Specification Issues in Panel Data Analyses

Short Seminar at UTD
2008

Thomas Plümper

University of Essex

Before we start: Specification versus (?) Econonometric Textbook Wisdom

Three Problems of Econometric Textbook Wisdom:

– consistency as criterion for the choice of estimators

– choice of estimator discussed without discussion of ‘what does the estimator do to the data’ (King 1980)

– the correct model is assumed to be known

Specification versus (?) Econonometric Textbook Wisdom

What discussion of specification can clarify

- small sample properties of estimators (unbiasedness versus efficiency)
  best estimator for data at hand

- theory first, estimator second
  best estimator for theory at test

- robustness and sensitivity analyses helps
  discussion of results given model uncertainty

What to do (and not to do) with Time-Series Cross-Section Data: A First and Some Second Opinions

Thomas Plümper

draws on joint work with Vera Troeger and Philip Manow

University of Essex
Department of Government

Before we get there…

Why everyone should use panel data.

Why using Panel Data?

Bad (= not entirely convincing) answers:

a)

The number of observations increases.

This makes estimation results more significant (not always, of course).

The probability of getting accepted in journals largely increases.

b)

Everyone (well, behavioralists) does it nowadays.

Thus, I do what everyone does. Can’t be that wrong.

At least: The probability of getting accepted in journals largely increases.

Why using Panel Data?

Some better answers:

a)

Theory makes predictions across space and over time.

Researchers therefore need to model both effects.

And they need to model them simultaneously, because they tend to be correlated (but also partly independent).

Thus, to obtain unbiased (this means: probably less biased) results, one cannot exclude either effect.

b)

Human behavior tends to be ‘sticky’ (path-dependent if you wish to say so).

Hence, everything we’re interested in at period t is some function of the same thing at period t-1.

Accordingly, ALL cross-sectional analyses lead to largely biased results, because we cannot account for this temporal dependence.

Why using Panel Data?

One more answer:

With very little exceptions social science theories are formulated without spatial or temporal qualifications. They tend to demand validity at all times at all places.

I.e., there is no theory of voting behavior in the United States in 2008 (at least I have not seen one).

However, testing a general theory with spatially and temporally limited data is somewhat problematic even if researchers claim to analyze a “random” sample:

A random draw from a non-random distribution is not a random sample.

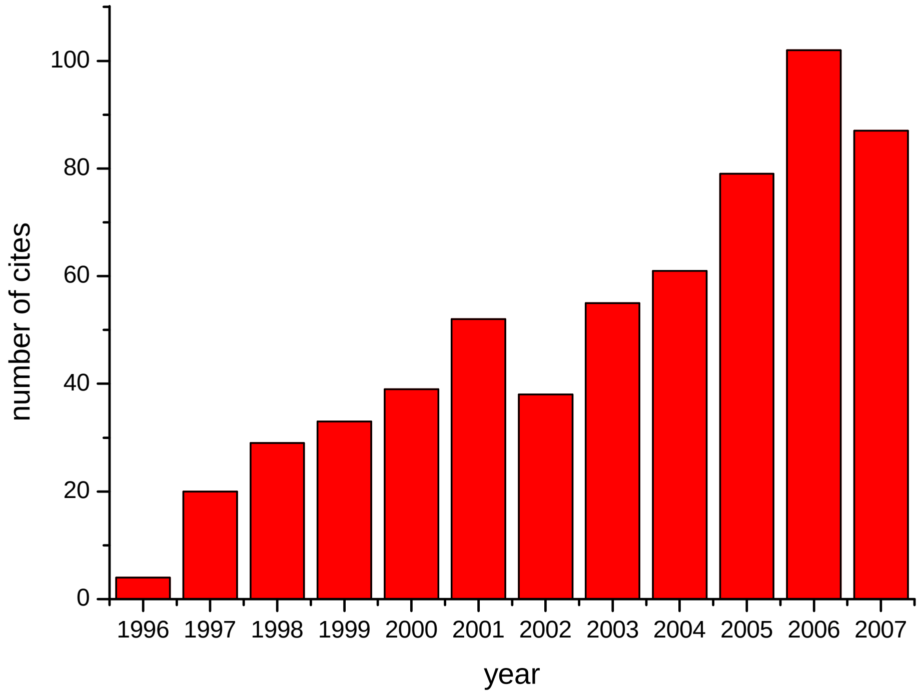
Summary: Why Panel Data Analysis?

Panel Data Analysis is the biggest success story in Political Science.

Four answers:

- Most theories make predictions across time and space.
- Most theories claim universal validity (well wrongly so, but still)
- Pooled Data allows to explicitly model dynamics.
- Pooled data allows controlling for unobserved unit heterogeneity.

First Views: The ‘de facto Beck-Katz-Standard



annual SSCI citations for b/k 95, 1996-2007

First Views: The ‘de facto Beck-Katz-Standard

Problem analyzed in b/k 1995:

Panel-Heteroscadasticity leads to overconfidence if T>N.

Solution: Panel-corrected standard errors (and OLS estimate).

Side effects of OLS estimation (Gauss-Markov conditions):

serial correlation of errors biases estimates (solution: LDV)

unobserved time-invariant unit heterogeneity biases estimates (solution: unit FE)

unobserved common shocks biases estimates (solution: period FE)

The de facto Beck-Katz-Standard: Second Views

Almost no justification for specification decisions in b/k 1995.

LDV because serially correlated errors.

Unit FE because unobserved unit heterogeneity.

Period FE because common shocks.

The de facto Beck-Katz-Standard: Second Views

All solutions suggested by b/k ’95 do the job.

BUT:

Neither of b/k’s solutions is particularly ‘cheap’ (in terms of efficiency).

Other solutions may be more appropriate given the data at hand and the theory which shall be tested.

Specification Issues not mentioned by b/k

Observations not spatially independent

Heterogeneous Dynamics

Heterogeneous Lag Structure

LDV

Dynamics after b/k 95

Chris Achen:

So if lagged budgets are not remarkably influential, why were the statistical studies so wrong? Essentially, the answer lies in the bias results derived earlier: The effects of lagged dependent variables are grossly overestimated when disturbances are heavily trended. Translated to budget data, this means that last year's budget will predict this year's budget very well even if it has little or no real causal impact, so long as other unmeasured influential political factors are strongly trended—much the same from one year to the next.[8] Heavy trending seems likely in budget data most of the time, and if so, the result would be an artificially inflated estimate of the causal power of last year's budget.

This paper has argued that in practice, the anomaly is often due to the combination of high serial correlation and heavy trending in the exogenous variables, which can jointly produce dominating autoregressive terms even when they have little or no real explanatory power.

The one blemish on this otherwise attractive proposal stems from the issue discussed throughout this paper: If serial correlation is present and autoregressive terms are entered into the regression equation, then all the coefficients are biased. In particular, when serial correlation is high and the exogenous variables are heavily trended, as will happen frequently in panel data, the lagged variable will falsely dominate the regression and suppress the legitimate effects of the other variables.

Plümper et al. 2005

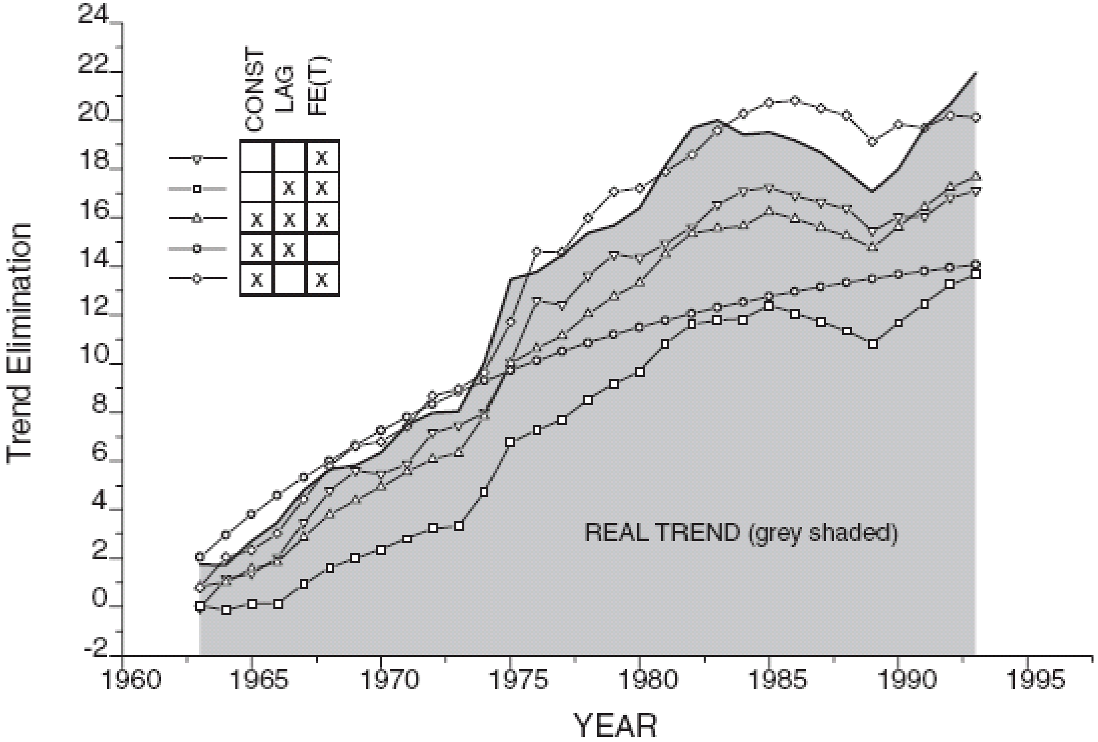


Figure 1. Trend 'elimination' in panel data.

The five variants are highly correlated with each other and with the observed values (all correlation coefficients are larger than 0.90). What does this exercise tell us? First of all, it is important to see that the estimates of the lagged dependent variable, the constant and the period dummies depend to a very high degree on the model specification. Perhaps most importantly for the theoretical discussion of exogenous variables, the inclusion of the lagged dependent variable and/or time dummies leaves very little variance for the explanatory variables.

Plümper et al. 2005 suggest Prais-Winston transformation.

Example of Prais-Winston versus LDV/PW Approach

	Model 2.5	Model 2.6	Model 2.7
	FE(C) $\tilde{\beta}_0$ =0.0	FE(C) $\tilde{\beta}_0$ =0.3	FE(C) $\tilde{\beta}_0$ =0.7
Unemployment	1.041 (0.0978) ***	0.773 (0.0759) ***	0.294 (0.0442) ***
GDP per Capita Growth	-0.339 (0.0309) ***	-0.380 (0.0361) ***	-0.452 (0.0315) ***
Dependency Ratio	0.5566 (0.0444) ***	0.4109 (0.0322) ***	0.2159 (0.0176) ***
Left Cabinet Portfolios	0.0051 (0.0045)	0.0042 (0.0038)	0.0025 (0.0026)
Christian Democrat Portfolios	-0.0088 (0.0081)	-0.0075 (0.0070)	-0.0043 (0.0046)
Trade Openness	0.0308 (0.0245)	0.0265 (0.0216)	0.0103 (0.0149)
Low Wage Imports	-0.0156 (0.0300)	0.0028 (0.0245)	0.0203 (0.0140)
Foreign Direct Investment	0.1150 (0.0979)	0.1095 (0.0843)	0.0583 (0.0583)
N	529	529	529
R^2	0.822	0.805	0.770
Wald χ^2	279377.87	385942.12	348954.87
prob> χ^2	0.000	0.000	0.000
PCSE	yes	yes	yes
time dummies	no	no	no
country dummies	yes	yes	yes
error correction	AR1	AR1	AR1
rho	0.707	0.620	0.313

Adolph, Butler and Wilson 2005

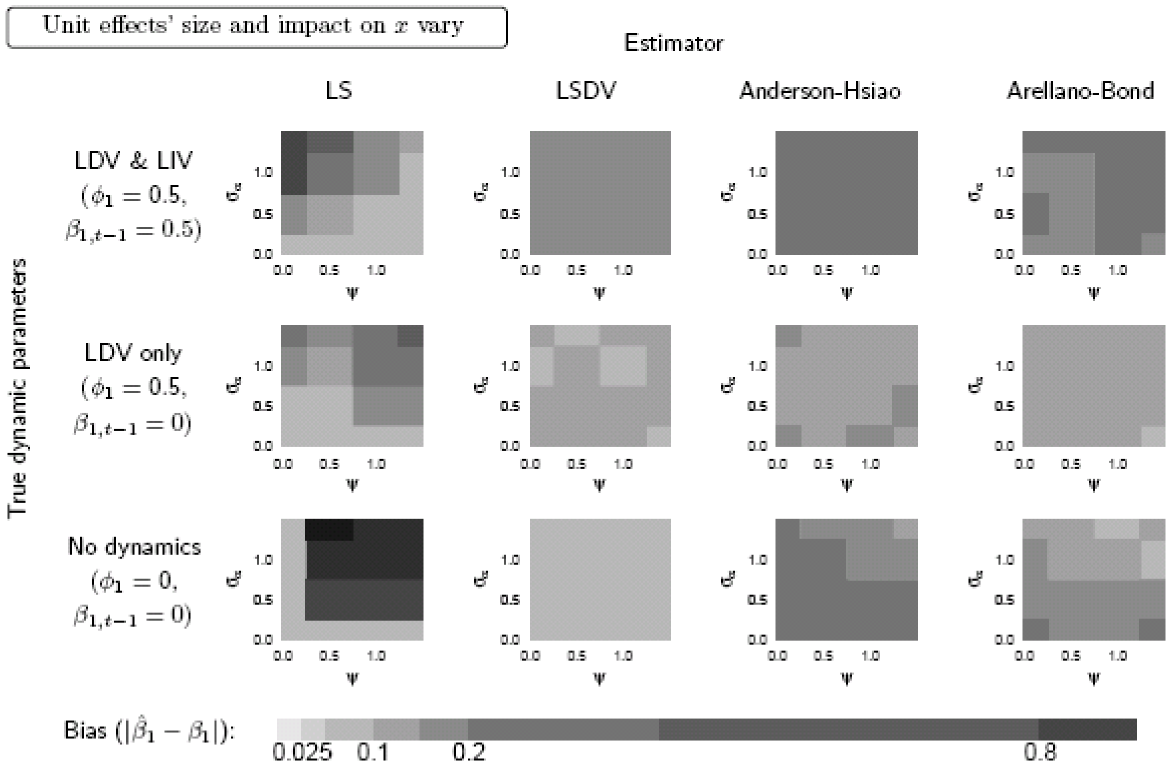


Figure 3: Bias ($|\hat{\beta}_1 - \beta_1|$) in various time series cross-section estimators (columns) under different dynamic parameters (rows), different impacts of unit effects on x (horizontal axes of plots), and different sizes of unit effects (vertical axes). Each cell in each plot shows the bias in a different scenario for the given combination of estimator and dynamics (all models are correctly specified). For all scenarios, 1000 simulations were drawn from a TSCS process with $N = 10$, $T = 20$, $\beta_{1,t} = 1$, $\sigma_\varepsilon = 1$, $\delta = 0.5$, $\sigma_\omega = 0.6$, and $\sigma_\eta = 1$. Each of the N time series begins with a burn-in period of 50 observations, which is discarded.

where

4.1 Data generating process

We model the TSCS data, y_{it}, $i = 1, \ldots, N$, $t = 1, \ldots, T$ as follows:

$$\begin{aligned}
y_{it} &= \phi_1 y_{i,t-1} + \phi_2 y_{i,t-2} + \beta_1 x_{it} + \beta_{1,t-1} x_{i,t-1} + \beta_{1,t-2} x_{i,t-2} + \alpha_i + \varepsilon_{it} \\
x_{it} &= \delta x_{i,t-1} + \psi \alpha_i + \omega_{i,t} \\
\varepsilon_{it} &\sim \mathcal{N}(0, \sigma^2_\varepsilon) \\
\omega_{it} &\sim \mathcal{N}(0, \sigma^2_\omega) \qquad (1) \\
\alpha_i &\sim \mathcal{N}(0, \sigma^2_{\alpha,i}) \\
\sigma^2_{\alpha,i} &= \sigma^2_\alpha + \eta_i \\
\eta_i &\sim \mathcal{N}(0, \sigma^2_\eta)
\end{aligned}$$

Note:

Anderson-Hsiao:

first difference in y on first difference in x’s and lagged differences in y using lagged level of y as instrument

Arellano-Bond: GMM estimator, all available lags as instrument
estimation in differences

Comment on AH and AB: differencing usually eliminates serial correlation of errors AND unit effects. Why then do we need instruments?

Dynamics: A preliminary Conclusion

The LDV eliminates serial correlation of errors, which otherwise would bias OLS estimates.

The LDV coefficient is biased upwards if dynamics of regressors are incorrectly specified and if data is trended and if this trend is not exogenous.

LDVs are difficult to interpret (and usually incorrectly interpreted)

$$y(x)_{t_1 \to t_p} = \beta_1 x_{it} + \sum_{p=1}^{t_p} \left(\beta_0^{t-p} \beta_1 x_{it} \right) \quad (1)$$

$$s.d.\left(\hat{\beta}_1\right) + \sum_{p=1}^{t_p} \left[s.d.\left(\hat{\beta}_1\right) \right]^{t-p} \quad (2)$$

Thus, the odds are that substantive variables have a larger effect (and relatively smaller s.e.'s than we would believe by looking at the regression results!

Alternatives to the LDV Approach

Other ways of dealing with serial correlation are apparently less costly, but not always successful.

Differencing: problematic because time lags need to be correctly specified.

Unit FE/ Period FE: in itself costly, does not necessarily eliminate serial correlation, captures common trends

Distributed Lag Model: if LDV included: see above, if LDV excluded: autocorrelation often not fully eliminated

Prais-Winston: less costly than LDV, not always as successful in eliminating serial correlation of errors.

DLM/Prais-Winston: unknown small sample properties, some MC analysis needed

Additional Complications

Unit Root (non-stationarity)

Heterogeneous Lag Structure (we get to that a little later)

Time-variant Coefficients

Unit Root

testing for unit root: Levin and Lin, Im, Peseran and Shin, Maddala, …

very unreliable tests: biased against rejecting unit root hypothesis

simple alternative: xtreg y l.y, fe,
convention: unit root if confidence band of beta(l.y) includes the 1.

Remedies:

estimate in differences (what about the theory???)

panel co-integration (in its infancy), T usually too small, time-series in Comparative Politics often not co-integrated

Slope Heterogeneity

the effect of a specific independent variable is not stable over time

Example: preferences and behaviour of left parties have changed over time
(labor-new labor, SPD of Willy Brandt and Schröder)

How to estimate?

xtreg y x*period1 x*period2 x*periodT

then do a Chow-test for the differences of period estimates

Example: Party Positions and Government Spending

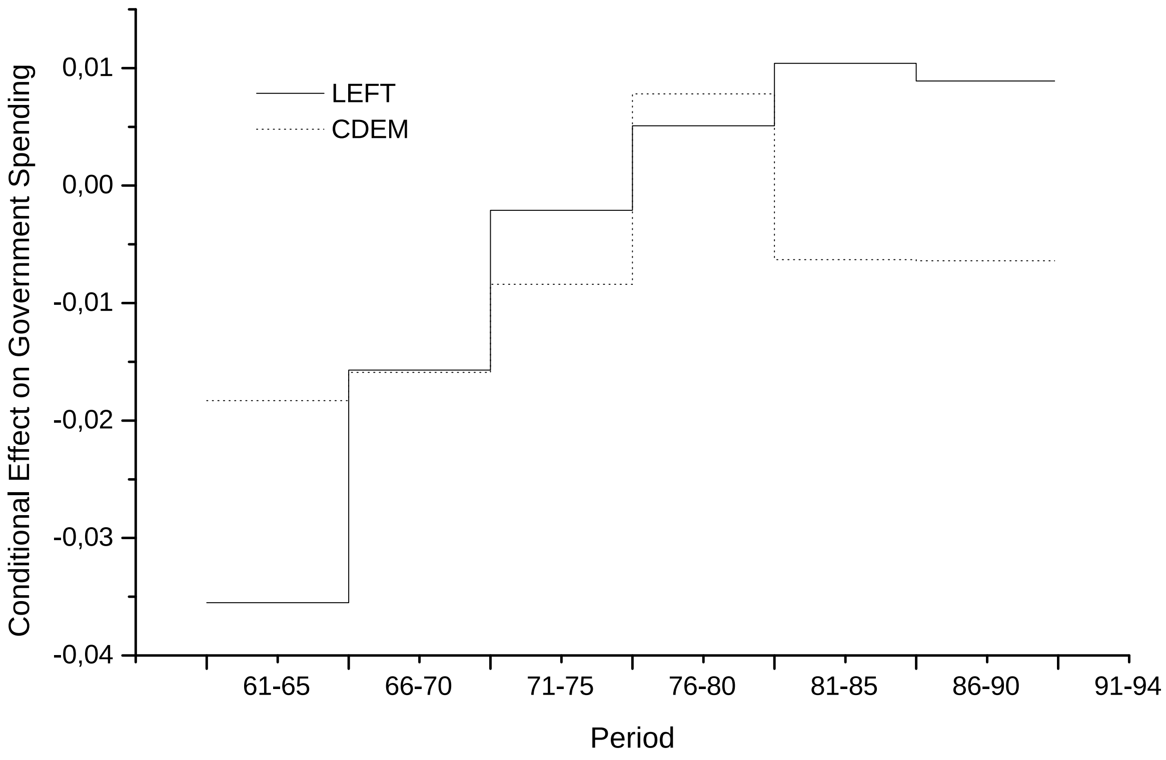


Figure 2: The Conditional Effects of LEFT and CDEM in Government Spending

Heterogeneity: A Summary of Issues

Take for example

$$y_t = \alpha + \rho y_{t-1} + \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t$$

and ‘pool’ it across space so that we get

$$y_{it} = \alpha + \rho y_{i,t-1} + \beta_1 x_{it} + \beta_2 x_{it-1} + \varepsilon_{it}$$

But is that already appropriate? Isn’t the world more complicated?

These are questions for the theory.

So think about heterogeneity.

A Simple Look at Heterogeneity

What could be heterogeneous?

$$y_{it} = \alpha + \rho y_{i,t-1} + \beta_1 x_{it} + \beta_2 x_{it-1} + \varepsilon_{it}$$

Let's start with: What is not problematic:

If var(x) increases, that's called variance. Do not believe that we have a problem because the units look differently. We want that. We need that.

A Simple Look at Heterogeneity

1. The intercept:

$$y_{it} = \alpha_i + \rho y_{i,t-1} + \beta_1 x_{it} + \beta_2 x_{it-1} + \varepsilon_{it}$$

This is the standard fixed effects unit heterogeneity. I will come back to that.

2. The betas

$$y_{it} = \alpha_i + \rho y_{i,t-1} + \beta_{i1} x_{it} + \beta_{i2} x_{it-1} + \varepsilon_{it}$$

Beck and Katz suggest doing random coefficients. I suggest modeling this heterogeneity explicitly and do fixed coefficients (estimate N coefficients, for each unit one).

3. The dynamics (rho)

$$y_{it} = \alpha_i + \rho_i y_{i,t-1} + \beta_{i1} x_{it} + \beta_{i2} x_{it-1} + \varepsilon_{it}$$

I do not know of papers doing this. However, Stata allows to do unit-specific Prais-Winston transformation.

Random coefficient models with LDV would do it.

4. The lag structure

$$y_{it} = \alpha_i + \rho_i y_{i,t-1} + \beta_{i1} x_{it} + \beta_{g1} x_{it-1} + \beta_{g2} x_{it-2} + \varepsilon_{it}$$

Not much used, but see Plümper, Troeger and Manow EJPR 2005 for a brief discussion.

We can deal with all these complications by using an appropriately specified estimation model. Perhaps we cannot deal with all complications simultaneously. But researchers have to learn how to think about specification and heterogeneity.

Solutions to Different Types of Heterogeneity

1. Different intercepts

   first difference models, fixed effects model

2. Different lag structures

   no textbook solution available, a data-mining exercise is discussed in Plümper et al. 2005

3. Different slopes

   random coefficient model, SUR model

4. Different dynamics

   hardly discussed in textbooks, random coefficient model with LDV, xtreg with psar(1)

Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects

Political Analysis 15:2, 2007, 124-139.

Thomas Plümper and Vera E. Troeger

University of Essex
Department of Government

A Closer Look at the Title

Efficient Estimation: minimum sampling variance of an estimate/ estimator

Time-Invariant Variables: Regressors with a within variance of 0

Rarely Changing Variables: Regressors with a small positive within variance

Finite Sample: Sample size smaller than infinity

Panel Data Analyses: Regression that simultaneously uses time-series and cross-sectional information (that ‘pools’ various cases over various periods)

Unit Fixed Effects: unobserved unit-specific heterogeneity

The Complication(s) this Presentation is Dealing with…

assume units are heterogeneous
(unobserved time-invariant variance exists)

this unobserved heterogeneity is correlated with some of the regressors
hence: doing nothing leads to biased estimates)

some of the variables included are time-invariant or have very little within variance

Why is this at all a problem?

the solution to the problem of correlated unit effects (a fixed effects model) precludes the estimation of coefficients for time-invariant variables

AND

render the estimation of almost time-invariant variables inefficient.

A quote:

“Although we can estimate (…) with slowly changing independent variables, the fixed effect will soak up most of the explanatory power of these slowly changing variables. Thus, if a variable (…) changes over time, but slowly, the fixed effects will make it hard for such variables to appear either substantively or statistically significant.“ (Beck 2001: 285)

How typical is the combination of these three conditions?

That is: how important is the problem we are solving?

unit effects almost ever exist since the time-series information social scientists are able to gather hardly ever begins at t_0

in almost all data sets, the unit effects are correlated with some of the regressors
(Hausman-test)

and: in almost all data sets almost time-invariant variables do exist, often time-invariant variables exist

Time-Invariant Variables

Two categories:

1) time-invariant by definition:

- geography (being a European country, being landlocked),
- inheritance (former colony, sex, …)

2) time-invariant for the period under analysis or because of researchers’ selection of cases:

- constitutions
- institutions
- number of siblings
- sex
- education level

Rarely Changing Variables / Almost Time-Invariant Variables

A small change in the sample can turn time-invariant variables of the second category into a variable with very low within variation – an almost time-invariant or rarely changing variable.

1) variables that change only once in a while

- level of democracy
- status of the president
- electoral rules
- central bank independence
- federalism
- family income
- marital status

2) variables that change continuously, but depending on the sample, the between variance can still exceed the within variance

- government spending
- per capita income
- household income

If problems are that common, solutions must exist…

what textbooks recommend:

time-invariant:	Hausman-Taylor
rarely changing:	?

what applied researchers do

time-invariant:	random effects
	pooled-OLS
	Hausman-Taylor (some economists)
rarely changing:	fixed effects
	random effects (if they do not like the coefficients of FE estimation)
	pooled OLS

Unfortunately, these 'solutions' are pretty bad…

(as I will demonstrate pretty soon, I hope)

Hausman-Taylor: inefficient, biased if instruments are poor

random effects: biased

pooled-OLS: as biased as random effects, slightly less efficient if uncorrelated unit effects exist

A superior solution: fixed effects vector decomposition

the xtfevd procedure:

stage 1:

estimation of the unit fixed effects by the baseline panel fixed effects model excluding the time-invariant right hand side variables

stage 2:

regression of the unit effects on the time-invariant and/or almost time-invariant variables

stage 3:

re-estimation of stage 1 model by pooled OLS including
- time-varying variables
- time-invariant variables
- the unexplained part of the fixed effects vector (residuals from stage 2)

Additional features

– correction of standard errors
– dynamics (AR1)
– robust standard errors
– panel-corrected standard errors
– instrumental estimation for endogenous variables

Some Equations and all that Math Jazz

The DGP

$$y_{it} = \alpha + \sum_{k=1}^{K} \beta_k x_{kit} + \sum_{m=1}^{M} \gamma_m z_{mi} + u_i + \varepsilon_{it}$$

Averaging:

$$\bar{y}_i = \sum_{k=1}^{K} \beta_k \bar{x}_{ki} + \sum_{m=1}^{M} \gamma_m z_{mi} + \bar{e}_i + u_i$$

where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$, $\bar{e}_i = \frac{1}{T}\sum_{t=1}^{T} e_{it}$

Demeaning (Stage 1 estimation):

$$y_{it} - \bar{y}_i = \beta_k \sum_{k=1}^{K} \left(x_{kit} - \bar{x}_{ki} \right) + \gamma_m \sum_{m=1}^{M} \left(z_{mi} - z_{mi} \right) + \left(e_{it} - \bar{e} \right)_i + \left(u_i - u_i \right)$$

$$\equiv \ddot{y}_{it} = \beta_k \sum_{k=1}^{K} \ddot{x}_{kit} + \ddot{e}_{it}$$

We obtain:

$$\hat{u}_i = \overline{y}_i - \sum_{k=1}^{K} \beta_k^{FE} \overline{x}_{ki} - \overline{e}_i$$

which we decompose in stage 2 according to:

$$\hat{u}_i = \sum_{m=1}^{M} \gamma_m z_{mi} + h_i$$

so that

$$h_i = \hat{u}_i - \sum_{m=1}^{M} \gamma_m z_{mi}$$

In stage 3 we then regress

$$y_{it} = \alpha + \sum_{k=1}^{K} \beta_k x_{kit} + \sum_{m=1}^{M} \gamma_m z_{mi} + \delta h_i + \varepsilon_{it}$$

Do you really believe me?

On page 44, I have claimed that fevd is superior to all alternatives.

Do you really believe me?

On page 44, I have claimed that fevd is superior to all alternatives.

Those of you who fully trust me: You can go now. Thank you.

Do you really believe me?

For those of you, who don’t, I demonstrate that the claim is correct.

A Monte Carlo Analysis of Competing Procedures' Finite Sample Properties

Monte Carlos…

- Define a Data Generating Process (A TRUE MODEL).
- Add a stochastic element to the dependent variable.
- Estimate the model by various estimators.
- Replace the stochastic element by another random draw.
- Re-estimate the models (well, do that 1000 times).
- Systematically change the DGProcess (redo everything as often as necessary)

Design

DGP:

$$y_{it} = \alpha + \beta_1 x1_{it} + \beta_2 x2_{it} + \beta_3 x3_{it} + \beta_4 z1_i + \beta_5 z2_i + \beta_6 z3_i + u_i + \varepsilon_{it}$$

x1, x2, x3 are time varying variables
z1, z2 and z3 are the time invariant variables
z3 and x3 are correlated with the unit effects – analytically interesting

True parameter values:

$$\alpha = 1,\ \beta_1 = 0.5,\ \beta_2 = 2, \beta_3 = -1.5,\ \beta_4 = -2.5,\ \beta_5 = 1.8,\ \beta_6 = 3$$

Corr(x3,u)={0.01, 0.1,…, 0.9, 0.99), Corr(z3,u)={0.01, 0.1,…, 0.9, 0.99)

N={15, 30, 50, 70, 100}, T={20, 40, 70, 100} (not reported here)

1000 experiments for different correlations of z3,u and x3,u

Criterion

The mean deviation of the estimated coefficient from the true value of that coefficient:

$$\mathrm{RMSE}\left(\hat{\beta}\right) = \frac{1}{K}\sum_{k=1}^{K}\sqrt{\left(\hat{\beta}-\beta_{\mathrm{true}}\right)^2}$$

Needlees to say: a higher RMSE is worse…

Time-Invariant Variables

Competing Estimators:

- Pooled OLS
- Random Effects
- Hausman-Taylor (Instrumental Equation RE Model)
- Fixed Effects Vector Decomposition

Summary of Results

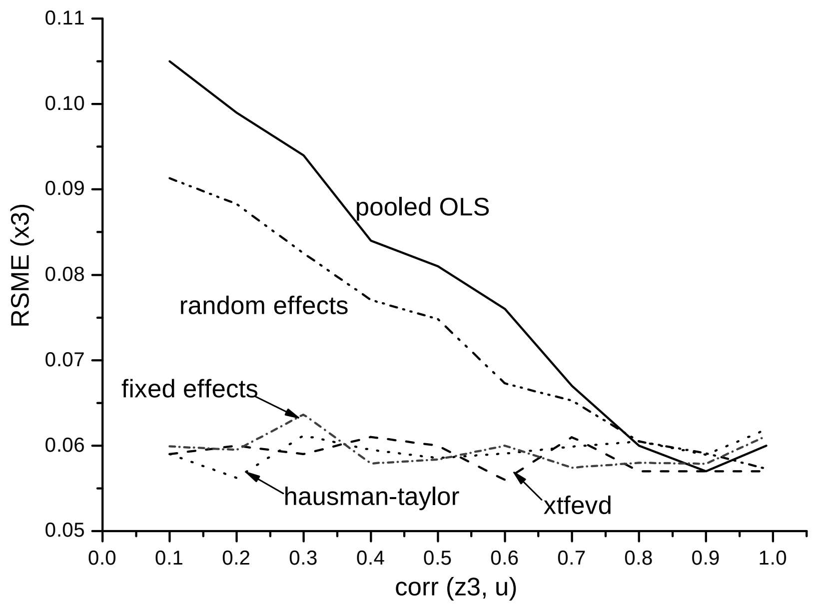
	p-ols		*fevd*		Hausmann-Taylor		RE		FE	
	RMSE	average bias	RMSE	average bias	RMSE	average bias	RMSE	average bias	RMSE	average bias
time-varying variable x3	0.187	-0.167	0.103	0.001	0.105	-0.003	0.173	-0.149	0.103	-0.001
time-invariant variable z3	0.494	-0.470	0.523	-0.548	1.485	-1.128	0.506	-0.481	.	.

Settings of the parameter held constant:
N=30, T=20
Corr(u,x1)=corr(u,x2)=corr(u,z1)=corr(u,z2)=0
Corr(u,x3)=0.5

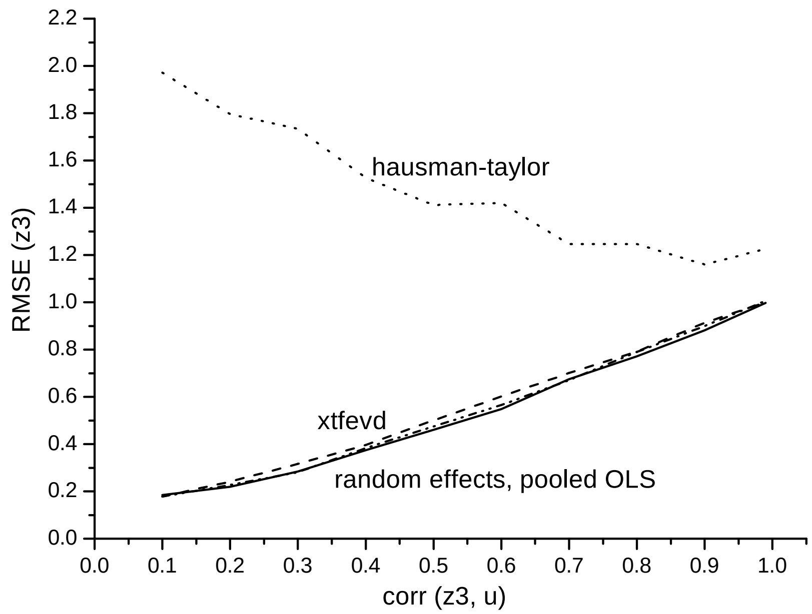
Settings of the varying parameter:
Corr(u,z3)={0.1, 0.2,…, 0.9, 0.99}

Changes in corr(z3, u_i)

affect



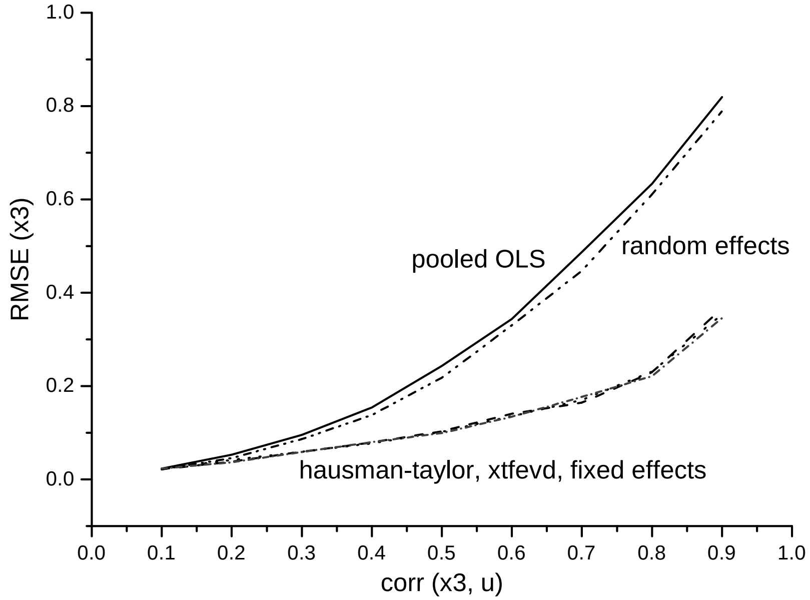
the RMSE of x3



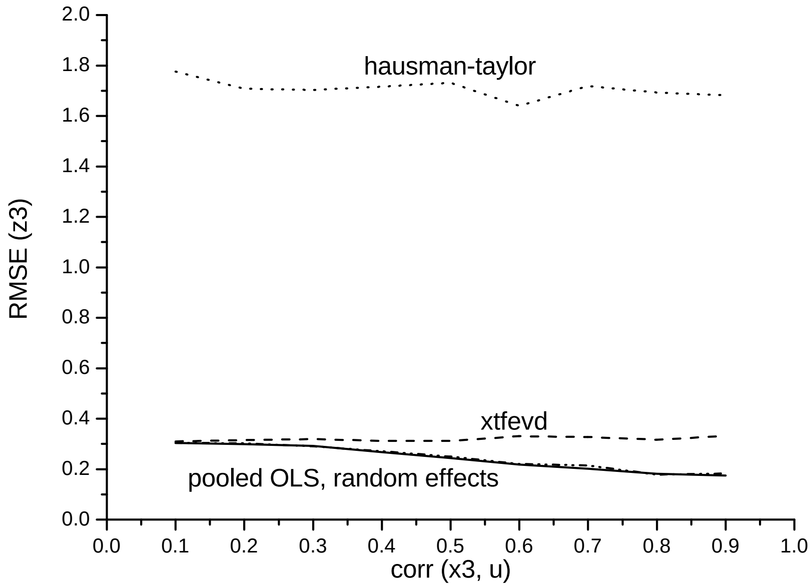
the RMSE of z3

Changes in corr(x3, u_i)

affect



the RMSE of x3



the RMSE of z3

Summary of Results

fevd, random effects and pooled OLS perform best in estimating the coefficient of the correlated time-invariant variable z3.

fixed effects, Hausman-Taylor and fevd perform vest in estimating the coefficient of time-varying variable x3.

Hence, there is only one optimal procedure if both, time-varying and time-invariant variables are correlated with the unit effects: fevd.

Almost Time-Invariant Variables

Competing Estimators:

- Fixed Effects
- Fixed Effects Vector Decomposition

Previous results for random effects and pooled-OLS carry over to rarely changing variables.

One step back…

The fixed effects model undoubtly computes coefficients for almost time-invariant variables.
So what is the problem?

FE models are inefficient, because

The unit dummies reduce the degrees of freedom (important if T very small).

The ‘within transformation’ ignores the between variation and thus does not take all the available information into account. (important if the between variation > within variation).

Summary of Results

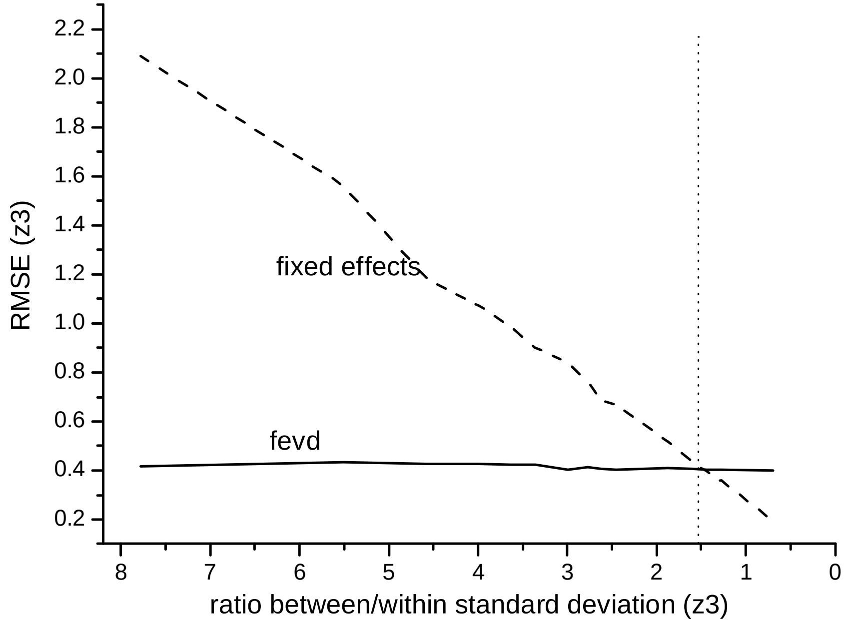
	p-ols		*fevd*		RE		FE	
	RMSE	average bias	RMSE	average bias	RMSE	average bias	RMSE	average bias
time-varying variable x3	0.265	-0.265	0.069	0.001	0.230	-0.230	0.069	0.000
Rarely changing variable z3	0.133	0.028	0.131	0.001	0.133	0.027	0.858	0.008

parameters held constant:	varied parameters:
N=30, T=20 corr(u,x1)=corr(u,x2)=corr(u,z1) =corr(u,z2)=0 corr(u,z3)=0.3 corr(u,x3)=0.5 Between SD (z3)=1.2	Within SD (z3)={0.04,…,0.94}

Table 2: Average RMSE and bias over 10 permutations à 1000 estimations

Ratio Between/Within Standard Deviation (z3)

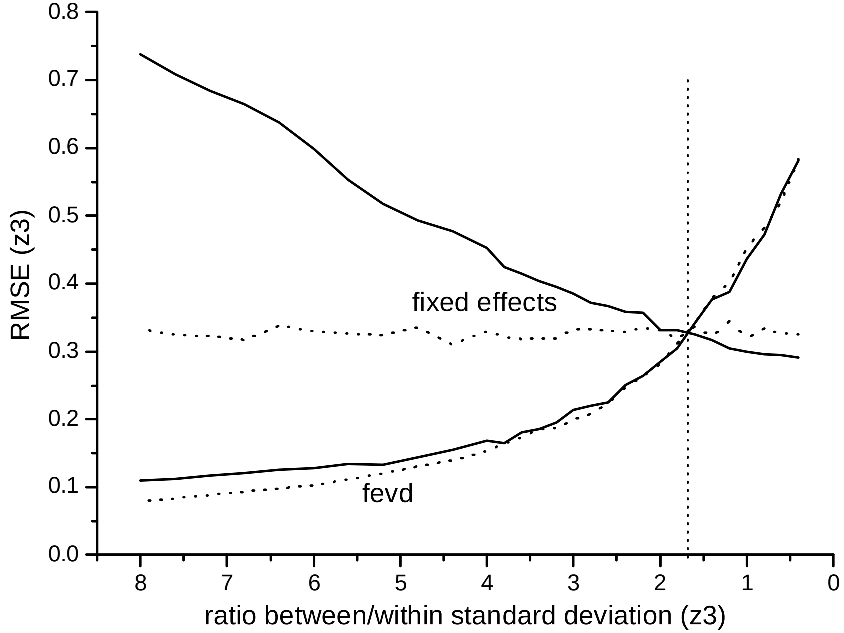
varying the within variance



Parameter settings:
N=30
T=20
rho(u,x3)=0; rho(u,z3)=0.3
between SD (z3): 1.2
within SD (z3) 0.15…1.73

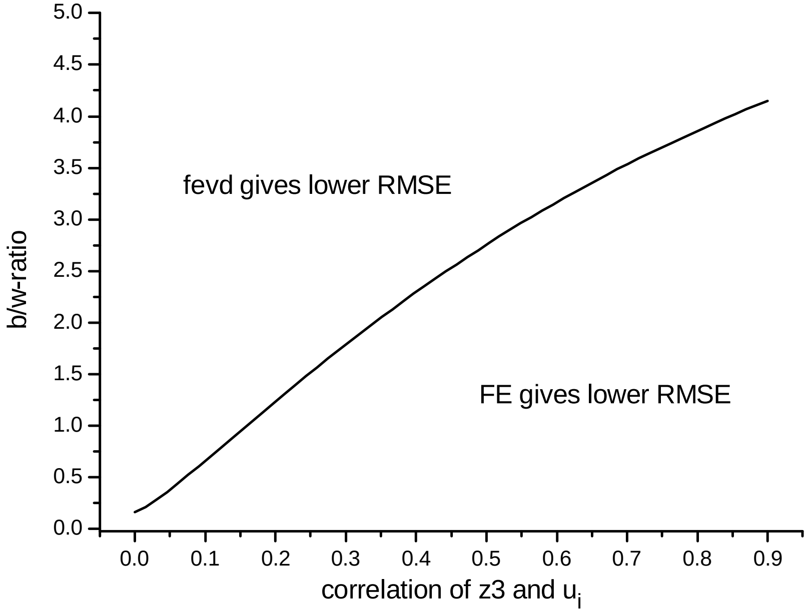
Ratio Between/Within Standard Deviation (z3)

varying the between variance



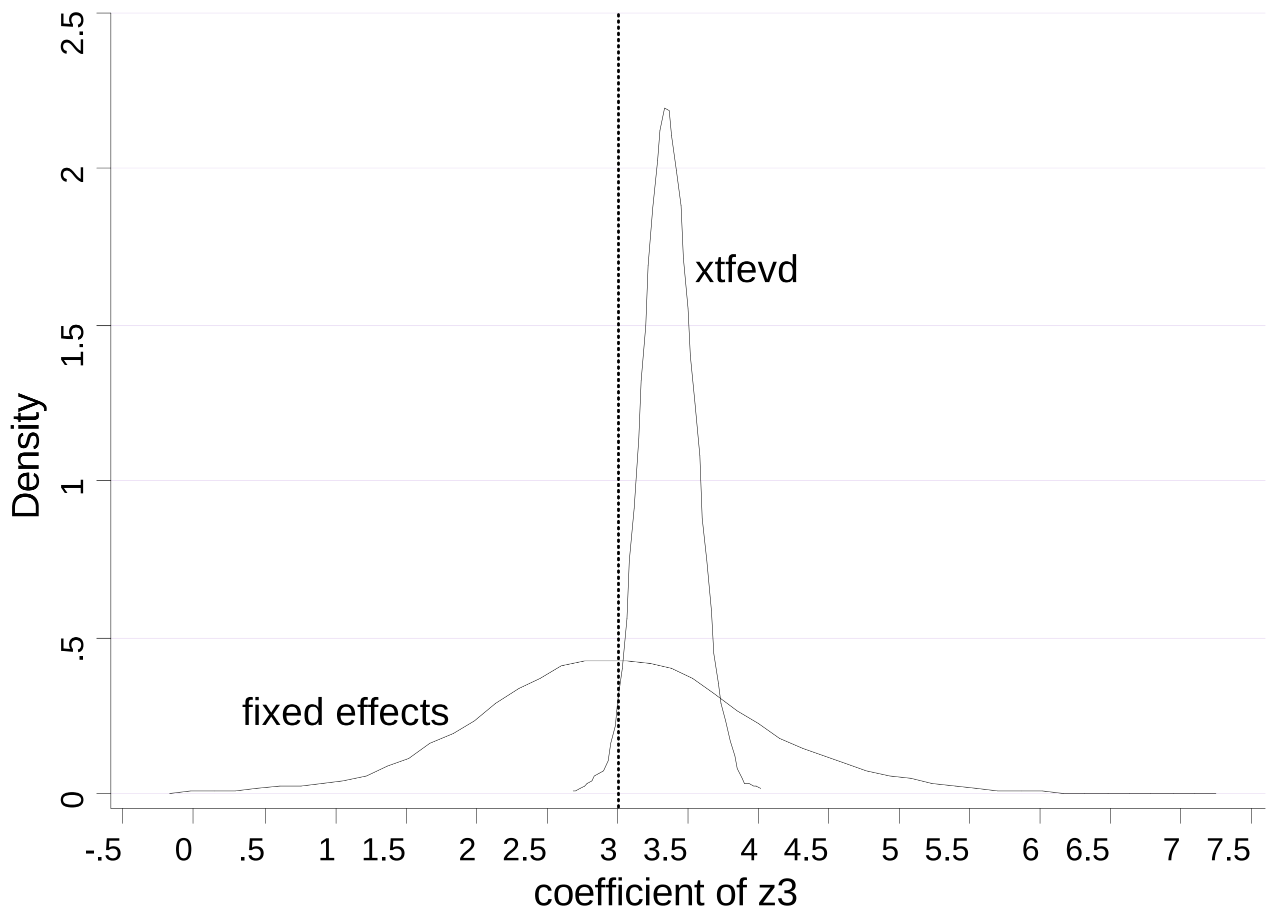
Parameter settings:
N=30
T=20,
rho(u,x3)=0
rho(u,z3)=0.3
between SD (z3): 0.4…8
within SD (z3): 1

When to prefer fevd over fe (and vice versa)?



Parameter settings:
N=30
T=20,
$R^2 = 0.5$
rho(u,x3)=0,
rho(u,z3)= {0, 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 0.9}
between SD (z3): 0.4…8
within SD (z3): 1

Trade-Off between Efficiency and Bias



Bias, Efficiency, and Point Estimates

	estimator	beta (z3)	confidence	interval
# 247	fixed effects	5.3604	3.5486	7.1722
# 247	xtfevd	3.3913	3.0541	3.7286
# 521	fixed effects	0.8269	-0.9035	2.5573
# 512	xtfevd	3.2143	2.8920	3.5367

Summary of Results

Inefficiency translates into unreliable point estimates for the FE model.

The decision whether to treat a variable as time invariant or varying depends on the ratio of between and within variation of this variable.

If the within variation is quite small and the between variation very large, treating the variable as time-invariant in an fevd model gives more efficient and probably less biased point estimates.

A Point of Caution

Contrary to what some want to believe, fevd computes correct standard errors if the model is correctly specified

However, for the model to be correctly specified, it is required that a time-invariant variable or a slowly changing variable exerts a continuing effect on the dependent variable – an effect which can be and in fact is observed in every period.

Examples:

being landlocked on economic growth
size (and quality) of basketball player of # of points
latitude on temperature

If time-invariant variables are observed less often observed than there periods, the fevd standard errors become too small.

Final Remarks

xtfevd is a useful estimation procedure

and superior to alternative estimators when

u is correlated with z

and

z is time-invariant or rarely changing

and (in the latter case)

the between variance is larger than the within variance.

Example

	pooled OLS, ldv	FE, ldv	xtvefd, ldv	FE, ar1	*xtfevd*, ar1
gdp	0.250 (0.006)***	0.342 (0.013)***	0.342 (0.005)***	0.890 (0.015)***	1.050 (0.009)***
population	-0.059 (0.006)***	0.143 (0.068)**	-0.008 (0.006)	-0.823 (0.024)***	-0.290 (0.012) ***
distance	-0.328 (0.012)***	Dropped	-0.577 (0.011)***	Dropped	-1.287 (0.022)***
allied	-0.247 (-0.027)***	0.419 (0.121) ***	-0.396 (0.026)***	-0.040 (0.210)	-0.658 (0.051)***
min democ.	0.022 (0.001)***	-0.009 (0.002)***	-0.009 (0.001)***	-0.008 (0.004)*	-0.008 (0.002)***
lag. trade	0.736 (0.002)***	0.533 (0.003)***	0.533 (0.003)***		
eta			1.000 (0.009)***		0.982 (0.006)***
cons	-3.046 (0.177)***	-13.745 (1.676)***	-4.047 (0.167)***	-0.025 (0.091)	-14.773 (0.326)***
Nobs	88946	88946	88946	85867	88946
rho				0.589	0.532
F		14649.22***	14649*** 5343***	2282.00***	10497.33***
adj. R^2	0.73	0.692	0.692	0.221	0.414

Table 4: Re-analysis of Green et al. 2001

Advertisement

you can download a Stata ado and a Stata helpfile at

www.polsci.org/pluemper/xtfevd.htm

and do not forget to cite

Thomas Plümper and Vera E. Troeger (2007): “Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects”

Political Analysis 15:2, 124-139

Model Specification in the Analysis of Spatial Dependence

Thomas Plümper and Eric Neumayer

University of Essex
Department of Government

University of Essex

What is Spatial Dependence?

“Political units often spatially depend on each other in their policy choices. For example, capital tax rates in one country are typically affected by tax policies in other countries. Similar patterns of spatial dependence have been studied in areas as diverse as social policies (Franzese and Hays 2006, Brooks 2007, Cho 2003, Bailey and Rom 2004; Jahn 2006), monetary policies (Simmons and Elkins 2004; Plümper and Troeger 2008), tax and fiscal policies (Basinger and Hallerberg 2004; Hays 2003, 2008; Swank 2006), trade and investment policies (Mansfield and Reinhardt 2003; Elkins, Guzman and Simmons 2006), military spending and armed conflict (Shin and Ward 1999; Salehyan and Gleditsch 2006), democratization (Gleditsch and Ward 2006), and many others.”

Theories:

– policy spill-over

– policy diffusion

– policy clustering (?)

– policy convergence

Spatial Lags are also important for making correct Inferences

If spatial dependence is ignored, studies will be biased toward finding domestic, internal factors to be more important than international, external ones.

Franzese and Hays Political Analysis 2007

3 different processes:

$$y_{it} = \alpha + \rho \mathbf{W} y_{-it} + ... + \varepsilon_{it} \quad \forall \; i \neq -i \quad . \qquad (1)$$

$$y_{it} = \alpha + \rho \mathbf{W} x_{-it} + ... + \varepsilon_{it} \quad \forall \; i \neq -i \quad . \qquad (2)$$

$$y_{it} = \alpha + \rho \mathbf{W} \hat{u}_{-it} + ... + \varepsilon_{it} \quad \forall \; i \neq -i \quad . \qquad (3)$$

where $W = w_{i-i}$

Estimation of Spatial Dependence

Spatial-OLS

Spatial-2SLS

Spatial-ML

Franzese and Hays (personal communication) nowadays claim:

S-ML most efficient.

S-2SLS often required, because when policy in country A partly depends on policy in country B, then policy in country B is also likely to depend on policy in country A.

è textbook endogeneity

Monte Carlo Results (Franzese and Hays PA 2007)

Table 1 Average coefficient estimates across 1000 trials (bias)

			Coefficient	*OLS*	*S-OLS*	*S-2SLS*	*S-ML*
$\rho = .1$	$N = 5$	$T = 20$	$\hat{\beta}_s$	*1.112*	1.027	***1.003***	1.048
			$\hat{\rho}$	—	0.078	***0.097***	*0.063*
		$T = 40$	$\hat{\beta}_s$	*1.112*	0.991	***1.001***	1.021
			$\hat{\rho}$	—	0.108	***0.099***	*0.082*
	$N = 40$	$T = 20$	$\hat{\beta}_s$	*1.112*	*1.049*	***0.994***	*1.050*
			$\hat{\rho}$	—	*0.055*	***0.105***	*0.054*
		$T = 40$	$\hat{\beta}_s$	*1.112*	***0.999***	1.003	1.026
			$\hat{\rho}$	—	***0.101***	.098	*0.077*
$\rho = .5$	$N = 5$	$T = 20$	$\hat{\beta}_s$	*1.999*	*0.837*	***0.998***	*1.050*
			$\hat{\rho}$	—	*0.579*	***0.499***	*0.475*
		$T = 40$	$\hat{\beta}_s$	*2.001*	*0.826*	***1.000***	1.029
			$\hat{\rho}$	—	*0.587*	***0.500***	0.487
	$N = 40$	$T = 20$	$\hat{\beta}_s$	*2.004*	*0.861*	***1.008***	*1.050*
			$\hat{\rho}$	—	*0.570*	***0.497***	*0.474*
		$T = 40$	$\hat{\beta}_s$	*2.000*	*0.844*	***1.002***	1.025
			$\hat{\rho}$	—	*0.578*	***0.499***	0.487
Number of unbiased wins/ties				0	2	14	0
Number of times noticeable ($\geq$5%) bias				16	10	0	10

Note. Unbiasedness winner in bold italics and notable or appreciable biases in italics.

Textbook Endogeneity in Spatial Econometrics

$$y_{it} = \alpha + \gamma y_{jt} + \beta_k \sum_{k=1}^{K} x_{kit} + \varepsilon_{it} \qquad (1)$$

$$y_{jt} = \alpha + \gamma y_{it} + \beta_k \sum_{k=1}^{K} x_{kjt} + \varepsilon_{jt} \qquad (2)$$

$$\rightarrow \quad E\left(\varepsilon_{it} \middle| y_{jt}\right) \neq 0 \qquad (3)$$

where $i \neq j$

When does Spatial Dependence NOT cause Endogeneity?

- in spatial-x models
- unclear in spatial-e models (if error in one country results from error processes in other countries, then endogeneity, instrumenting predicted residuals sounds strange, though)
- in spatial-y models, when there are senders s and receivers r where $s \neq r$.
- temporally lagging the spatial lag (???) the poor man's instrument

Examples:

effect of civil war on economic growth in neighboring country (Murdoch and Sandler, JCR 2002)

monetary policy spill-over from key currencies to trading partners (Plümper and Troeger, AJPS 2008)

A Note on the Replication Study

Hays: World Politics 2002
Hays: Book manuscript 2008

tax competition

Hays explains this observation by arguing that a country's tax mix is crucially determined by its political institutions – with consensus democracies having lower tax rates than majoritarian democracies because their political institutions 'constrain political majorities from choosing transfer maximizing capital tax rates' (Hays 2008: 136).

Spatial Lag Specification

Yet, the theory does not concern us much. What matters most is that Hays specifies the spatial lag as

$$sl_i = \left[\sum_{j \neq i}^{J} contiguity_{ij}\right]^{-1} \left(contiguity_{ij} \cdot tax_j\right)$$

- effective capital tax rate in –i countries
- weighted by contiguity
- row standardized

	model 1 replication S-ML	model 2 S-OLS with robust s.e.
temporal lag	0.772	0.771
	(0.025) ***	(0.034) ***
spatial lag	0.040	0.047
	(0.010) ***	(0.026) *
capital mobility	0.088	0.088
	(0.038) *	(0.035) *
union density	0.037	0.037
	(0.059)	(0.053)
left government	-0.018	-0.018
	(0.019)	(0.025)
european union	6.670	6.613
	(2.723) *	(3.214)
capital mobility interacted with		
capital endowment	-0.004	-0.004
	(0.001) **	(0.001) ***
consensus democracy	0.016	0.017
	(0.010)	(0.016)
union density	-0.000	-0.000
	(0.000)	(0.001)
left government	0.000	0.000
	(0.000)	(0.000)
european union	-0.074	-0.074
	(0.030) *	(0.035) *
unit fixed effects	yes	yes
period fixed effects	yes	yes
W row-standardized	yes	yes
weight	contiguity	contiguity
Nobs	581	581

Note: results reported in Hays (forthcoming) are not exactly replicable due to minor changes in the data structure. * statistically significant at .1 level ** at .01 level *** at .001 level

Specification Issues in the Analysis of Spatial Dependence

Dynamics, Common Trends, and Common Shocks

Spatial Clustering and Unit Fixed Effects

Row Standardization and the Weighting Matrix

Functional Form of the Weighting Matrix

Dynamics, Common Trends, and Common Shocks

Not controlling for common trends and common shocks is likely to bias the coefficient of the spatial lag.

Common trends exert an upward bias, common shocks exert an upward bias unless governments respond differently to common shocks.

Example: Governments command over different tax instruments (capital taxation, labor taxes, VAT). If governments respond differently to a common revenue shock, the spatial lag can be downward biased if the common shock is not accounted for.

Common Trends in Capital Taxation

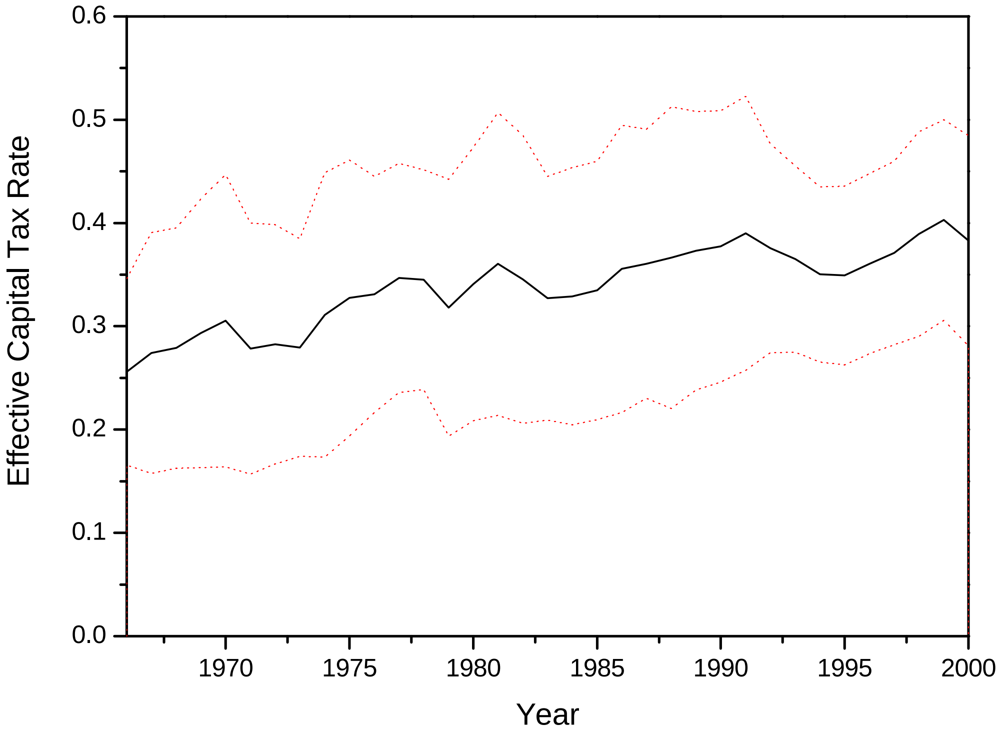


Table 2: Different Treatments for Common Trends and Common Shocks.

	model 3 ldv excluded	model 4 period fe excluded	model 5 ldv and period fe ex-cluded
temporal lag		0.776	
		(0.023) ***	
spatial lag	0.124	0.078	0.257
	(0.016) ***	(0.012) ***	(0.039) ***
unit fixed effects	yes	yes	yes
period fixed effects	yes	no	no
W row-standardized	yes	yes	yes
weight	contiguity	contiguity	contiguity
Nobs	581	581	581

Note: all models include the full battery of control variables reported in table 1.

Results

The size of the spatial lag coefficient varies largely with the treatment of dynamics and common trends.

Controlling common trends away?

The results reported so far demonstrate the importance of controlling for common shocks and common trends, especially when the data is so obviously trended as it is for capital taxation.

Importantly, for capital taxation we are on safe grounds arguing that the common trend is not caused by the spatial effect, because according to all theories, tax competition should not lead to the common increase in capital taxation, which can be observed in the data, but to a decrease instead.

If, however, the common trend is partly due to the spatial lag, then inclusion of period fixed effects or lagged dependent variable can downward bias the coefficient of the spatial lag.

Spatial Clustering

The odds are that contiguous or geographically close political units are more similar than more distant units.

cultures

customs

preferences

constitutions

institutions

are spatially clustered.

These clustering effects are likely to be captured by a spatial lag.

no easy solution

unit fixed effects eliminate spatial clustering, but they transform the estimation equation to

$$y_{it} - \bar{y}_i = \alpha + \rho\left(\mathbf{W}y_{-it} - \overline{\mathbf{W}y_i}\right) + \ldots + \varepsilon_{it} - \bar{\varepsilon}_i \quad \forall\ i \neq -i \quad . \quad (4)$$

where

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \quad \overline{\mathbf{W}y_i} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{W}y_{-it} \quad (5)$$

and likewise for all control variables and the stochastic error.

In many cases the advantages of the spatial FE model will outweigh the disadvantages.

Table 3: Excluding Unit Fixed Effects.

	model 1 (repeated)	model 6 no fixed effects
temporal lag	0.772	0.945
	(0.025) ***	(0.014) ***
spatial lag	0.040	-0.018
	(0.010) ***	(0.011)
unit fixed effects	yes	no
period fixed effects	yes	yes
W row-standardized	yes	yes
weight	contiguity	contiguity
Nobs	581	581

Note: all models include the full battery of control variables reported in table 1.

contrary to expectations?

Interpretation: governments respond to changes in neighboring countries capital tax rates, but obviously not to differences in levels.

This implies that the US tax reforms in 83-85 rather than the abolition of capital controls triggered tax competition.

Row-Standardization of the Weighting Matrix

Gleditsch and Ward (2008): Row standardization has the advantage that the spatial lag will have the same potential metric or units' as the dependent variable itself."

Spatial econometricians typically treat row-standardization as something that is
'commonly' (Franzese and Hays 2006: 174; Franzese and Hays 2008: 29),
'generally' (Darmofal 2006: 8),
'typically' (Anselin 2002: 257) or
'usually' (Beck et al. 2006: 28)
done.

Hmm…

However, row standardization needs theoretical justification.

Row Standardization matters…

… because it changes the relative importance of observations.

Take a weighting matrix that measures contiguity. It has cell entries of one for observations that are contiguous, and zero otherwise. If country *i* has two contiguous countries whereas country *j* has six contiguous countries, then both of *i*'s neighbors and all six of *j*'s neighbors exert the same influence each on the spatial lag variable. After row-standardization, however, the two neighbors of *i* now exert an influence on the spatial lag that is three times larger than the influence of the six neighbors of *j.*

Table 4: Weighting Matrix Not Row-Standardized.

	model 1 (repeated)	model 7 not row-standardized
temporal lag	0.772	0.762
	(0.025) ***	(0.025) ***
spatial lag	0.040	0.078
	(0.010) ***	(0.011) ***
unit fixed effects	yes	yes
period fixed effects	yes	yes
W row-standardized	yes	no
weight	contiguity	contiguity
Nobs	581	581

Note: all models include the full battery of control variables reported in table 1.

Results: Row standardization cannot be justified by common practice or convenience. Its effect on results is significant and possibly important.

Functional Form of Weighting Matrix

What weighting matrices are feasible?

popular:

contiguity

distance

“Space is more than geography.” Beck et al. ISQ 2006

less popular:

trade relations

capital flows

cultural proximity

Example (Beck et al.)

TABLE 1. Democracy and Social Requisites, 1998

		Spatial Autoregressive Estimates		
Variable	*OLS (1)*	*(2)*	*(3)*	*(4)*
Constant	– 19.71 (3.66)	– 12.96 (3.35)	– 19.39 (3.39)	– 13.24 (3.11)
Ln(GDPPC)	2.66 (0.42)	1.72 (0.41)	2.22 (0.41)	1.53 (0.37)
κ_1 (distance)		0.48 (0.09)		0.89 (0.19)
κ_2 (trade)			0.51 (0.14)	0.59 (0.43)
N	170	170	170	170

Standard errors in parentheses.

Correctly specified?
Trade is a function of distance. (See gravity models)

More importantly: The functional form matters as well…

Table 5: Different Functional Forms of the Weighting Matrix.

	model 8 1/(distance)	model 9 1/ln(distance)
temporal lag	0.730	0.808
	(0.034) ***	(0.031) ***
spatial lag	3.392	-0.181
	(1.606) *	(0.050) ***
unit fixed effects	yes	yes
period fixed effects	yes	yes
W row-standardized	no	no
weight	d^{-1}	$(\ln d)^{-1}$
Nobs	581	581

Note: all models include the full battery of control variables reported in table 1.

Interpretation

The functional form of the weighting matrix can determine the results and thus inferences.

Yet, there hardly ever is a theoretical justification of the weighting matrix, leave alone the functional form.

Anything goes?

Perhaps even worse…

	model 8 repeated 1/(distance)	model 10 1/(distance reversed)
temporal lag	0.730	0.775
	(0.034) ***	(0.025) ***
spatial lag	3.392	1.354
	(1.606) *	(0.998)
unit fixed effects	yes	yes
period fixed effects	yes	yes
W row-standardized	no	yes
weight	d^{-1}	p^{-1}
Nobs	581	581

Note: all models include the full battery of control variables reported in table 1.

Intuitively, the coefficients should point in opposite directions.

But they do not, which implies that the coefficient of the spatial lag cannot be interpreted as support of the weighting matrix (perhaps not even as evidence for spatial dependence).

two potential remedies:

– better specified theory

– robustness test (natural log and squared term?)

However: what do we learn about the weighting matrix if we find robust results? That it does not matter?

Conclusion

It seems fair to say that specification issues are much more important as the choice between S-OLS, S-2SLS and S-ML.

Unfortunately, theory often offers not enough guidance for a sufficiently specified spatial model.

Period dummies, some form of temporal error correction, perhaps unit dummies seem straightforward remedies for the problems of common trends and spatial clustering.

Yet, solutions to these problems throw away more variation than is likely to be generated by these processes.

If the cure is worse than the disease, however, it is certainly not worth it.

The choice of the functional form of the weighting matrix seems even more important.

Without theoretical guidance, researchers are actually likely to have control over the estimation result.

In the absence of better specified theories, only robustness tests can help.

Analysis of Spatial Lags in (directed) Dyadic Data

Thomas Plümper and Eric Neumayer

University of Essex
Department of Government

What is (directed) dyadic data?

dyadic data is relational

- wars between countries
- marriage
- treaties
- and so on

directed data distinguishes between sender and receiver

sender: the source, origin, giver, aggressor and so on

receiver: the target, destination, recipient, taker, victim and so on

- terrorist attacks
- coercive treaties/ treaties between unequal
- economic exchange between seller and buyer, exporter and importer

Is modeling spatial lags in dyadic data different?

No, if you just look at the estimation.

Yes, if you look at specification.

The specification of spatial lags in (directed) dyadic data is more flexible.

type of contagion	*(partial) model*	*specification of weighting matrix*
monadic data:		
unit contagion	$y_i = \rho \mathbf{W} y_{-i} + \varepsilon_i$	$\mathbf{W} = w_{i-i}$
undirected dyads:		
unit contagion	$y_{ij} = \rho \mathbf{W} y_k + \varepsilon_{ij} \Leftrightarrow y_{ji} = \rho \mathbf{W} y_k + \varepsilon_{ji}$	$\mathbf{W} = \begin{cases} w_{ik} \\ w_{jk} \end{cases}$
undirected dyad contagion	$y_{ij} = \rho \mathbf{W} y_{-i-j} + \varepsilon_{ij} \Leftrightarrow y_{ji} = \rho \mathbf{W} y_{-j-i} + \varepsilon_{ji}$	$\mathbf{W} = \begin{cases} w_{ik} \\ w_{jk} \\ f\left(w_{ik}; w_{jk}\right) \\ w_{(ij)(-i-j)} \Leftrightarrow w_{(ji)(-j-i)} \end{cases}$
directed dyads:		
source contagion	$y_{ij} = \rho \mathbf{W} y_{-i} + \varepsilon_{ij}$	$\mathbf{W} = \begin{cases} w_{i-i} \\ w_{(ij)(-ij)} \end{cases}$
target contagion	$y_{ij} = \rho \mathbf{W} y_{-j} + \varepsilon_{ij}$	$\mathbf{W} = \begin{cases} w_{j-j} \\ w_{(ij)(i-j)} \end{cases}$
source-to-dyad contagion	$y_{ij} = \rho \mathbf{W} y_{-ij} + \varepsilon_{ij}$	$\mathbf{W} = \begin{cases} w_{i-i} \\ w_{(ij)(-ij)} \end{cases}$
target-to-dyad contagion	$y_{ij} = \rho \mathbf{W} y_{i-j} + \varepsilon_{ij}$	$\mathbf{W} = \begin{cases} w_{j-j} \\ w_{(ij)(i-j)} \end{cases}$
directed dyad contagion	$y_{ij} = \rho \mathbf{W} y_{-i-j} + \varepsilon_{ij}$	$\mathbf{W} = \begin{cases} \textit{function of any combination of} \\ \textit{the above weighting matrices} \\ w_{(ij)(-i-j)} \end{cases}$

Monadic Data

y(-i)

W(i-i)

Undirected Dyads

unit contagion

y(k)

w(ik), w(jk)

where k≠i,j

undirected dyad contagion

y(-i-j)

w(ik), w(jk), f(wik,wjk), w(ij)(-i-j)

Directed Dyads

source contagion

y(-i)

w(i-i), w(ij)(-i-j)

target contagion

y(-j)

w(j-j), w(ij)(-i-j)

source-to-dyad contagion

y(-ij)

w(i-i), w((ij)(-i-j)

target-to-dyad contagion

y(i-j)

w(j-j), w(ij)(-i-j)

directed-dyad contagion

y(-i-j)

highly flexible (any of the above)

OK, specification of spatial lags in dyadic data is more flexible

… but does it matter?

Depends what you mean by matter!

We learn more – if learning matters…

Application: Spatial Dependence in a Directed Country Dyad Sample of BIT Diffusion

Replication of Elkins et al. 2006

Spread of Bilateral Investment Treaties

theory: different variants of competition for FDI

	model 1 target contagion	model 2 source contagion	model 3 target-to-dyad contagion
$w_{j-j}\sum_{s=1}^{t-1} y_{-js}$	1.006 (0.003) *		
$w_{i-i}\sum_{s=1}^{t-1} y_{-is}$		1.002 (0.003)	
$w_{j-j}y_{i-jt-1}$			1.062 (0.003) ***
$w_{i-i}y_{-ijt-1}$			
$w_{i-i}w_{j-j}y_{-i-jt-1}$			
host extractive industries /exports	0.998 (0.002)	0.998 (0.002)	0.997 (0.002) *
common law	0.706 (0.076) **	0.699 (0.076) **	0.671 (0.072) ***
IMF credit (dummy)	1.498 (0.170) ***	1.509 (0.171) ***	1.223 (0.063) ***
GDP (host) ln	1.258 (0.065) ***	1.258 (0.064) ***	1.453 (0.165) **
per capita income (host)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
GDP growth (host)	1.045 (0.011) ***	1.044 (0.010) ***	1.047 (0.011) ***
FDI inflow	1.025 (0.016)	1.030 (0.016) *	1.028 (0.016) *
capital account (host) (% of GDP)	3.388 (1.882) *	3.291 (1.818) *	2.648 (1.479) *
level of democracy	1.010 (0.008)	1.009 (0.008)	1.008 (0.008)
diplomatic repre-sentation (host)	1.007 (0.003) *	1.007 (0.003) *	1.010 (0.003) **
bilateral trade to GDP	21.946 (33.949) *	18.879 (29.387) *	6.854 (11.803)
colonial ties	2.868 (0.685) ***	2.873 (0.688) ***	1.769 (0.436) *
common language	0.892 (0.167)	0.898 (0.168)	1.030 (0.197)
N | Nobs	2400 | 38291	2400 | 38291	2400 | 38291
chi²	228.98 ***	224.69 ***	513.92 ***
-ll	3611.3	3613.4	3468.8

	model 4 source-to-dyad contagion	model 5 directed dyad contagion	model 6 multiple forms of contagion
$w_{j-j}\sum_{s=1}^{t-1} y_{-js}$			1.073 (0.020) ***
$w_{i-i}\sum_{s=1}^{t-1} y_{-is}$			0.921 (0.019) ***
$w_{j-j}y_{i-jt-1}$			1.058 (0.004) ***
$w_{i-i}y_{-ijt-1}$	1.029 (0.002) ***		1.032 (0.003) ***
$w_{i-i}w_{j-j}y_{-i-jt-1}$		1.005 (0.004)	1.008 (0.006)
host extractive industries /exports	0.998 (0.002)	0.998 (0.002)	0.997 (0.002) *
common law	0.776 (0.084) **	0.705 (0.076) **	0.711 (0.077) **
IMF credit (dummy)	1.238 (0.142) *	1.502 (0.170) ***	1.197 (0.138)
GDP (host) ln	1.191 (0.063) **	1.250 (0.064) ***	1.163 (0.062) **
per capita income (host)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
GDP growth (host)	1.033 (0.010) **	1.044 (0.010) ***	1.035 (0.011) **
FDI inflow	1.021 (0.016)	1.027 (0.016) *	1.026 (0.016) *
capital account (host) (% of GDP)	2.989 (1.633) *	3.301 (1.818) *	2.262 (1.262)
level of democracy	1.007 (0.008)	1.009 (0.008)	1.008 (0.008)
diplomatic repre-sentation (host)	1.006 (0.003) *	1.007 (0.003) *	1.010 (0.003) **
bilateral trade to GDP	21.761 (33.819) *	18.473 (28.499) *	3.325 (5.953)
colonial ties	2.912 (0.696) ***	2.869 (0.686) ***	1.692 (0.422) *
common language	0.921 (0.173)	0.909 (0.171)	1.112 (0.215)
N | Nobs	2400 | 38291	2400 | 38291	2400 | 38291
chi^2	340.86 ***	226.37 ***	661.45 ***
-ll	3555.3	3612.6	3395.0

What do we find?

As Elkins et al.,: similarity between capital importers

increases the probability of a BIT between an capital importer and a capital exporter if the similar capital importer has also signed BITs

In addition, we find:

similarity between capital importers
increases the probability of a BIT between an capital importer and a specific capital exporter if the similar capital importer has also signed a BIT with the very same capital exporter

similarity between capital exporters
increases the probability of a BIT between an capital exporter and a specific capital importer if the similar capital exporter has also signed a BIT with the very same capital importer

similarity between dyads
two countries are more likely to sign a BIT if a similar dyad has already signed one

New Venues for Research

true, this is not the philosopher's stone,

but we offer more flexible ways of testing theories (which may also lead to the formulation of novel theories).