Estimating bilateral relationships from aggregated data

Harald Badinger
WU, WIFO
Jesús Crespo Cuaresma
WU, IIASA, WIC, WIFO

wiiw - International Seminar
Vienna, December 2011

A more pretentious title unveiling the pseudo-alchemist nature of the paper

De aestimo aequationes bilateralis a data aggregata

Harald Badinger
WU, WIFO
Jesús Crespo Cuaresma
WU, IIASA, WIC, WIFO

How it all began ...

- We are interested in contagion effects of financial instability (measured by some variable z) and aim at estimating a model such as, for example,
$$z = \mathbf{W}z + \mathbf{X}_z\beta_z + \varepsilon$$
- We want to create spatial weighting matrices which are related to the (exogenously given) financial linkages between units, which may be related to geographical distance, but also to other exogenous variables
- In order to unveil the nature of such exogenous drivers of financial linkages, we estimate models such as
$$y = \mathbf{X}\beta + u,$$
where y measures, for example, bilateral portfolio flows

How it all began ...

- Big problem: bilateral data on financial linkages are not existing, only aggregated data are available
- Big solution: Estimating bilateral data from (nonlinearly) aggregated variables
- The method proposed can be used to:
  - Perform inference on bilateral data when only aggregated data are available
  - Create (time-varying) weight matrices for spatial models which go beyond geographical distance
  - Specify models with spillover effects for phenomena for which data on linkages are not available

An example: bilateral trade relationships

- The bilateral gravity model

$$\log T_{ij} = \mathbf{X}_{ij}\beta + u_{ij}$$

- The observed data

$$T_i = \sum_j \exp(\log T_{ij})$$

- The model on aggregated/bilateral data

$$T_i = \sum_j \exp(\log T_{ij}) = \sum_j \exp(\mathbf{X}_{ij}\beta + u_{ij})$$

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

The general setting

- The bilateral model

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$$

- The nonlinear aggregation constraint

$$\mathbf{Y} = f(\mathbf{y}),$$

where $\mathbf{Y}$ is $N \times 1$ (aggregate), observed (e.g., total financial openness), $\mathbf{X}$ is $(N-1)N \times k$ (bilateral), observed (e.g., size, distance, common border ...) $\mathbf{y}$ is is $(N-1)N \times 1$ (bilateral), unobserved (e.g., bilateral financial openness)

- Goal: Estimation of β from observed data

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

The case of gravity equations

- For our bilateral trade model

$$f(\mathbf{y}) = \mathbf{S}\exp(y)$$

where

$$\mathbf{S} = \begin{bmatrix} \\ 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 \end{bmatrix}_{N\times(N-1)N}$$

- Bilateral data are nonlinearly transformed (exponentiated) and summed over partner countries

Estimation

- Following Proietti (JCGStat, 2006), we linearize the aggregation constraint around some value $\mathbf{y}^*$,

$$\mathbf{Y} \approx \mathbf{Y}^* + \mathbf{A}^*(\mathbf{y} - \mathbf{y}^*),$$

  where $\mathbf{A}^*$ is the Jacobian evaluated at $\mathbf{y}^*$, with a typical element $a_{ij} = \partial f_i / \partial y_j$
- The model can be estimated in a straightforward manner using linear methods,

$$\begin{aligned}
\mathbf{Y} &\approx \mathbf{Y}^* + \mathbf{A}^*(\mathbf{y} - \mathbf{y}^*), \\
\mathbf{Y} - \mathbf{Y}^* + \mathbf{A}^*\mathbf{y}^* &\approx \mathbf{A}^*\mathbf{y}, \\
\mathbf{Y} - \mathbf{Y}^* + \mathbf{A}^*\mathbf{y}^* &\approx \mathbf{A}^*(\mathbf{X}\beta + \mathbf{u}), \\
\underbrace{\mathbf{Y} - \mathbf{Y}^* + \mathbf{A}^*\mathbf{y}^*}_{\tilde{\mathbf{Y}}^*} &\approx \underbrace{\mathbf{A}^*\mathbf{X}}_{\tilde{\mathbf{X}}^*}\beta + \underbrace{\mathbf{A}^*\mathbf{u}}_{\tilde{\mathbf{u}}^*}, \\
\tilde{\mathbf{Y}}^* &\approx \tilde{\mathbf{X}}^*\beta + \tilde{\mathbf{u}}^*
\end{aligned}$$

Estimation

- Iterative procedure
  - Estimate β for the trial $\mathbf{y}_0^*$
  - Construct artificial bilateral data as $\mathbf{y}_1^* = \mathbf{X}^*\hat{\beta} + \hat{\tilde{\mathbf{u}}}_0^*$
  - Reestimate the model using $\mathbf{y}_1^*$ as trial value
  - Iterate until convergence
- How much voodoo is involved? A simulation study
  - Simulated data using

$$y_{ij} = 0.5 + 0.1x_{ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathsf{NID}(0, \sigma^2),$$

  - $x_{ij} \sim \mathsf{NID}(0, 1)$
  - Aggregated data obtained as $Y_i = \sum_{j=1}^{J} \exp(y_{ij})$ for $i = 1, \ldots, I$
  - Settings for size of dataset: $I = J = 10$, $I = J = 50$ and $I = J = 100$
  - Settings for error variance: $\sigma = 0.1$ and $\sigma = 0.25$
  - Results based on 1000 replications

Simulation results

Dimension	σ	Mean	Median	Std. Dev.	Skew.
10 × 10	0.1	0.102	0.101	0.037	0.197
10 × 10	0.25	0.102	0.098	0.109	0.079
50 × 50	0.1	0.100	0.100	0.015	0.021
50 × 50	0.25	0.108	0.107	0.040	-0.024
100 × 100	0.1	0.106	0.106	0.027	0.120
100 × 100	0.25	0.108	0.107	0.027	0.195

A small-scale application: Intra-EU trade

- Bilateral trade flows for 14 EU countries (Austria, Belgium, Denmark, France, Germany, Italy, Netherlands, Sweden, Finland, Greece, Ireland, Portugal, Spain and the UK)
- The bilateral model

$$\log T_{ij} = \beta_0 + \beta_1 \log (GDP_i \times GDP_j) + \beta_2 \log d_{ij} + \varepsilon_{ij},$$

- Aggregated model based on $T_i = \sum_j \exp(\log T_{ij})$

	Bilateral data		**Aggregated data**	
Variable	Estimate	St. dev.	Estimate	St. dev
Intercept	-19.26	2.407	-19.23	0.734
$\log (GDP_i \times GDP_j)$	0.798	0.039	0.787	0.012
$\log d_{ij}$	-1.010	0.095	-0.924	0.029
R-squared	0.902		-	
Obs.	91		14 (aggregated data)	

A small-scale application: Intra-EU trade

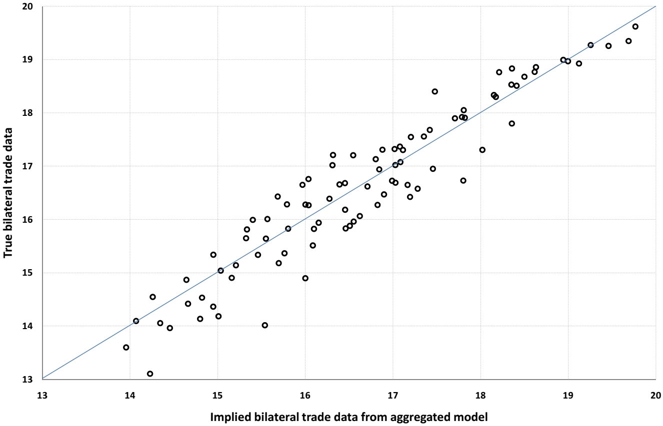

Figure: True versus fitted values of (log) bilateral trade based on the model with aggregated data

Towards a measure of bilateral financial openness

Total	*trade*		*portfolio*		*fdi*	
	coeff.	se	coeff.	se	coeff.	se
const	-5.521	0.414	-9.230	1.075	-3.482	0.773
$\ln(Y_i Y_j)$	0.833	0.011	1.013	0.0298	0.913	0.0214
$\ln D_{ij}$	-0.969	0.027	-1.166	0.071	-1.563	0.051

EU-14	*trade*		*portfolio*		*fdi*	
	coeff.	se	coeff.	se	coeff.	se
const	-2.396	0.541	-1.943	1.072	1.733	0.803
$\ln(Y_i Y_j)$	0.778	0.015	0.837	0.030	0.791	0.022
$\ln D_{ij}$	-1.264	0.036	-1.578	0.071	-1.900	0.053

Conclusions

- We present a method to estimate bilateral models when bilateral data are not available, but some (nonlinear) aggregation of the dependent variable exists
- The method can be used to construct weighting matrices for spatial econometric models where "space" is understood as eventually encompassing other exogenous characteristics different from pure geographical distance
- Our method opens the door to the quantitative (spatial) analysis of socio-economic relationships whose study was hitherto impossible due to data constraints
- Research in progress: Use estimated time-varying exogenous bilateral financial openness as a building block for spatial models of financial instability contagion
- Forthcoming research questions: Migration models
- Model uncertainty can be built in the method in a relatively straightforward manner