Hierarchical modeling of longitudinal trade networks

Peter Hoff

Statistics, Biostatistics and the CSSS
University of Washington
Longitudinal trade data

- GDP for 1990-2000 (WDI online)

Let \( y_{i,j,t} = \log \text{trade between countries } i \text{ and } j \text{ at time } t \):

\[
y_{i,j,t} = \beta_1 + \beta_2 y_{i,j,t-1} + \beta_3 x_{i,t} + \beta_4 x_{j,t} + \sigma \epsilon_{i,j,t}
\]

We’ll consider modeling each cell as
USA Exports to top 20 importers
China Exports to top 20 importers
Denmark Exports to top 20 importers

- Log exports from DEN
- Residuals

Graphs show changes in log GDP and lagged log exports from DEN, with residual plots indicating the model's performance.
The gravity model of trade

By analogy with Newton’s law of gravity,

\[ \text{trade}_{i,j} \approx c_0 \times \frac{(\text{gdp}_i^{c_1} \cdot \text{gdp}_j^{c_2})}{\text{dist}_{i,j}^{c_3}}. \]
The gravity model of trade

By analogy with Newton’s law of gravity,

\[ \text{trade}_{i,j} \approx c_0 \times \frac{(\text{gdp}_i^{c_1} \text{gdp}_j^{c_2})}{\text{dist}_{i,j}^{c_3}}. \]

Trade between countries

- increases with the economic “mass” of the two countries;
- decreases in trade costs or “distance”.
The gravity model of trade

By analogy with Newton’s law of gravity,

\[ \text{trade}_{i,j} \approx c_0 \times \frac{(gdp_i^{c_1} gdp_j^{c_2})}{\text{dist}_{i,j}^{c_3}}. \]

Trade between countries

- increases with the economic “mass” of the two countries;
- decreases in trade costs or “distance”.

Taking logs,

\[ \text{trade}_{i,j} = \beta_{\text{int}} + \beta_{gdp} \text{gdp}_i + \beta_{gdp} \text{gdp}_j + \beta_{\text{dist}} \text{dist}_{i,j} + \sigma \epsilon_{i,j} \]

where now all variables are on the log scale.
Generalizing to a longitudinal model

Using g20 countries only, pooled across years:

<table>
<thead>
<tr>
<th></th>
<th>exporter gdp</th>
<th>importer gdp</th>
<th>distance</th>
<th>error sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>estimate</td>
<td>1.14</td>
<td>1.07</td>
<td>-0.89</td>
<td>1.31</td>
</tr>
<tr>
<td>std error</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td></td>
</tr>
</tbody>
</table>
Generalizing to a longitudinal model

Using g20 countries only, pooled across years:

<table>
<thead>
<tr>
<th></th>
<th>exporter gdp</th>
<th>importer gdp</th>
<th>distance</th>
<th>error sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>estimate</td>
<td>1.14</td>
<td>1.07</td>
<td>-0.89</td>
<td>1.31</td>
</tr>
<tr>
<td>std error</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td></td>
</tr>
</tbody>
</table>

Implications for changes across years:

If the parameters in the gravity model are constant across countries and years,
Generalizing to a longitudinal model

Using G20 countries only, pooled across years:

<table>
<thead>
<tr>
<th></th>
<th>exporter gdp</th>
<th>importer gdp</th>
<th>distance</th>
<th>error sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>estimate</td>
<td>1.14</td>
<td>1.07</td>
<td>-0.89</td>
<td>1.31</td>
</tr>
<tr>
<td>std error</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td></td>
</tr>
</tbody>
</table>

Implications for changes across years:

If the parameters in the gravity model are constant across countries and years,

\[
\begin{align*}
(trade_{i,j,t} - trade_{i,j,t-1}) &= \beta_{gdpe}(gdp_{i,t} - gdp_{i,t-1}) + \beta_{gdpi}(gdp_{j,t} - gdp_{j,t-1}) + \epsilon_{i,j,t} \\
y_{i,j,t} &= 0 + y_{i,j,t-1} + \beta_3 x_{i,t} + \beta_4 x_{j,t} + \epsilon_{i,j,t},
\end{align*}
\]

where \( x_{i,t} \) is the change in log gdp of country \( i \) from year \( t-1 \) to \( t \).
Does the static gravity model fit generalize over time? Consider the model

$$y_{i,j,t} = \beta_1 + \beta_2 y_{i,j,t-1} + \beta_3 x_{i,t} + \beta_4 x_{j,t} + \sigma \epsilon_{i,j}.$$ 

This is a first-order autoregressive (AR1) model, which contains the lagged gravity model as a submodel ($\beta_1 = 0$ and $\beta_2 = 1$).
Gravity model

Does the static gravity model fit generalize over time? Consider the model

\[ y_{i,j,t} = \beta_1 + \beta_2 y_{i,j,t-1} + \beta_3 x_{i,t} + \beta_4 x_{j,t} + \sigma \epsilon_{i,j}. \]

This is a first-order autoregressive (AR1) model, which contains the lagged gravity model as a submodel \((\beta_1 = 0 \text{ and } \beta_2 = 1)\).

<table>
<thead>
<tr>
<th></th>
<th>intercept</th>
<th>lag trade</th>
<th>exporter gdp</th>
<th>importer gdp</th>
<th>error sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>estimate</td>
<td>0.54</td>
<td>0.93</td>
<td>0.51</td>
<td>2.60</td>
<td>0.47</td>
</tr>
<tr>
<td>std error</td>
<td>0.03</td>
<td>0.004</td>
<td>0.194</td>
<td>0.194</td>
<td></td>
</tr>
</tbody>
</table>
Within and between variability

Compare:

<table>
<thead>
<tr>
<th></th>
<th>intercept</th>
<th>lag trade</th>
<th>exporter gdp</th>
<th>importer gdp</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1</td>
<td>0.54</td>
<td>0.93</td>
<td>0.51</td>
<td>2.60</td>
</tr>
<tr>
<td>lagged gravity</td>
<td>0.00</td>
<td>1.00</td>
<td>1.14</td>
<td>1.07</td>
</tr>
</tbody>
</table>
Within and between variability

Compare:

<table>
<thead>
<tr>
<th></th>
<th>intercept</th>
<th>lag trade</th>
<th>exporter gdp</th>
<th>importer gdp</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1</td>
<td>0.54</td>
<td>0.93</td>
<td>0.51</td>
<td>2.60</td>
</tr>
<tr>
<td>lagged gravity</td>
<td>0.00</td>
<td>1.00</td>
<td>1.14</td>
<td>1.07</td>
</tr>
</tbody>
</table>

\(\beta_{egdp}, \beta_{igdp}\) represent

- across-pair covariance of trade and gdp in the static model;
- within-pair, across year covariance of trade and gdp in the AR1 model.

Both of these approaches ignore across-pair heterogeneity in trade patterns.
Modeling decision:
We would like to obtain a trade model for each pair \(\{i,j\}\), but we don’t have much pair-specific data. Should we

- fit a single AR1 model, using data from all pairs?
- fit a separate model for each pair?
- assume some similarity of regression coefficients across pairs?
The model comparison game

\[ Y_t = B_1 + B_2 Y_{t-1} + B_3 X_{\text{exp},t} + B_4 X_{\text{imp},t} + E_t \]

where \( \beta_{i,j}^T = (B_{1,[i,j]}, B_{2,[i,j]}, B_{3,[i,j]}, B_{4,[i,j]}) \)
The model comparison game

\[ Y_t = B_1 + B_2 Y_{t-1} + B_3 X_{exp,t} + B_4 X_{imp,t} + E_t \]

where \( \beta_{i,j}^T = (B_{1,[i,j]}, B_{2,[i,j]}, B_{3,[i,j]}, B_{4,[i,j]}) \)

The game:

- estimate \( \beta_{i,j} \) for each \( \{i,j\} \) using data \( \{Y_t, X_t\} \) for \( t \in \{1991, \ldots, 1999\} \).
- Make predictions for 2000 and evaluate:

\[ \hat{Y}_{2000} = \hat{B}_1 + \hat{B}_2 Y_{1999} + \hat{B}_3 X_{exp,2000} + \hat{B}_4 X_{imp,2000} \]

\[ \text{MSE} = ||Y_{2000} - \hat{Y}_{2000}|| \]
The model comparison game

\[ Y_t = B_1 + B_2 Y_{t-1} + B_3 X_{\text{exp},t} + B_4 X_{\text{imp},t} + E_t \]

where \( \beta_{i,j}^T = (B_{1,[i,j]}, B_{2,[i,j]}, B_{3,[i,j]}, B_{4,[i,j]}) \)

The game:
- estimate \( \beta_{i,j} \) for each \( \{i,j\} \) using data \( \{Y_t, X_t\} \) for \( t \in \{1991, \ldots, 1999\} \).
- Make predictions for 2000 and evaluate:
  \[
  \hat{Y}_{2000} = \hat{B}_1 + \hat{B}_2 Y_{1999} + \hat{B}_3 X_{\text{exp},2000} + \hat{B}_4 X_{\text{imp},2000} \\
  \text{MSE} = ||Y_{2000} - \hat{Y}_{2000}||
  \]

We will do this for a variety of estimation strategies:
- separate fits for each pair
- common fit for all pairs
- a pair-exchangeable hierarchical model for the \( \beta_{i,j} \)'s
- a node-exchangeable hierarchical model for the \( \beta_{i,j} \)'s
Separate and common fits

To accommodate pairs having zero trade in some years, we use a tobit model:

\[
y_{i,j,t}^* = \beta_1 + \beta_2 y_{i,j,t-1} + \beta_3 x_{i,t} + \beta_4 x_{j,t} + \epsilon_{i,j,t}
\]

\[
= \beta^T x_{i,j,t} + \epsilon_{i,j,t}
\]

\[
y_{i,j,t} = y_{i,j,t}^* \lor 0
\]

Bayesian estimation is done using a Gibbs sampler, with prior distributions that are weakly centered around 1989-1990 estimates.

- independent fit procedure: \{\beta_{i,j}, \sigma_{i,j}\} estimated separately for each pair;
- common fit procedure: \{\beta_{i,j}, \sigma_{i,j}\} = \{\beta, \sigma\} for all pairs \{i, j\}
Separate fit estimation
Common fit estimation
Comparison in terms of prediction

For each estimation procedure, we calculate

\[
\frac{\sum_{i,j} (y_{i,j,2000} - \hat{y}_{i,j,2000})^2}{\sum_{i,j} (y_{i,j,2000} - y_{i,j,1999})^2}
\]
Comparison in terms of prediction

For each estimation procedure, we calculate

\[
\frac{\sum_{\{i,j\}} (y_{i,j,2000} - \hat{y}_{i,j,2000})^2}{\sum_{\{i,j\}} (y_{i,j,2000} - y_{i,j,1999})^2}
\]

<table>
<thead>
<tr>
<th>estimation</th>
<th>all pairs</th>
<th>g20 pairs</th>
<th>g20 exporter</th>
<th>g20 importer</th>
<th>non g20 pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>separate fits</td>
<td>1.16</td>
<td>1.06</td>
<td>1.14</td>
<td>1.11</td>
<td>1.18</td>
</tr>
<tr>
<td>common fit</td>
<td>1.00</td>
<td>1.11</td>
<td>0.99</td>
<td>0.99</td>
<td>1.01</td>
</tr>
</tbody>
</table>
Comparison in terms of prediction

For each estimation procedure, we calculate

$$\frac{\sum_{i,j} (y_{i,j,2000} - \hat{y}_{i,j,2000})^2}{\sum_{i,j} (y_{i,j,2000} - y_{i,j,1999})^2}$$

<table>
<thead>
<tr>
<th>estimation</th>
<th>all pairs</th>
<th>g20 pairs</th>
<th>g20 exporter</th>
<th>g20 importer</th>
<th>non g20 pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>separate fits</td>
<td>1.16</td>
<td>1.06</td>
<td>1.14</td>
<td>1.11</td>
<td>1.18</td>
</tr>
<tr>
<td>common fit</td>
<td>1.00</td>
<td>1.11</td>
<td>0.99</td>
<td>0.99</td>
<td>1.01</td>
</tr>
</tbody>
</table>

A good statistical model should

- represent patterns in the observed data;
- generalize to other datasets.

The “separate fits” procedure does the first, fails the second.

The “common fit” procedure fails the first, “doesn’t fail” the second.
Pair-exchangeable model

\[ \beta_{1,2}, \ldots, \beta_{136,135} \overset{iid}{\sim} \text{mvn}(\beta_0, \Sigma) \]

\[ y_{i,j,t} = \beta_{i,j}^T x_{i,j,t} + \sigma_{i,j} \epsilon_{i,j,t} \]

- \( \beta_0 \) represents the average regression coefficient;
- \( \Sigma \) represents the across-pair heterogeneity in regression coefficients.
### Pair-exchangeable model

\[
\beta_{1,2}, \ldots, \beta_{136,135} \overset{iid}{\sim} \text{mvn}(\beta_0, \Sigma) \\
y_{i,j,t} = \beta_{i,j}^T x_{i,j,t} + \sigma_{i,j} \epsilon_{i,j,t}
\]

- \(\beta_0\) represents the average regression coefficient;
- \(\Sigma\) represents the across-pair heterogeneity in regression coefficients.

**Regularization:**
If \((\beta_0, \Sigma)\) were given, then our uncertainty about \(\beta_{i,j}\) would be given by

\[
\begin{align*}
\text{V[} \beta_{i,j} | Y, X, \beta_0, \Sigma \text{]} &= [\Sigma^{-1} + X_{i,j}^T X_{i,j} / \sigma_{i,j}^2]^{-1} \\
\text{E[} \beta_{i,j} | Y, X, \beta_0, \Sigma \text{]} &= [\Sigma^{-1} + X_{i,j}^T X_{i,j} / \sigma_{i,j}^2]^{-1} [\Sigma^{-1} \beta_0 + X_{i,j}^T y_{i,j} / \sigma_{i,j}^2]
\end{align*}
\]
Pair-exchangeable model

\[ \beta_{1,2}, \ldots, \beta_{136,135} \overset{\text{iid}}{\sim} \text{mvn}(\beta_0, \Sigma) \]
\[ y_{i,j,t} = \beta_{i,j}^T x_{i,j,t} + \sigma_{i,j} \epsilon_{i,j,t} \]

- \( \beta_0 \) represents the average regression coefficient;
- \( \Sigma \) represents the across-pair heterogeneity in regression coefficients.

**Regularization:**
If \((\beta_0, \Sigma)\) were given, then our uncertainty about \(\beta_{i,j}\) would be given by

\[ V[\beta_{i,j}|Y, X, \beta_0, \Sigma] = [\Sigma^{-1} + X_{i,j}^T X_{i,j} / \sigma_{i,j}^2]^{-1} \]
\[ E[\beta_{i,j}|Y, X, \beta_0, \Sigma] = [\Sigma^{-1} + X_{i,j}^T X_{i,j} / \sigma_{i,j}^2]^{-1}[\Sigma^{-1} \beta_0 + X_{i,j}^T y_{i,j} / \sigma_{i,j}^2] \]

\( \hat{\beta}_{i,j} \) is pulled towards \( \beta_0 \) by an amount determined by \( \sigma_{i,j}^2 \) and \( \Sigma \):
- pairs with high \( \sigma_{i,j}^2 \) are shrunk more;
- pairs with low \( \sigma_{i,j}^2 \) are shrunk less.
Exchangeable model
## Predictive comparison

<table>
<thead>
<tr>
<th>estimation</th>
<th>all pairs</th>
<th>g20 pairs</th>
<th>g20 exporter</th>
<th>g20 importer</th>
<th>non g20 pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>separate fits</td>
<td>1.16</td>
<td>1.06</td>
<td>1.14</td>
<td>1.11</td>
<td>1.18</td>
</tr>
<tr>
<td>common fit</td>
<td>1.00</td>
<td>1.11</td>
<td>0.99</td>
<td>0.99</td>
<td>1.01</td>
</tr>
<tr>
<td>exchangeable fit</td>
<td>1.00</td>
<td>0.80</td>
<td>0.98</td>
<td>0.99</td>
<td>1.01</td>
</tr>
</tbody>
</table>

We seem to be on the right track.
Exchangeability

Does pair-exchangeability make sense? Exchangeability implies

\[ \{\beta_{i,j}\} \stackrel{d}{=} \{\beta_{\pi(i,j)}\} \]
Exchangeability

Does pair-exchangeability make sense? Exchangeability implies

- \( \{\beta_{i,j}\} \overset{d}{=} \{\beta_{\pi(i,j)}\} \)
- \( \text{Cov}(\beta_{USA,CAN}, \beta_{USA,MEX}) = \text{Cov}(\beta_{SIN,ICE}, \beta_{URU,ETH}) \)
Does pair-exchangeability make sense? Exchangeability implies

- \[ \{ \beta_{i,j} \} \overset{d}{=} \{ \beta_{\pi(i,j)} \} \]
- \[ \text{Cov}(\beta_{\text{USA,CAN}}, \beta_{\text{USA,MEX}}) = \text{Cov}(\beta_{\text{SIN,ICE}}, \beta_{\text{URU,ETH}}) \]
- All \( \beta_{i,j} \)'s get shrunk towards the same value.
Exchangeability

Does pair-exchangeability make sense? Exchangeability implies

- \( \{\beta_{i,j}\} \overset{d}{=} \{\beta_{\pi(i,j)}\} \)
- \( \text{Cov}(\beta_{\text{USA,CAN}}, \beta_{\text{USA,MEX}}) = \text{Cov}(\beta_{\text{SIN,ICE}}, \beta_{\text{URU,ETH}}) \)
- all \( \beta_{i,j} \)'s get shrunk towards the same value.

Lack of fit of exchangeable model: Consider the \( n \times (n - 1) \) different \( \beta_{i,j,1} \)'s

\[
B_1 = \begin{pmatrix}
\beta_{3,1} & \beta_{2,1} & \beta_{1,3} & \cdots \\
\beta_{2,1} & \beta_{3,1} & \beta_{1,2} & \cdots \\
\beta_{1,1} & \beta_{2,2} & \beta_{1,3} & \cdots \\
\cdots & \cdots & \cdots & \cdots
\end{pmatrix}
\]

Exchangeable model:

\[
B_1 = \beta_{0,1} \mathbf{1}\mathbf{1}^T + \mathbf{E}, \text{ where } \mathbf{E} \text{ is patternless noise}
\]

Empirical evaluation:

\[
\hat{B}_1 = \hat{\beta}_{0,1} \mathbf{1}\mathbf{1}^T + \hat{\mathbf{E}}, \text{ look for patterns in } \hat{\mathbf{E}}.
\]
Singular value decomposition

Any matrix $M$ can be decomposed as follows:

$$M = UDV^T$$

where

- $U$ represents variability among the rows of $M$;
- $V$ represents variability among the columns of $M$;
- $D$ is a diagonal matrix of weights.

In particular, $m_{i,j} = u_i^T D v_j$. 
Singular value decomposition

Any matrix $M$ can be decomposed as follows:

$$M = UDV^T$$

where

- $U$ represents variability among the rows of $M$;
- $V$ represents variability among the columns of $M$;
- $D$ is a diagonal matrix of weights.

In particular, $m_{i,j} = u_i^T D v_j$.

In terms of residual deviations from the mean coefficient,

$$\hat{E} = UDV^T$$

$$\hat{e}_{i,j} = u_i^T D v_j$$

Think of $\{u_i, v_i\}$ as latent country-specific characteristics: The residual $\hat{e}_{i,j}$ is positive if $u_i, v_j$ in same direction.
Decompositions of coefficient matrices
Node exchangeable effects

\[ y_{i,j,t} = \beta_{i,j}^T x_{i,j,t} + \epsilon_{i,j,t} \]

\[ \beta_{i,j} = (\beta_{i,j,1}, \beta_{i,j,2}, \beta_{i,j,3}, \beta_{i,j,4})^T \]

\[ \begin{align*}
\beta_{i,j,1} &= b_1 + u_i^T D_1 v_j + \gamma_{i,j,1} \\
\beta_{i,j,2} &= b_2 + u_i^T D_2 v_j + \gamma_{i,j,2} \\
\beta_{i,j,3} &= b_3 + u_i^T D_3 v_j + \gamma_{i,j,3} \\
\beta_{i,j,4} &= b_4 + u_i^T D_4 v_j + \gamma_{i,j,4}
\end{align*} \]

Thinking of \( \{u_i, v_j\} \) as random effects,

\[ \{\beta_{i,j}\} \overset{d}{\neq} \{\beta_{\pi(i,j)}\} \]

\[ \{\beta_{i,j}\} \overset{d}{=} \{\beta_{\pi_1 i, \pi_2 j}\} \]

- Allows shrinkage to depend on the exporter and importer.
- More general than additive effects, i.e. \((u_i + v_j)d_k\)
Country exchangeable model
Country exchangeable model

<table>
<thead>
<tr>
<th>dimension</th>
<th>intercept</th>
<th>lag trade</th>
<th>gdp exporter</th>
<th>gdp importer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.95</td>
<td>8.75</td>
<td>0.13</td>
<td>0.73</td>
</tr>
<tr>
<td>2</td>
<td>1.19</td>
<td>12.74</td>
<td>0.07</td>
<td>0.36</td>
</tr>
</tbody>
</table>
### Predictive comparison

<table>
<thead>
<tr>
<th>estimation</th>
<th>all pairs</th>
<th>g20 pairs</th>
<th>g20 exporter</th>
<th>g20 importer</th>
<th>non g20 pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>separate fits</td>
<td>1.15</td>
<td>1.06</td>
<td>1.14</td>
<td>1.11</td>
<td>1.18</td>
</tr>
<tr>
<td>common fit</td>
<td>1.00</td>
<td>1.11</td>
<td>0.99</td>
<td>0.99</td>
<td>1.01</td>
</tr>
<tr>
<td>exchangeable fit</td>
<td>1.00</td>
<td>0.80</td>
<td>0.98</td>
<td>0.99</td>
<td>1.01</td>
</tr>
<tr>
<td>factor model</td>
<td>0.97</td>
<td>0.77</td>
<td>0.95</td>
<td>0.96</td>
<td>0.98</td>
</tr>
</tbody>
</table>
Discussion

Shrinkage can greatly improve out-of-sample model validity
- not enough data for a separate model for each pair;
- ignoring across-pair differences is too extreme;
- shrinkage without regard to exporter/importer identity is too simplistic.

Shrinkage should be done with the proper exchangeability structure
- country-specific effects provide realistic shrinkage, improved predictions.
Future work

What about the variance?
Thanks

- NSF
- Mike Ward, UW Polysci
- Xiaoyue Niu, UW Stat
- John Ahlquist, UCLA Polysci