Sessions 1 & 2: Gravity and Panel Data

Ben Shepherd

Niehaus Center, Princeton University
GEM, Sciences Po

ARTNeT Capacity Building Workshop for Trade Research: “Behind the Border” Gravity Modeling

Wednesday, December 17, 2008
Introduction

Fixed Versus Random Effects

- Introduction
- Fixed Effects Estimation
- Random Effects Estimation
- Choosing Between Fixed and Random Effects
- Panel Data Dimensionality and the AvW Model

Additional Panel Data Issues

- Sample Splitting
- Aggregation and Clustering

Summary
The AvW Gravity Model–Aggregate Data

\[
\ln (X_{ij}) = \ln (Y_i) + \ln (E_j) - \ln (Y) + (1 - \sigma) \left[ \ln \tau_{ij} + \ln \Pi_i + \ln P_j \right]
\]

- One very common way of taking account of inward and outward multilateral resistance is to include dummy variables by exporter and by importer.
- The model can be estimated as usual by OLS, even if computations are sometimes cumbersome.
- In fact, we can analyze the model in more detail by treating it as an example of panel data: it is a pseudo-panel in which exporters index one dimension, and importers index the other.
The AvW Gravity Model–Aggregate Data

\[
\ln \left( X_{ij}^k \right) = \\
\ln \left( Y_i^k \right) + \ln \left( E_j^k \right) - \ln \left( Y^k \right) + (1 - \sigma_k) \left[ \ln \tau_{ij}^k + \ln \Pi_i^k + \ln \Pi_j^k \right]
\]

The sectoral gravity model is another example of panel data: it is a multi-dimensional pseudo panel, with dimensions indexed by exporters, importers and sectors.
In general, panel data refers to any situation in which we have a single dimension (the cross-section) observed a number of different times.

The most usual situation is to have the cross-sectional units as units of observation (individual, country), and for repeated observations to be drawn over time. But other configurations are possible, and that is typically what we are dealing with in gravity models.
Econometricians have developed special techniques for dealing with panel data, and taking advantages of the possibilities it offers.

In particular, there are simple ways of accounting for unobserved heterogeneity between cross-sectional units:

- Fixed effects.
- Random effects.

We will look at how to estimate these models, and then how to test for which one is more appropriate in a given setting.

Finally, we will deal with some practical issues that often arise in panel gravity model contexts.
Outline

1. Introduction

2. Fixed Versus Random Effects
   - Introduction
   - Fixed Effects Estimation
   - Random Effects Estimation
   - Choosing Between Fixed and Random Effects
   - Panel Data Dimensionality and the AvW Model

3. Additional Panel Data Issues
   - Sample Splitting
   - Aggregation and Clustering

4. Summary
Traditionally, panel data models have been used as a way of controlling for unobserved heterogeneity among cross sectional units.

Cross-sectional heterogeneity = There is something “different” about different units, but we may not be able to reduce these differences 100% to observable data.
Fixed and random effects are both aimed at controlling for unobserved heterogeneity, but take fundamentally different approaches: fixed effects allow for free or structureless variation, whereas random effects require that unobserved heterogeneity obey some probability constraints. Gravity theory strongly suggests a role for unobserved heterogeneity through the Multilateral Resistance terms in the AvW model.
Outline

1. Introduction
2. Fixed Versus Random Effects
   - Introduction
   - Fixed Effects Estimation
   - Random Effects Estimation
   - Choosing Between Fixed and Random Effects
   - Panel Data Dimensionality and the AvW Model
3. Additional Panel Data Issues
   - Sample Splitting
   - Aggregation and Clustering
4. Summary
Fixed Effects Estimation

“Fixed effects” are largely synonymous with categorical dummy variables covering one or more dimensions of a panel dataset.

In general terms, a panel data gravity model with importer and exporter fixed effects (like the one implied by the AvW model with aggregate data) can be written as follows:

\[ X_{ij} = A_i^1 + A_j^2 + A_{ij}^3 B^3 + E_{ij} \]

Trade flows, some covariates \((A_{ij}^3)\) and the error term are all defined at the country pair level. Distance and other bilateral factors would be part of \(A_{ij}^3\).

The matrix \(A_i^1\) contains the exporter fixed effects, and \(A_j^2\) contains the importer fixed effects.
Fixed Effects Estimation

Fixed Effects Gravity Model–Aggregate Data

\[ X_{ij} = A_i^1 + A_j^2 + A_{ij}^3 B^3 + E_{ij} \]

Note that \( A_{ij}^3 \) can only contain data that vary bilaterally, i.e. according to importer-exporter pairs.

Data that vary only in the importer dimension are collinear with the importer dummies in \( A_j^2 \).

Data that vary only in the exporter dimension are collinear with the exporter dummies in \( A_i^1 \).

Thus, we need to drop all variables from the model that vary only in the exporter or importer dimension: GDP, MFN tariffs, infrastructure development, general measures of regulation, etc.
Fixed Effects Estimation

Fixed Effects Gravity Model–Aggregate Data

\[ X_{ij} = A_i^1 + A_j^2 + A_{ij}^3 B^3 + E_{ij} \]

What can we do if we want to estimate a gravity model with some data that vary only by importer or exporter, not by country pair?
Transform the data to vary by pair: often we are really concerned about the combined impact of importer and exporter effects (e.g., infrastructure). At the price of an assumption as to relative weights, we can include these data by taking, for example, the average.

Look for alternative data that do vary by country pair.

Choose an alternative estimation methodology that does not have this characteristic: Baier-Bergstrand, AvW NLS, random effects, differencing.
To estimate the fixed effects model, we do not necessarily need to use any additional econometrics: we can just use OLS, as per the discussion earlier in this workshop.

The fixed effects are just dummy variables, and can be estimated as usual.

However, the number of fixed effects can become very large as we increase the number of countries. So it is useful to apply an econometric trick in order to eliminate one set of fixed effects.
Fixed Effects Estimation

Fixed Effects Gravity Model–Aggregate Data

\[ X_{ij} = A_i^1 + A_j^2 + A_{ij}^3 B^3 + E_{ij} \]

- Consider the mean of the above model, by exporter:

\[ \frac{1}{N} \sum_j X_{ij} = \bar{X}_{ij} = \bar{A}_i^1 + \bar{A}_j^2 B + \bar{A}_{ij}^3 B^3 + \bar{E}_{ij} \]

- Note that \( \bar{A}_i^1 = A_i^1 \), because the exporter fixed effect is constant for each exporter across all importers.

- Subtract the mean from the model to get the “within” estimator:

\[ X_{ij} - \bar{X}_{ij} = \left( A_j^2 - \bar{A}_j^2 \right) B^2 + \left( A_{ij}^3 - \bar{A}_{ij}^3 \right) B^3 + (E_{ij} - \bar{E}_{ij}) \]
Fixed Effects Estimation

- Note that the coefficients from the within estimator can still be given straightforward interpretations in terms of the underlying model:

\[ X_{ij} - \bar{X}_{ij} = \left( A_{j}^2 - \bar{A}_{j}^2 \right) + \left( A_{ij}^3 - \bar{A}_{ij}^3 \right) B^3 + (E_{ij} - \bar{E}_{ij}) \]

- The dimensionality of the problem is greatly reduced by the transformation: instead of having to estimate 2N fixed effects, we only have to estimate N of them.

- This is the transformation Stata applies when you use areg, or xtreg fe (to be discussed later). This allows it to work much more quickly, and with less memory, than through OLS with dummy variables.
Outline

1. Introduction
2. Fixed Versus Random Effects
   - Introduction
   - Fixed Effects Estimation
   - Random Effects Estimation
   - Choosing Between Fixed and Random Effects
   - Panel Data Dimensionality and the AvW Model
3. Additional Panel Data Issues
   - Sample Splitting
   - Aggregation and Clustering
4. Summary
Random Effects Estimation

- Multi-dimensional random effects models are actually quite complex, so to simplify I will consider random effects in the exporter dimension only:

\[ X_{ij} = A_1^i + A_{ij}^3B^3 + E_{ij} \]

- This time, we assume that \( A_1^i \) is a collection of random terms, satisfying similar assumptions to the error \( E_{ij} \).

- It can be shown that a GLS estimator of such a model is OLS performed on the following transformation:

\[ X_{ij} - t\bar{X}_i = \left( A_{ij}^3 - t\bar{A}_i \right) B^3 + (1 - t) A_1^i + (E_{ij} - t\bar{E}_i) \]

where \( t = 1 - \sqrt{\frac{\sigma_e^2}{N_j\sigma_A^2 + \sigma_e^2}} \).
Random Effects Estimation

- Again, coefficients from a random effects regression can be interpreted in exactly the same way as in a standard gravity model.

- It is quite feasible to include independent variables that vary in the same dimension as the random effects. Thus, a random effects gravity model with exporter and importer effects does not stop us from including GDP terms, or data on policies that vary by country only.
Random Effects Estimation

- The price to pay for this advantage is that random effects rely on a very strong assumption: that the pattern of unobserved heterogeneity is distributed randomly with given mean and variance.

- AvW theory leads us to expect unobserved heterogeneity, but does it suggest that this variation should necessarily fit well within a given distributional assumption?
Introduction

1. Introduction

2. Fixed Versus Random Effects
   - Introduction
   - Fixed Effects Estimation
   - Random Effects Estimation
   - Choosing Between Fixed and Random Effects
   - Panel Data Dimensionality and the AvW Model

3. Additional Panel Data Issues
   - Sample Splitting
   - Aggregation and Clustering

4. Summary

Ben Shepherd
Sessions 1 & 2: Gravity and Panel Data
Choosing Between Fixed and Random Effects

Advantages of fixed over random effects include:

- Less restrictive assumptions
- Fits nicely with AvW gravity theory
- Easy to estimate, even in multiple dimensions

Advantages of random over fixed effects include:

- Does not eliminate country-level variables
- No dimensionality constraint
The choice between these two estimators is not just one of convenience, however.

Random effects estimation imposes considerably more structure on the data (unobserved heterogeneity) than does fixed effects estimation.

Because of this, FE estimates are consistent regardless of whether the “true” model is FE or RE. If the true model is FE, they are also efficient; if it is RE, they are inefficient.

However, RE is inconsistent unless the “true” model is RE.
Choosing Between Fixed and Random Effects

- The Hausman test exploits the difference between coefficient and variance estimates in the two cases.
- Simplifying a little, H0 for the Hausman test is that the true model is random effects.
- A high test statistic (low prob. value) suggests that H0 is rejected. Most people would then conclude that a fixed effects model would be more appropriate.
- Note that the Hausman test can be unstable in practice, and usually cannot even be calculated with a robust variance-covariance estimator. Treat it as a rough indication only.
Outline

1 Introduction

2 Fixed Versus Random Effects
   - Introduction
   - Fixed Effects Estimation
   - Random Effects Estimation
   - Choosing Between Fixed and Random Effects
   - Panel Data Dimensionality and the AvW Model

3 Additional Panel Data Issues
   - Sample Splitting
   - Aggregation and Clustering

4 Summary
Panel Data Dimensionality and the AvW Model

- We have been concentrating on the aggregate version of the AvW model and its implications for panel data.
  - Aggregate export flows and bilateral data $\Rightarrow$ Fixed effects by exporter and importer
- Many other configurations of data are possible, however.
- To see what different data setups imply for estimation, we need to go back to the AvW model and think about the way in which the MR terms might change in particular dimensions.
Aggregate Data, Multiple Years

- The AvW model for a pure panel, i.e. aggregate bilateral trade data observed over a number of years, is:

\[
\ln(X_{ijt}) = \ln(Y_{it}) + \ln(E_{jt}) - \ln(Y_t) + (1 - \sigma) \left[ \ln(\tau_{ijt}) + \ln(\Pi_{it}) + \ln(P_{jt}) \right]
\]

- GDP and expenditure terms definitely change over time, as do trade flows.

- On the (reasonable) assumption that trade costs also change over time, then the MR terms must do likewise.

- Hence, we need fixed effects by importer-year, exporter-year, and year.
We’ve already seen the AvW model for many sectors observed only once in time:

\[
\ln \left( X_{ij}^k \right) = \ln \left( Y_i^k \right) + \ln \left( E_j^k \right) - \ln \left( Y^k \right) + (1 - \sigma_k) \left[ \ln \tau_{ij}^k + \ln \Pi_i^k + \ln P_j^k \right]
\]

If estimation is undertaken by pooling all observations into a single regression, we need fixed effects by importer-sector, exporter-sector, and sector.

We also need sector dummies interacted with trade costs.
Sectoral Data, Single Year

\[
\ln \left( X_{ij}^k \right) = \ln \left( Y_i^k \right) + \ln \left( E_j^k \right) - \ln \left( Y^k \right) + \\
(1 - \sigma_k) \left[ \ln \tau_{ij}^k + \ln \Pi_i^k + \ln P_j^k \right]
\]

- If estimation is undertaken separately using sector-by-sector regressions, each one needs fixed effects by importer and exporter.
An extension of the previous model to many time periods is:

$$\ln(X_{ijt}^k) = \ln(Y_{it}^k) + \ln(E_{jt}^k) - \ln(Y_t^k) + (1 - \sigma_k) \left[ \ln(\tau_{ijt}^k) + \ln(\Pi_{it}^k) + \ln(P_{jt}^k) \right]$$

Pooled estimation requires fixed effects by importer-sector-year, exporter-sector-year, and sector-year. Trade costs need to be interacted with sectoral dummies.

Estimation separately for each sector requires fixed effects by importer-year, exporter-year, and year.
Dealing with Long Time Series

- In the previous analysis, we implicitly assumed stability over time of:
  - The elasticity of substitution;
  - The parameters (elasticities) of the trade costs function.
- These should be reasonable assumptions over a short time period, but might be open to doubt over the longer term. Think of the “death of distance” hypothesis.
- One way to investigate this issue in long panels is by interacting trade costs with year dummies, in addition to the sector dummies. Alternatively, the model can be estimated separately for each year.
Outline

1. Introduction
2. Fixed Versus Random Effects
   - Introduction
   - Fixed Effects Estimation
   - Random Effects Estimation
   - Choosing Between Fixed and Random Effects
   - Panel Data Dimensionality and the AvW Model
3. Additional Panel Data Issues
   - Sample Splitting
   - Aggregation and Clustering
4. Summary
Sample Splitting

Although not strictly limited to panel data, the question of whether or not to split samples, and if so, how, is particularly important in this context.

For example, a panel data gravity model might group together data:

- For a number of disparate sectors (agriculture, manufacturing, even services);
- Across a number of highly disparate countries (low income vs. OECD; landlocked vs. others; etc.)

Splitting the sample can be a way of making these differences “talk”.

Ben Shepherd
Sessions 1 & 2: Gravity and Panel Data
### Sample Splitting

WMO (2005) - Splitting the Country Sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Full Sample</th>
<th>South-to-North Trade</th>
<th>South-to-South Trade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tariff Rates</td>
<td>-1.555***</td>
<td>-1.512</td>
<td>-1.5***</td>
</tr>
<tr>
<td>Port Efficiency of Importing Country</td>
<td>0.307*</td>
<td>0.344</td>
<td>-0.283</td>
</tr>
<tr>
<td>Port Efficiency of Exporting Country</td>
<td>0.924***</td>
<td>0.845***</td>
<td>0.949***</td>
</tr>
<tr>
<td>Customs Environment of Importing Country</td>
<td>0.472**</td>
<td>1.041</td>
<td>0.202</td>
</tr>
<tr>
<td>Regulatory Environment of Importing Country</td>
<td>0.281*</td>
<td>-1.120*</td>
<td>0.816***</td>
</tr>
<tr>
<td>Regulatory Environment of Exporting Country</td>
<td>0.620***</td>
<td>2.437***</td>
<td>0.827***</td>
</tr>
<tr>
<td>Service Sector Infrastructure of Importing Country</td>
<td>0.729***</td>
<td>2.134***</td>
<td>0.866</td>
</tr>
<tr>
<td>Service Sector Infrastructure of Exporting Country</td>
<td>1.943***</td>
<td>2.124***</td>
<td>3.133***</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.758</td>
<td>0.702</td>
<td>0.649</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>7,904</td>
<td>2,188</td>
<td>3,094</td>
</tr>
</tbody>
</table>
### Sample Splitting

Djankov et al. (2008) - Splitting the Country Sample

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Regional Agreement Sample</th>
<th>Regional Agreement and Income Group</th>
<th>Landlocked Country Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Ratio_time</td>
<td>-0.484 ***</td>
<td>-0.412 ***</td>
<td>-0.559</td>
</tr>
<tr>
<td></td>
<td>(-7.17)</td>
<td>(-5.34)</td>
<td>(-1.43)</td>
</tr>
<tr>
<td>Ratio_export time in neighbors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratio_GDP</td>
<td>1.146 ***</td>
<td>1.170 ***</td>
<td>1.134 ***</td>
</tr>
<tr>
<td></td>
<td>(41.38)</td>
<td>(43.09)</td>
<td>(33.92)</td>
</tr>
<tr>
<td>Ratio_GDPC</td>
<td>0.315 ***</td>
<td>0.118 *</td>
<td>0.446 ***</td>
</tr>
<tr>
<td></td>
<td>(5.82)</td>
<td>(1.81)</td>
<td>(2.93)</td>
</tr>
<tr>
<td>Distance</td>
<td>-1.272 ***</td>
<td>-1.255 ***</td>
<td>-1.296 ***</td>
</tr>
<tr>
<td></td>
<td>(-23.05)</td>
<td>(-22.06)</td>
<td>(-20.53)</td>
</tr>
<tr>
<td>Contiguity</td>
<td>0.533 ***</td>
<td>0.533 ***</td>
<td>0.471 ***</td>
</tr>
<tr>
<td></td>
<td>(6.41)</td>
<td>(6.40)</td>
<td>(4.59)</td>
</tr>
<tr>
<td>Language</td>
<td>0.720 ***</td>
<td>0.758 ***</td>
<td>0.670 ***</td>
</tr>
<tr>
<td></td>
<td>(8.84)</td>
<td>(9.13)</td>
<td>(8.42)</td>
</tr>
<tr>
<td>Colony</td>
<td>0.503 ***</td>
<td>0.566 ***</td>
<td>0.528 ***</td>
</tr>
<tr>
<td></td>
<td>(5.49)</td>
<td>(6.38)</td>
<td>(6.03)</td>
</tr>
<tr>
<td>Landlocked</td>
<td>-0.387 ***</td>
<td>-0.340 ***</td>
<td>-0.341 ***</td>
</tr>
<tr>
<td></td>
<td>(-4.14)</td>
<td>(-3.82)</td>
<td>(-2.83)</td>
</tr>
</tbody>
</table>
Sample Splitting
Djankov et al. (2008) - Splitting the Product Sample

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Manufacturing Products</th>
<th></th>
<th>Agricultural Products</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SITC 2-digit</td>
<td>HS 6-digit</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Ln(Time) * Ln(Time Sensitivity)</td>
<td>-0.260***</td>
<td>-0.148***</td>
<td>-0.430***</td>
<td>-0.366***</td>
</tr>
<tr>
<td></td>
<td>(-8.12)</td>
<td>(-3.20)</td>
<td>(-4.94)</td>
<td>(-3.42)</td>
</tr>
<tr>
<td>Ln(Time)^2(TimeSensitivityDummy)</td>
<td></td>
<td>-0.430***</td>
<td>-0.366***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-4.94)</td>
<td>(-3.42)</td>
<td></td>
</tr>
<tr>
<td>Ln(Skill Intensity) * Ln(Skill Abundance)</td>
<td>0.557***</td>
<td>0.371***</td>
<td>0.342***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.76)</td>
<td>(7.25)</td>
<td>(6.51)</td>
<td></td>
</tr>
<tr>
<td>Ln(Capital Intensity) * Ln(Capital Abundance)</td>
<td>0.152**</td>
<td>2.124</td>
<td>2.002</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.02)</td>
<td>(1.20)</td>
<td>(1.08)</td>
<td></td>
</tr>
<tr>
<td>Ln(Business_Entry) * Ln(TimeSensitivityDummy)</td>
<td></td>
<td>-0.028</td>
<td></td>
<td>-0.219</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.21)</td>
<td></td>
<td>(-1.29)</td>
</tr>
<tr>
<td>Ln(Labor_Regulation) * Ln(Time_SensitivityDummy)</td>
<td></td>
<td>-0.065</td>
<td></td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.81)</td>
<td></td>
<td>(1.20)</td>
</tr>
<tr>
<td>R^2</td>
<td>0.87</td>
<td>0.88</td>
<td>0.88</td>
<td>0.87</td>
</tr>
<tr>
<td>Number of Obs.</td>
<td>3276</td>
<td>2366</td>
<td>2366</td>
<td>2288</td>
</tr>
</tbody>
</table>
Outline

1. Introduction
2. Fixed Versus Random Effects
   - Introduction
   - Fixed Effects Estimation
   - Random Effects Estimation
   - Choosing Between Fixed and Random Effects
   - Panel Data Dimensionality and the AvW Model
3. Additional Panel Data Issues
   - Sample Splitting
   - Aggregation and Clustering
4. Summary
Particularly in panel data contexts, it is very common to use data at different levels of aggregation.

For example:

- Trade data that varies by importer-exporter-sector;
- Distance that varies by country pair;
- Infrastructure that varies by importer and by exporter;
- Effectively applied tariffs that vary by importer-exporter-sector.
Use of data at different levels of aggregation is very common, but can create problems in some cases. In particular, if errors tend to be correlated according to some cross-sectional unit (importers, exporters, etc), then an important OLS assumption is violated. Such “clustering” of errors is very common when different aggregation schemes are used, and can lead to significant underestimation of the true standard errors.
Aggregation and Clustering

- Stata contains a correction for clustering of the errors, but you need to tell it which variable to cluster on.
- To check that results are robust to clustering, experiment with different levels.
- For example, in an aggregate gravity model, check clustering by:
  - Exporter
  - Importer
  - Country pair
As a general rule, we should check that results are robust to clustering at the highest level of aggregation present in the data. So if we are using importer-level data with sectoral data, we should be sure to check for clustering by importer.

An alternative approach is to aggregate the data up to the highest level consistent with the data we are interested in.

For example, if we are interested in the effect of importer-level variables in an aggregate trade gravity model, we would aggregate the trade data so as to be worldwide exports rather than bilateral exports.
### Aggregation and Clustering

**WMO (2005) - Clustering Adjustments**

<table>
<thead>
<tr>
<th>Variable</th>
<th>OLS</th>
<th>WLS</th>
<th>WLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tariff Rates</td>
<td>-1.155***</td>
<td>-1.467***</td>
<td>-0.483**</td>
</tr>
<tr>
<td>(0.318)</td>
<td>(0.343)</td>
<td>(0.246)</td>
<td></td>
</tr>
<tr>
<td>Port Efficiency of Importer</td>
<td>0.307*</td>
<td>0.246</td>
<td>0.473***</td>
</tr>
<tr>
<td>(0.163)</td>
<td>(0.157)</td>
<td>(0.119)</td>
<td></td>
</tr>
<tr>
<td>Port Efficiency of Exporter</td>
<td>0.924****</td>
<td>0.913***</td>
<td>0.537***</td>
</tr>
<tr>
<td>(0.148)</td>
<td>(0.142)</td>
<td>(0.137)</td>
<td></td>
</tr>
<tr>
<td>Customs Environment of Importer</td>
<td>0.472**</td>
<td>0.472**</td>
<td>1.112***</td>
</tr>
<tr>
<td>(0.199)</td>
<td>(0.193)</td>
<td>(0.147)</td>
<td></td>
</tr>
<tr>
<td>Regulatory Environment of Importer</td>
<td>0.281*</td>
<td>0.288**</td>
<td>-0.069</td>
</tr>
<tr>
<td>(0.144)</td>
<td>(0.138)</td>
<td>(0.107)</td>
<td></td>
</tr>
<tr>
<td>Regulatory Environment of Exporter</td>
<td>0.620****</td>
<td>0.594***</td>
<td>0.180</td>
</tr>
<tr>
<td>(0.132)</td>
<td>(0.127)</td>
<td>(0.118)</td>
<td></td>
</tr>
<tr>
<td>Service Sector Infrastructure of Importer</td>
<td>0.729****</td>
<td>0.647***</td>
<td>0.494***</td>
</tr>
<tr>
<td>(0.224)</td>
<td>(0.227)</td>
<td>(0.166)</td>
<td></td>
</tr>
<tr>
<td>Service Sector Infrastructure of Exporter</td>
<td>1.943****</td>
<td>1.831***</td>
<td>2.336***</td>
</tr>
<tr>
<td>(0.216)</td>
<td>(0.208)</td>
<td>(0.189)</td>
<td></td>
</tr>
<tr>
<td>GNP of Importer</td>
<td>0.915***</td>
<td>0.931***</td>
<td>0.892***</td>
</tr>
<tr>
<td>(0.014)</td>
<td>(0.014)</td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>Per Capita GNP of Importer</td>
<td>-0.182***</td>
<td>-0.183***</td>
<td>-0.227***</td>
</tr>
<tr>
<td>(0.037)</td>
<td>(0.037)</td>
<td>(0.028)</td>
<td></td>
</tr>
<tr>
<td>GNP of Exporter</td>
<td>1.246****</td>
<td>1.239***</td>
<td>1.169***</td>
</tr>
<tr>
<td>(0.014)</td>
<td>(0.014)</td>
<td>(0.012)</td>
<td></td>
</tr>
<tr>
<td>Per Capita GNP of Exporter</td>
<td>-0.226***</td>
<td>-0.231***</td>
<td>-0.153***</td>
</tr>
<tr>
<td>(0.029)</td>
<td>(0.028)</td>
<td>(0.022)</td>
<td></td>
</tr>
<tr>
<td>Geographical Distance</td>
<td>-1.258***</td>
<td>-1.238***</td>
<td>-1.143***</td>
</tr>
<tr>
<td>(0.025)</td>
<td>(0.025)</td>
<td>(0.018)</td>
<td></td>
</tr>
</tbody>
</table>

| Robust Standard Error                   | Cluster No | Importer Yes | Exporter Yes |
| Weighted Least Squares                  | Cluster     | Importer     | Exporter     |
### Aggregation and Clustering

Djankov et al. (2008) - Aggregation

#### Aggregate Trade Data to the World

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Regional Agreement Sample</th>
<th>Regional Agreement &amp; Income Group</th>
<th>Landlocked Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Ratio_Time</td>
<td>-0.307 ***</td>
<td>-0.362 ***</td>
<td>-0.749 **</td>
</tr>
<tr>
<td></td>
<td>(-3.26)</td>
<td>(-2.97)</td>
<td>(-2.06)</td>
</tr>
<tr>
<td>Ratio_export time in neighbors</td>
<td></td>
<td></td>
<td>-0.588*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-1.84)</td>
</tr>
<tr>
<td>Ratio_GDP</td>
<td>0.933 ***</td>
<td>0.943 ***</td>
<td>1.124 ***</td>
</tr>
<tr>
<td></td>
<td>(26.20)</td>
<td>(26.51)</td>
<td>(5.24)</td>
</tr>
<tr>
<td>Ratio_GDPC</td>
<td>0.363 ***</td>
<td>0.257***</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td>(7.54)</td>
<td>(4.30)</td>
<td>(0.55)</td>
</tr>
<tr>
<td>Landlocked</td>
<td>0.024</td>
<td>0.070</td>
<td>0.089</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.50)</td>
<td>(0.53)</td>
</tr>
<tr>
<td>Instrument: Transit time in neighbors</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>R²</td>
<td>0.82</td>
<td>0.83</td>
<td>0.79</td>
</tr>
<tr>
<td>No of Obs</td>
<td>333</td>
<td>333</td>
<td>220</td>
</tr>
</tbody>
</table>

Sessions 1 & 2: Gravity and Panel Data
Panel data techniques provide a simple and effective way for analyzing the gravity model in terms of unobserved heterogeneity, due theoretically to the MR terms.

The fixed effects estimator is much more common in gravity models than is the random effects estimator. RE imposes much tighter conditions on the data, and experience suggests they are usually rejected in practice.

However, RE has the advantage of not requiring exclusion of data varying in the same dimensions as the panel effects.
In FE gravity models, it takes considerable care to get the right configuration of fixed effects in light of theory.

- See Baldwin & Taglioni (2007) for a critical overview of past papers—including some very well known ones—that did not get this right.

- Aggregation and clustering are also important issues in panel data contexts. They can have a significant effect on final results.
As a robustness check, it is important to be sure that results continue to hold:

- When different patterns of error clustering are allowed for; and
- With data aggregated to the highest level of any indicator of interest.