: min =        91
       between = 0.9841                                        avg =      91.0
       overall = 0.9030                                        max =        91

                                                F(9,3771)          =   3337.37
corr(u_i, Xb)  = 0.3126                         Prob > F           =    0.0000

------------------------------------------------------------------------------
       trade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gdp |   1.537816   .0130462   117.88   0.000     1.512238    1.563395
         sim |   .8386929   .0170965    49.06   0.000     .8051737    .8722121
         rlf |   .0205122   .0083251     2.46   0.014     .0041901    .0368343
         rer |   .0878235   .0038776    22.65   0.000      .080221     .095426
         cee |   .1669132   .0263631     6.33   0.000     .1152259    .2186005
         emu |    .210372   .0702206     3.00   0.003     .0726979    .3480461
        dist |  -.6975799   .0224123   -31.12   0.000    -.7415213   -.6536385
         bor |   .5358077   .0333704    16.06   0.000     .4703819    .6012336
         lan |   .2598762   .0336271     7.73   0.000     .1939471    .3258053
        rert |          0  (omitted)
       _cons |  -9.939758   .2598431   -38.25   0.000     -10.4492   -9.430311
-------------+----------------------------------------------------------------
     sigma_u |  .17190857
     sigma_e |  .57276029
         rho |  .08263978   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(41, 3771) =     5.54             Prob > F = 0.0000
```

APPENDIX 4 RESULTS FROM BASIC MODEL

```
. estimates tab Ols Ols_robust Random_eff Betw Fixed_eff,b(%7.2f)se(%7.2f)stats(N r2 r2_a rmse)
```

Variable	Ols	Ols_rob~t	Random_~f	Betw	Fixed_eff
gdp	1.33	1.33	1.24	2.20	1.18
	0.01	0.01	0.01	0.05	0.01
dist	-1.20	-1.20	-1.24	(omitted)	-1.27
	0.02	0.02	0.02		0.02
_cons	-4.37	-4.37	-2.90	-24.67	-1.87
	0.28	0.29	0.28	0.73	0.27
N	3822	3822	3822	3822	3822
r2	0.80	0.80		0.98	0.80
r2_a	0.80	0.80		0.98	0.80
rmse	0.85	0.85	0.80	0.13	0.77

legend: b/se

APPENDIX 5 ESTIMATING RANDOM EFFECTS BY MLE

```
Random-effects ML regression                    Number of obs      =      3822
Group variable: date                            Number of groups   =        42

Random effects u_i ~ Gaussian                   Obs per group: min =        91
                                                               avg =      91.0
                                                               max =        91

                                                LR chi2(9)         =   8428.59
Log likelihood  = -3330.6246                    Prob > chi2        =    0.0000
```

trade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
gdp	1.548023	.0130392	118.72	0.000	1.522466	1.573579
sim	.8503014	.0170996	49.73	0.000	.8167868	.883816
rlf	.0230638	.0083224	2.77	0.006	.0067522	.0393754
rer	.0903842	.0038841	23.27	0.000	.0827714	.097997
cee	.2049284	.0264906	7.74	0.000	.1530077	.2568491
emu	.2333633	.0652822	3.57	0.000	.1054126	.3613141
dist	-.684861	.022433	-30.53	0.000	-.7288288	-.6408931
bor	.5319779	.0333292	15.96	0.000	.4666539	.5973019
lan	.2543451	.033609	7.57	0.000	.1884726	.3202175
rer	0	(omitted)				
_cons	-10.18672	.2614237	-38.97	0.000	-10.69911	-9.674343
/sigma_u	.1439933	.0199835			.1097014	.1890048
/sigma_e	.5723507	.006589			.559581	.5854117
rho	.0595261	.0156322			.0345651	.0968071

Likelihood-ratio test of sigma_u=0: chibar2(01)=   97.33 Prob>=chibar2 = 0.000

APPENDIX 6 Mundlak's Approach with more variables

```
. test mean_gdp mean_sim mean_rlf mean_rer mean_cee mean_emu

 ( 1)  mean_gdp = 0
 ( 2)  mean_sim = 0
 ( 3)  mean_rlf = 0
 ( 4)  mean_rer = 0
 ( 5)  mean_cee = 0
 ( 6)  mean_emu = 0

           chi2(  6) = 3.0e+05
         Prob > chi2 =    0.0000
```