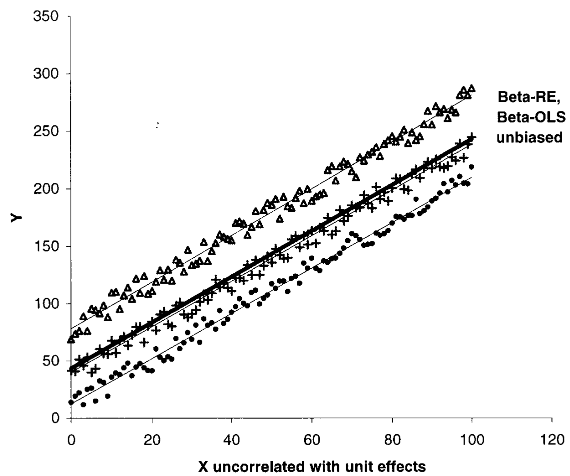
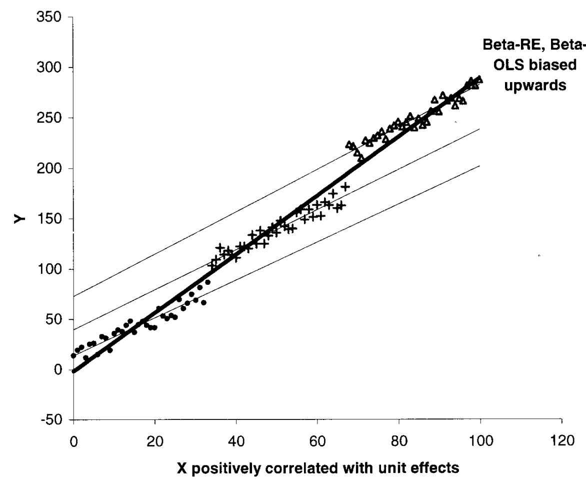
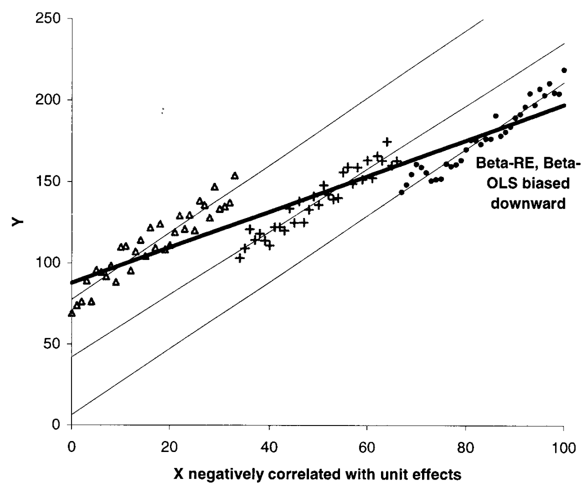
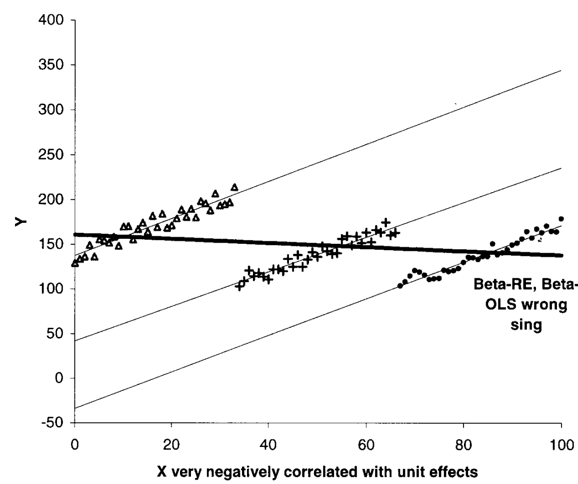

350
300
250
200
150
100
50
0
Y
0
20
40
60
80
100
120
X uncorrelated with unit effects
Beta-RE,
Beta-OLS
unbiased


350
300
250
200
150
100
50
0
-50
Y
0
20
40
60
80
100
120
X positively correlated with unit effects
Beta-RE, Beta-
OLS biased
upwards

(i)



Y
250
200
150
100
50
0
0
20
40
60
80
100
Beta-RE, Beta-OLS biased downward
X negatively correlated with unit effects



Y
400
350
300
250
200
150
100
50
0
-50
0
20
40
60
80
100
Beta-RE, Beta-OLS wrong sing
X very negatively correlated with unit effects

Some Notes on Bias (based on Doug Rivers MS)

(simplest case: single covariate, balanced data, one-way (space) effects)

- Model

  $y_{it} = \mu_i + \beta x_{it} + \varepsilon_{it} \qquad (i=1,\ldots N;\ t=1,\ldots T)$

- ε_{it} assumed IID $N(0, \sigma^2)$
- x_{it} a non-stochastic sequence, independent of ε_{it}

- OLS ("Total" estimator in old, ANOVA terminology)
  - regress y_{it} on x_{it} and an intercept

$$\hat{\beta} = \frac{\sum_{i,t} (x_{it} - \bar{x}) y_{it}}{\sum_{i,t} (x_{it} - \bar{x})^2}$$

Fixed Effects Estimation ("Within" or LSDV estimator)

- Introducing N unit dummies for varying intercepts, run OLS

or

- Compute unit means: $\bar{y}_{i\cdot} = \sum_i y_{it} / T$, $\bar{x}_{i\cdot} = \sum_i x_{it} / T$, etc.

- Within transformation: $(y_{it} - \bar{y}_{i\cdot}) = \beta(x_{it} - \bar{x}_{i\cdot}) + (\varepsilon_{it} - \bar{\varepsilon}_{i\cdot})$

- $\tilde{\beta}$ obtained from OLS on transformed equation or LSDV specification

Random Effects: FGLS dominates OLS under conditions when OLS assumptions are met, but it is simpler to contrast FE and OLS, since RE shares OLS bias

ANOVA Decomposition

$TSS_X = WNSS_X + BNSS_X$

$$T_{XX} = \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_{..})^2$$

$$WN_{XX} = \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_{i.})^2$$

$$BN_{XX} = T\sum_{i=1}^{n}(\bar{x}_{i.} - \bar{x}_{..})^2$$

$\lambda = BN_{XX}/T_{XX}$, the proportion of the variance in x_{it} that is between the units in the ANOVA decomposition

Bias Results

- Under our usual assumptions, FE is unbiased
- Unobserved heterogeneity will cause OLS to be biased

$$E(\hat{\beta}) = \beta + \lambda b$$

where b is the coefficient of an auxiliary regression of μ_i on x_{it} means

$$b = \frac{\sum_i (\bar{x}_i. - \bar{x}..)\mu_i}{\sum_i (\bar{x}_i. - \bar{x}..)^2}$$

- version of Theil-Griliches left-out-variables (LOV) formula

Bias of OLS:

$$E(\hat{\beta}) - \beta = \lambda b$$

- Two factors determine bias of OLS
  1. How strongly $\bar{x}_{i\cdot}$ is correlated with unit effects (measured by b)
  2. How much of the variance is between the units (vs. within) (measured by λ)

- OLS performs well enough if *either* b or λ is small

Variance of OLS and FE

- OLS: $V(\hat{\beta}) = \sigma^2 / T_{XX}$
- FE: $V(\tilde{\beta}) = \sigma^2 / WN_{XX}$
- Relative efficiency

$$\frac{SD(\tilde{\beta})}{SD(\hat{\beta})} = \sqrt{\frac{\sigma^2 / WN_{XX}}{\sigma^2 / T_{XX}}} = \frac{1}{\sqrt{1-\lambda}}$$

- FE is very inefficient when λ is large (and impossible in the limit, when λ=1). FE has no advantages over OLS (or RE) when λ=0 (and little advantage when λ is small), because bias is small.

Mean Square Error Comparison

- Use MSE to choose optimal bias-variance tradeoff
- FE is unbiased so

  $MSE(\tilde{\beta}) = V(\tilde{\beta}) = \sigma^2 / WN_{XX}$

- For OLS

  $MSE(\hat{\beta}) = V(\hat{\beta}) + [bias(\hat{\beta})]^2 = \sigma^2 / T_{XX} + \lambda^2 b^2$

- FE has smaller MSE than OLS iff

  $\sigma^2 / WN_{XX} \leq \sigma^2 / T_{XX} + \lambda^2 b^2$ or, equivalently,

  $$|b| \geq \frac{\sigma}{\sqrt{\lambda(1-\lambda)T_{XX}}}$$

- Conclusion: should use FE for moderate λ

Rivers generalizes this result

- for multiple covariates (just algebra)
- for covariates which are time-invariant (and can thus be used to model unobserved heterogeneity, potentially offsetting advantage of FE)

$\Lambda = T_{XX|Z}^{-1} B_{XX|Z}$ analogue to λ

comparison of FE and OLS (or RE) is now comparison of vectors and associated covariance matrices

consider estimable functions of the form $\theta = c'\beta$

OSL: $MSE(\hat{\theta}) = \sigma^2 c'(T_{XX|Z}^{-1} + \Lambda b b' \Lambda')c$

FE: $MSE(\tilde{\theta}) = \sigma^2 c' W_{XX}^{-1} c$

prefer FE whenever $\sigma^2 \Lambda^{-1} T_{XX|Z}^{-1} (I_k - \Lambda)^{-1} - bb'$ is p.d.

APPLICATION EXAMPLE: "GRAVITY MODEL" OF TRADE

T_{jkt} trade between countries j and k in period t
GDP_{jt} real GDP in country *j* in period *t*
d_{jk} distance between countries j and k

by analogy with physics $Force = G\frac{Mass_1 Mass_2}{\text{distance}^2}$

$$T_{jkt} = (GDP_{jt} \times GDP_{kt}) / d_{jk}^2$$

Empirical Implementation

$$\log(T_{jkt}) = \beta \log(GDP_{jt} \times GDP_{kt}) + \gamma d_{jk} + \text{error}$$

or $$y_{it} = \beta x_{it} + \gamma z_i + \varepsilon_{it}$$

Andrew Rose, "Do We Really Know that the WTO Increases International Trade?" *AER* (March 2004)

- Trade data from IMF Direction of Trade Statistics (DOTS)
  1. *very* large data set with about 200 countries over about 50 years (1948-present)
  2. each country's reports of imports and exports with other countries in data set
  3. over 1 million dyad-years
- controls for a large number of other factors (language, colonial relationship, etc.)
- adds dummies for whether countries in a dyad were members of the GATT/WTO (both, one, neither) and finds **NO** effect (!)
- finds currency unions and generalized system of preferences (GSP) significantly increased trade
- Results *appear* robust to choice of sample period and estimation technique

Rose: Table 1—Benchmark Results (excerpt)

	Default (OLS†)	…	With Country Effects (FE)
Both in GATT/WTO	-0.04	…	0.15
	(0.05)		(0.05)
One in GATT/WTO	-0.06	…	0.05
	(0.05)		(0.04)
GSP	0.86	…	0.70
	(0.03)		(0.30)
Log distance	-1.12		-1.31
	(0.02)		(0.02)
Log product real GDP	0.92	…	0.16
	(0.01)		(0.05)
…	…	…	…

† with clustered standard errors, time FE
(roughly 250,000 observations)

Tomz, Goldstein, Rivers "Membership Has Its Privileges"

- data in DOTS is full of mistakes, inconsistencies

- Rose treats colonies, provisional and de facto members same as non-members
  -when GATT originally formed, countries had right to extend benefits to colonies and done by GB, France
  -independent ex-colonies went through process of joining, so there are stages of membership that Rose mis-coded

- Expect hierarchical effects, e.g. GATT effect weakest for countries that already have a free-trade agreement

1. ***Which*** **unit effects should we include?**

Egger & Pfaffermayr (2002)

country effects, dyad effects, time effects

$$y_{ijt} = \beta_0 + x_{ijt}'\beta + \mu_i + \gamma_j + \eta_t + \delta_{ij} + \varepsilon_{ijt}$$

with usual identification restrictions $\Sigma_i\mu_i = \Sigma_i\gamma_i = \Sigma_i\eta_i = \Sigma_{i,j}\delta_{ij} = 0$

Baltagi, Egger, & Pfaffermay (2003) add country-time interaction effects

$$y_{ijt} = x_{ijt}'\delta + \alpha_i + \beta_j + \gamma_t + (\alpha\beta)_{ij} + (\alpha\gamma)_{it} + (\beta\gamma)_{jt} + \varepsilon_{ijt}$$

2. Average imports and exports or use "directional" dyads ($y_{ijt} \neq y_{jit}$)?

TGR (2005): Table 2—Effect of GATT (excerpt)

		(OLS)	…	with country & dyad FE
Both in GATT	as formal members	-0.17 (0.03)	…	0.48 (0.06)
	as nonmember participants			0.88 (0.09)
	as formal + nonmember			0.56 (0.06)
One in GATT	formal member	-0.27 (0.04)		0.23 (0.06)
	nonmember participant			0.34 (0.07)
	GSP	0.86 (0.03)	…	0.18 (0.03)

With FE, you always got larger magnitudes than OLS – which should you believe?

- GATT participation dummies have λ =0.6
- Λ has similar values down the diagonal (roughly 0.4 to 0.7) *after* controlling for time-invariant variations
- Therefore the data favour FE not OLS or RE, and the impact of GATT is probably smallish (per TGR), but **not** zero (per Rose)

NOTES ON GRAVITY-MODEL FIXED EFFECTS

01.12.05 (BJG)

I've had several queries about the fixed effects in the gravity trade models as discussed in last week's articles. There's some confusion about how and when time-invariant variables get "swept out" by fixed effects, and I undoubtedly went over this point a bit too quickly and abruptly in class, since the dyad data structure complicates matters quite a bit as opposed to the more standard structure we've dealt with heretofore wherein we have simple units × time-waves.

One issue is whether we form one observation on the dependent variable for each dyad or whether we create "directional" dyads such that US-Japan-1980 and Japan-US-1980 are different observations. To begin, assume that we form dyads from a set of countries so that observations are dyad-years without order mattering (i.e. we do **not** have both US-Japan-[year] and Japan-US-[year] observations). Suppose that we have 3 countries, A, B, and C, and 3 times periods, 1, 2, and 3, and we have GDP (that varies by country and year) and distance between countries (that does not vary by year). The country and year fixed effects could be implemented with indicator variables, and to avoid the "dummy variable trap," we could omit one country and one year, or suppress the overall constant and omit one year (or one country), or impose summing constraints on the data matrix and estimate the constrained regression. But the key point is that those country and year fixed effects do not induce linear dependence with time-invariant distance variable (except in the case of a fluke such as x = y = z). Likewise, the GDP product is distinct across all observations except by coincidence.

obs.	A	B	C	1	2	3	GDPi	GDPj	G×G	distance
ab1	1	1	0	1	0	0	h	k	hk	x
ab2	1	1	0	0	1	0	i	l	il	x
ab3	1	1	0	0	0	1	j	m	jm	x
ac1	1	0	1	1	0	0	h	n	hn	y
ac2	1	0	1	0	1	0	i	o	io	y
ac3	1	0	1	0	0	1	j	p	jp	y
bc1	0	1	1	1	0	0	k	n	kn	z
bc2	0	1	1	0	1	0	l	o	lo	z
bc3	0	1	1	0	0	1	m	p	mp	z

On the other hand, we cannot estimate country-, time-, and dyad- fixed effects, because we run out of degrees of freedom. We have 9 data points, 2 covariates (G×G and distance) and 5 dummies already, so we don't have the degrees of freedom left to add 2 more dummies. We could, however, have dyad dummies if we omit the country dummies. Notice, however, that these are linearly dependent with distance, since distance is merely a constant (x, y, or z) multiplied by the dyad dummy. So if we implement dyad fixed effects with a "within transformation," we sweep out the time-invariant variables in doing so. Likewise, if we estimate the model with OLS and dyad dummies, we must omit the distance variable.

obs.	AB	AC	BC	1	2	3	GDPi	GDPj	G×G	distance
ab1	1	0	0	1	0	0	h	k	hk	x
ab2	1	0	0	0	1	0	i	l	il	x
ab3	1	0	0	0	0	1	j	m	jm	x
ac1	0	1	0	1	0	0	h	n	hn	y
ac2	0	1	0	0	1	0	i	o	io	y
ac3	0	1	0	0	0	1	j	p	jp	y
bc1	0	0	1	1	0	0	k	n	kn	z
bc2	0	0	1	0	1	0	l	o	lo	z
bc3	0	0	1	0	0	1	m	p	mp	z

When can we include country and time and dyad fixed effects? Dyads are formed as binomial coefficients, so the number of observations (dyad-years) increases faster than the number of parameters as our data set grows in number of countries. If you have 10 countries, that's 45 dyads, and you have 45T observations. You consume 44 (or 45) d.f. with dyad indicators, 9 (or 10) with country indicators, and T-1 or (T) with time-period indicators, but as long as T is large enough that 44T -53 > 0, you don't run out of degrees of freedom. To follow Baltagi, Egger, and Pfaffermayr (*Economics Letters* 2003) and implement the full FE model, you also include interactions, i.e. you have country indicators, time-period indicators, dyad indicators, and country-time-period indicators. Their equation (1) on p. 393 includes both α_i and β_j and both $(\alpha\gamma)_{it}$ and $(\beta\gamma)_{jt}$ so we want indicators for both importer and exporter and for both time interactions. To make this concrete, consider N=4 and T=5. There are 6 dyads, so 30 data points, and 6 dyad indicators, 4 country indicators, 5 time-period indicators, and 20 country-time indicators. Hence, we have more parameters than data and cannot estimate the model. With N=5, however, there are 10 dyads, and thus 50 observations (dyad-years), and we have 10 dyad indicators, 5 country indicators, 5 time indicators, and 25 country-time pairs. Now we have a few more data points than parameters: not a good situation for reliable estimates to be sure, but not a logical impossibility either. Note, though, that distance (constant for a dyad) again cannot be included in the model, since it is completely collinear with the dyad indicators. (Note too that country indicators are not linearly dependent with dyad indicators because the latter cannot be expressed as linear function of *sums* of the former (they're products).)

		5 country indicator var.s					10 dyad indicator variables						5 time indicators				25 country-time interactions				(time-invar.)
no.	obs	A	B	C	D	E	AB	AC	AD	AE	…	DE	1	2	…	5	A1	A2	…	E5	distance
1	ab1	1	1	0	0	0	1	0	0	0		0	1	0		0	1	0		0	x
2	ab2	1	1	0	0	0	1	0	0	0		0	0	1		0	0	1		0	x
3	ab3	1	1	0	0	0	1	0	0	0		0	0	0		0	0	0		0	x
4	ab4	1	1	0	0	0	1	0	0	0		0	0	0		0	0	0		0	x
5	ab5	1	1	0	0	0	1	0	0	0		0	0	0		1	0	0		0	x
6	ac1	1	0	1	0	0	0	1	0	0		0	1	0		0	1	0		0	y
7	ac2	1	0	1	0	0	0	1	0	0		0	0	1		0	0	1		0	y
8	ac3	1	0	1	0	0	0	1	0	0		0	0	0		0	0	0		0	y
9	ac4	1	0	1	0	0	0	1	0	0		0	0	0		0	0	0		0	y
10	ac5	1	0	1	0	0	0	1	0	0		0	0	0		1	0	0		0	y
….	….																				….
21	bc1	0	1	1	0	0	0	0	0	0		0	1	0		0	0	0		0	w
….	…																				….
45	cd5	0	0	1	1	0	0	0	0	0		0	0	0		1	0	0		0	u
46	de1	0	0	0	1	1	0	0	0	0		1	1	0		0	0	0		0	z
47	de2	0	0	0	1	1	0	0	0	0		1	0	1		0	0	0		0	z
48	de3	0	0	0	1	1	0	0	0	0		1	0	0		0	0	0		0	z
49	de4	0	0	0	1	1	0	0	0	0		1	0	0		0	0	0		0	z
50	de5	0	0	0	1	1	0	0	0	0		1	0	0		1	0	0		1	z

In the Rose *AER* piece, he reports, in Table 1, a column that has both country fixed effects and some variables that are dyad-specific and time-invariant, like "log distance" and "common colonizer". That's OK because he's not including dyad fixed effects alongside the time- and country fixed effects in the model. On page 105, in reference to Table 3, he writes, "Cognoscenti may prefer the fixed-effects estimation shown at the right of the table that focus even more exclusively on time-series variation, since any features are constant over time for a pair of countries (such as geography, culture, and history) are taken out." Here, he is referring to the columns marked "Fixed country-pair effects" which (obviously) do contain dyad indicators (unreported in the table).

In the Baltagi, Egger, and Pfaffermayr *Economics Letters* article, by contrast, the only covariates included in the 8 reported models (beyond the large set of fixed effects) are all time-and-dyad-varying. In Egger and Pfaffermayr's *Empirical Economics* piece, by contrast, some models do include time-invariant, dyad-specific variables such as distance. However, you'll note by checking the F-statistics at the bottom of the table that those models are the ones that omit the dyad fixed effects (see the "bilateral interaction" row).

One reason it was a bit hard to figure out exactly what is going on in the empirical sections of the papers we read is that none of them had completely balanced panels. In every case, a number of dyads were discarded because they reported very little trade or had other data problems. Rose mentions millions of observations in the DOTS data, but estimates the models on only about one-quarter to one-third of the data. Egger & Pfaffermayr came closest to having a complete panel. Note that they differ from Rose insofar as their dependent variable is directional, so that, e.g., exports by Australia to Canada in 1982 and exports by Canada to Australia in 1982 are included as distinct observations rather than being summed or averaged. E&P analyze trade amongst 11 nations, which implies 55 unordered country-pairs (dyads) and 110 ordered dyads. They have 17 years of data (1982-1998) so there should $17 \times 110 = 1870$ dyad-years from their main cases. Footnote 7 says they included the EU15 as an importer, but not as an exporter, which should add $11 \times 17 = 187$ more observations, for 2057 total. In the text, they mention omitting 28 outliers and arriving at 2029 observations, so the data are *very* slightly unbalanced because of the asymmetrical treatment of the EU and the few cases deemed to be problematic in terms of influence (probably because of measurement or recording error).