EC9AA: The Practice of Economics Research

14/15/ June 2017: 10am – 1pm

S 2.81

Vera E. Troeger

Office: S0.70

Email: v.e.troeger@warwick.ac.uk

Office hours: Thursday 10am – 12noon

Session 1: Endogeneity, Identification and Instruments

1. Endogeneity and Instruments

2. Fixed Effects and Dynamics

3. Institutional persistency: time invariant and slowly moving variables

1. Endogeneity and Instruments

Everything is endogenous to everything…

Social, political and economic processes are complex – this implies everything depends on everything, our theories and empirical models are plagued by under-determination and causal heterogeneity across units and over time

Endogeneity is ubiquitous: institutions and economic growth, voting and party ID, education and income, wage and performance…

A set of hierarchical solutions to the identification problem proposed by Angrist and Pischke in "Mostly harmless econometrics"

The identification revolution in economics – that has swept over to other social sciences, has understood this very problem.

1. Do an experiment – many social phenomena we want to study do not lend themselves to experimentation (e.g. randomly assign institutional settings or electoral rules to some communities but not others in a polity, corruption, anything related to international relations – terrorism, war; crises etc.)

2. Find a quasi-experimental setting: a natural experiment, a discontinuity (close elections, colonial heritage imposing specific institutional settings)

3. Find exogenous variation, i.e. an instrument (rainfall, pollution, geography)

4. Differences-in-differences-type strategies (e.g. fixed effects) that use repeated observations to control for unobserved omitted factors

Tests and solution to the problem of endogeneity, reversed causality, simultaneity are weak at best and harmful at worst

Endogeneity tests – usually no power against the alternative → exogenous variation usually theoretically justified

Design based approaches → trade-off between internal and external validity → Meta-analyses?

Instruments → trade-off between validity and strength, unbiasedness vs. efficiency

Fixed effects → suffer from dynamic miss-specification

Endogeneity Testing

Most tests for Endogeneity follow one of two underlying principles:

1. Exogenous instrument(s) are available to test for potential endogeneity of a subset of regressors (Wu-Hausman F-Test; Dubin-Wu-Hausman Chi²-Test, Sargan-Hansen etc.)

2. There exists an estimator that is consistent under both the H0 and HA (e.g. FE) that is compared to an estimator only consistent under the alternative (RE), (Hausman-type tests, Mundlak version of the Hausman test, Davidson-MacKinnon)

Both types of principles can be shown to be statistically linked (Kiviet and Pleus 2017)
Kiviet and Pleus (2017) compare a large number of tests for orthogonality and show that:

- All tests underperform when instruments are weak
- Show under what conditions LM type and Wald type tests perform better
- Bootstrapped Wald-type tests and Sargan test behave slightly more favorable

They don't show results for (slightly) invalid instruments, hierarchical (panel, pooled) data sets

Using Instruments

The usual suspects: Wu-Hausman F-Test; Dubin-Wu-Hausman Chi^2-Test

The principle: consider a regression

y = b0 + b1*x1 + b2*x2 + e

where x1 is endogenous. Suppose that z1 is an instrumental variable for x1.

One should decide whether it is necessary to use an instrumental variable, i.e., whether a set of estimates obtained by least squares is consistent or not.

An augmented regression test can easily be formed by including the residuals of each endogenous right-hand side variable, as a function of all exogenous variables, in a regression of the original model.   We would first perform a regression

x1 = c0 + **c1*z1** + c3*x2 + xsi

to get residuals xsi, then perform an augmented regression:

y = d0 + d1*x1 + d2*x2 + **d3*xsi** + ε

If d3 is significantly different from zero, then OLS is not consistent.

Performance of DWH type Tests

MC set up

DGP: y=x1+x2+e: endogenous(x1), instrument(z1), all N~(0,1)

N = (50, 100, 1000)

Endogeneity of x1: corr(x1,e)=(0,0.1,...,0.9)

Strength of instrument z1: corr(z1, x1)=(0,0.1,0.3,0.5,0.7)

Validity of instrument z1: corr(z1, e)=(0,0.1,0.3,0.5,0.7)

The Ratio of DWH Tests Suggesting Endogeneity of x1 with 95% Confidence

validity of instrument $corr(z_1,e)$	strength of instrument $corr(z_1,x_1)$	endogeneiy $corr(x_1,e)$					
		0	0.1	0.3	0.5	0.7	0.9
0	0	0.443	0.428	0.417	0.405	0.361	0.211
	0.1	0.413	0.459	0.433	0.400	0.400	0.386
	0.3	0.440	0.445	0.488	0.574	0.730	0.950
	0.5	0.478	0.470	0.596	0.796	0.978	..
	0.7	0.461	0.523	0.813	0.991	1.000	..
0.1	0	0.492	0.457	0.518	0.472	0.444	0.384
	0.1	0.475	0.478	0.466	0.417	0.360	0.207
	0.3	0.500	0.480	0.423	0.454	0.487	0.715
	0.5	0.537	0.460	0.441	0.612	0.867	1.000
	0.7	0.581	0.505	0.591	0.902	1.000	..
0.3	0	0.776	0.774	0.760	0.802	0.873	0.976
	0.1	0.756	0.756	0.740	0.720	0.763	0.832
	0.3	0.789	0.724	0.646	0.555	0.461	0.247
	0.5	0.866	0.755	0.593	0.432	0.406	0.684
	0.7	0.915	0.842	0.539	0.463	0.848	..
0.5	0	0.963	0.975	0.965	0.980	0.999	..
	0.1	0.977	0.973	0.972	0.970	0.987	1.000
	0.3	0.986	0.966	0.946	0.908	0.863	0.892
	0.5	0.987	0.981	0.924	0.814	0.619	0.310
	0.7	1.000	0.996	0.930	0.712	0.390	0.766
0.7	0	0.997	0.999	1.000	1.000	1.000	..
	0.1	0.999	1.000	0.999	0.999	1.000	..
	0.3	0.999	0.999	0.999	1.000	0.996	..
	0.5	1.000	1.000	0.999	0.987	0.983	0.964
	0.7	1.000	1.000	1.000	0.984	0.898	0.457

DGP: y=x1+x2+e: endogenous(x1), instrument(z1)

Comparison of Estimators

Hausman-Test for RE vs. FE

FE is very inefficient if model includes RHS variables with small within variation
Hence, testing for correlated unit specific effects in order to determine whether a fixed effects specification is actually required is indispensable

The Hausman specification test (Hausman 1978) seems to be the gold standard in social sciences

About 45% of articles employing FE use the Hausman test to justify model choice (this does not mean that the other 55% use other tests, they don't test at all…)

Logic: since the RE estimator is biased if unit specific effects are correlated with any of the RHS variables, differences between FE and RE estimates are interpreted as evidence against the random effects assumption of strict exogeneity

$$\chi^2(df) = \left(\hat{\beta}_{FE} - \hat{\beta}_{RE}\right)' \left[A\text{var}\left(\hat{\beta}_{FE}\right) - A\text{var}\left(\hat{\beta}_{RE}\right)\right]^{-1} \left(\hat{\beta}_{FE} - \hat{\beta}_{RE}\right)$$

Performance of Hausman-Test

Good asymptotic properties and performs well with low frequency of type 1 errors (Baltagi 2001)

Different variants of the Hausman Test that allow for serial correlation, non-stationarity, heteroskedasticity and other violations of GM assumptions.

BUT: Test results are influenced by the trade-off between bias and efficiency
Hausman test is only powerful in the limit, since FE is consistent (in the limit) – difference of RE and FE estimates only result from biased RE estimates

In finite samples: differences can result from two sources: biased RE estimates and unreliable FE estimates (because of inefficient estimation)

Hausman-test mirrors this trade-off: difference of RE and FE estimates / difference in asymptotic variance of RE and FE estimates

→ Test results are especially unreliable if regressors are both correlated with the unit specific effects and rarely changing

Set up of the MC Analysis

DGP:

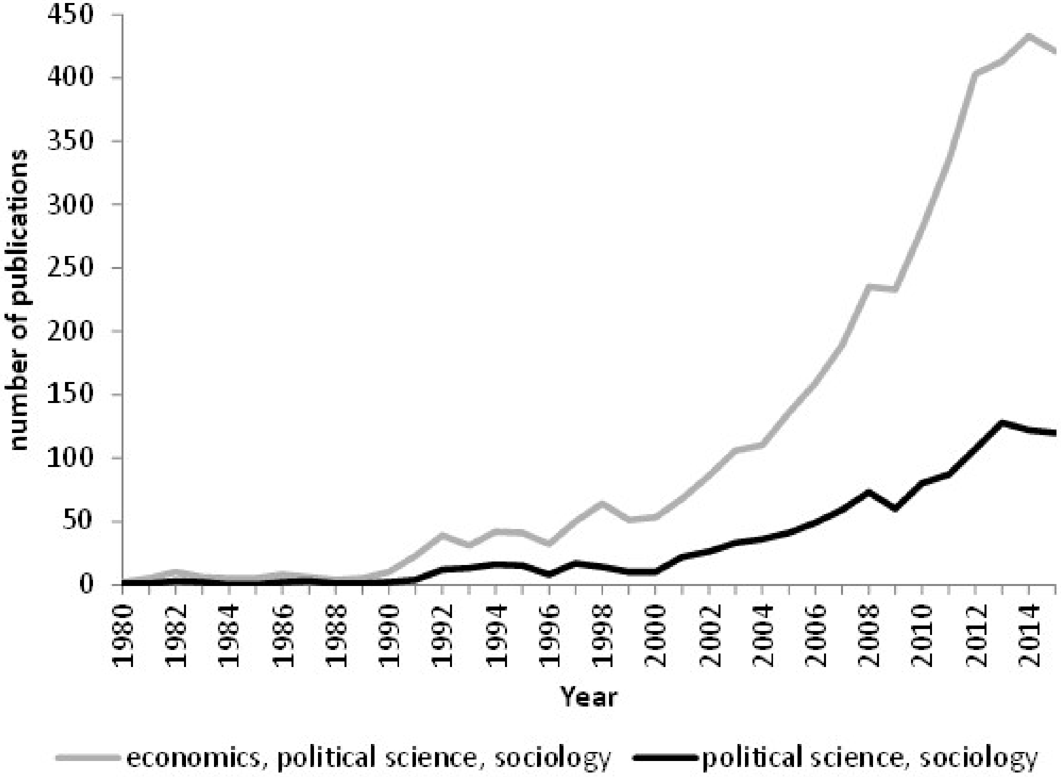
$$y_{it} = \alpha + \beta_k \sum_{k=1}^{K} x_{kit} + u_i + \varepsilon_{it}$$

- unit specific effects N ~ (0, sigma), the size of the unit specific effects: standard deviation = {0,1,3,5},
- the correlation between the RHS variables and the unit specific effects: corr(x,u)={0,0.2,0.4,0.6,0.8},
- efficiency of the fixed effects model: ratio of between to within standard deviation of the RHS variables: b/w(x)={0.15,1,2,3},
- Number of observations: all permutations of N and T = 10, 30, 70.
- Number of repetitions: 1000

N/T	Size b/w rho	0.05 0.2	0.1	0.05 0.4	0.1	0.05 0.6	0.1	0.05 0.8	0.1
10_10	0.15	25	31	26	36	28	40	22	33
	1	6	13	4	11	12	20	8	15
	2	5	7	7	13	7	14	3	7
	3	8	15	7	10	4	6	4	8
10_30	0.15	41	51	61	70	61	78	44	51
	1	4	11	15	25	25	35	11	19
	2	2	11	9	13	14	18	15	25
	3	3	7	4	9	13	18	12	20
10_70	0.15	37	41	100	100	100	100	90	96
	1	5	5	16	27	28	46	49	60
	2	0	3	8	15	17	24	19	32
	3	1	3	7	11	13	19	13	21
30_30	0.15	95	96	100	100	99	99	95	98
	1	6	11	34	46	53	67	58	72
	2	6	15	13	23	30	38	27	36
	3	3	8	14	21	12	19	20	24
30_70	0.15	94	94	100	100	100	100	100	100
	1	3	11	62	78	91	95	90	91
	2	8	13	27	40	47	61	42	62
	3	6	9	19	29	29	35	26	41
70_10	0.15	54	63	94	97	96	99	72	82
	1	6	14	32	48	51	65	47	59
	2	5	11	10	22	17	23	22	34
	3	3	6	7	15	14	21	16	24
70_30	0.15	100	100	100	100	100	100	100	100
	1	16	25	76	86	95	98	84	95
	2	7	13	25	34	56	66	51	60
	3	7	17	15	28	32	39	23	33
70_70	0.15	100	100	100	100	100	100	100	100
	1	23	41	97	99	100	100	100	100
	2	10	24	63	73	83	87	86	93
	3	12	18	36	47	55	70	54	62
Settings: 1 variable, corr (u,x) = {0.2, 0.4, 0.6, 0.8}, SD of unit specific effects = 3, N,T = {10,30,70}									

Instrumental Variables: Improving Strength

Move in social sciences to design based identification strategies
Exponential increase in the use of IV estimation approaches



Caveat: good instruments are notoriously hard to find – trade-off between strength and validity

The Idea

Think of the typical regression setup:

$$Y_i = \alpha + \beta X_i + e_i$$

where $Cov(X_i, e_i) \neq 0$

A potential solution is to use some Z with the following properties:

1. Knowledge of Z helps us predict X (strength)
2. Z exerts no effect on Y other than through its effect on X (validity: ignorability and exclusion)

If (1) & (2) can be established, beta is given by:

$$\widehat{\beta_1} \equiv IV_{Wald} = \frac{Cov(Z_i, Y_i)}{Cov(Z_i, X_i)}$$

The stronger the effect of X on Z, the more precise and least biased our estimates of beta.

- How do we estimate $\widehat{X}_i$?
- With an OLS regression: $X_i = \alpha + \gamma Z_i + u_i$

A linear structure imposed by the first-stage OLS may be a poor approximation of how Z affects X:

1. The effect may hold only locally
2. The relationship may be non-monotonic

Even when there is a non-zero effect of Z on X, if their correlation is small, the IV estimator may still be biased: probability limit of the IV estimator: $\lim_{\widehat{\beta_1}} = \beta_1 + \frac{r_{Z_i Z e_i}}{r_{Z_i X_i}}$
even when Z is only trivially correlated with the unmeasured causes of Y, this correlation can create substantial bias in finite samples, if the correlation between X and Z is also very low.

We can obtain improved $\widehat{X}_i$ based on Z_i without violating exclusion by using a locally weighted regression (lowess) curve or other non-parametric estimation procedures (Kernel smoothing, Kernel Regularized Least Squares):

$$X_i = f(z_i) + v_i$$

Statistical Properties

Why has the 2SLS prevailed over other alternative even when X is binary?

1. It guarantees that $Cov(Z_i,u_i)=0$
2. Logit and probit models borrow identification from the imposed functional form

None of these criticisms apply to the non-parametric first stage! Moreover:

1. Local Linear Regressions (Lowess, Kernel, KRLS) are non-parametric, hence no modeling assumption is needed, so there is no further uncertainty needs to be incorporated

2. The 2SLS is biased when instruments are weak and when there are many instruments. NPFS increases efficiency and thus makes it more likely that one IV will suffice

The uncertainty of the first stage can be easily incorporated at the second stage through bootstrapping:

$SE_{\beta_1} = \sqrt{1/k \sum_{k=1}^{K} SE_{\beta_1}^2 + S_{\beta_1}^2}$, where $S_{\beta_1}^2$ is the sample variance across all $k = 1, \cdots, K$

MC Analyses

DGP:

Stage 2: (we also run MCs with exogenous covariates in Stage 2)

$$y_i = x_i'\beta + \varepsilon_i, \quad i = 1, \ldots, N$$

Stage 1:

$$x_i = z_i'\gamma_1 + z_i^2{}'\gamma_2 + \xi_i, \quad i = 1, \ldots, N$$

Variables are drawn as follows:

$$x_i, z_i, \varepsilon_i, \xi_i \sim N(0,1)$$

Decision Criterion – RMSE:

$$RMSE = \sqrt{\frac{\sum\left(\hat{\beta} - \beta_{true}\right)^2}{N}} = \sqrt{Var\left(\hat{\beta}\right) + \left[Bias\left(\hat{\beta}, \beta_{true}\right)^2\right]}$$

Comparison: Simple OLS of stage 2, 2SLS, 2 stage IV-Lowess, 2 stage IV-Kernel, 2 stage IV-KRLS

Set up:

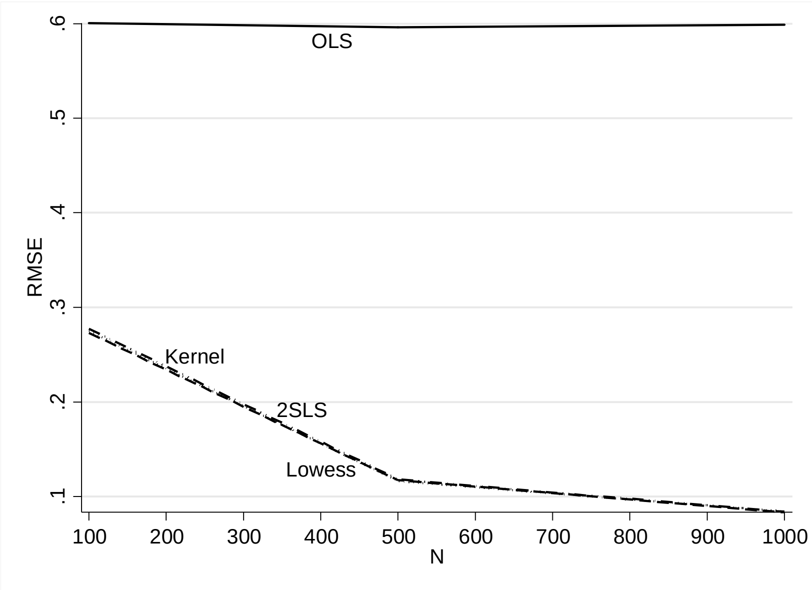
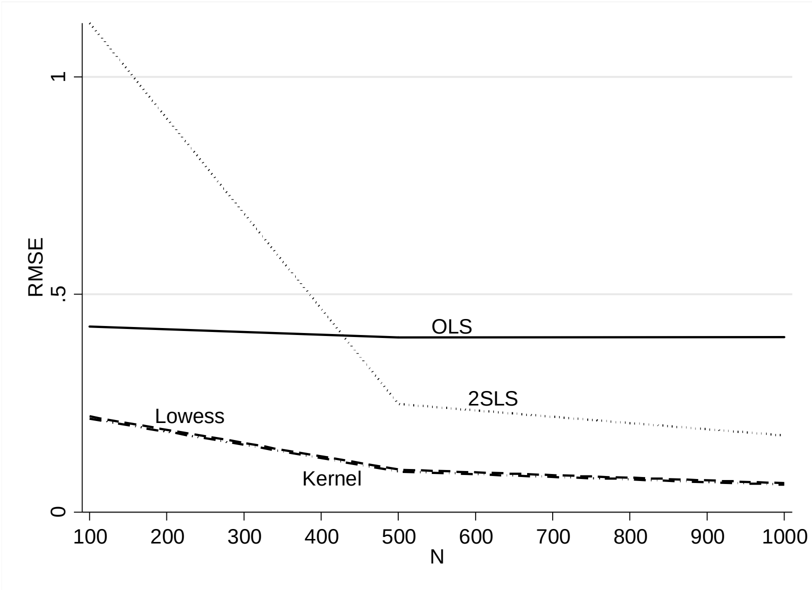
1. Number of observations: N = [100, 500, 1000] → efficiency for all estimators should increase with larger number of observations, N = [100, 300, 500] for KRLS
2. Degree of endogeneity of X: corr(x,e) = [0.3, 0.6]
3. Strength of the instrument Z: corr(x,z) = [0.3, 0.6]
4. Validity of the instrument Z: corr(z,e) = [0, 0.3, 0.6]
5. Linearity of the relationship between X and Z: gamma2=0, gamma2>0
6. Bandwidth of the loess smoothing in the first stage: bw = [0.2, 0.5, 0.8] → the greater the bandwidth the greater the smoothing.

1000 repetitions for each permutation of different features

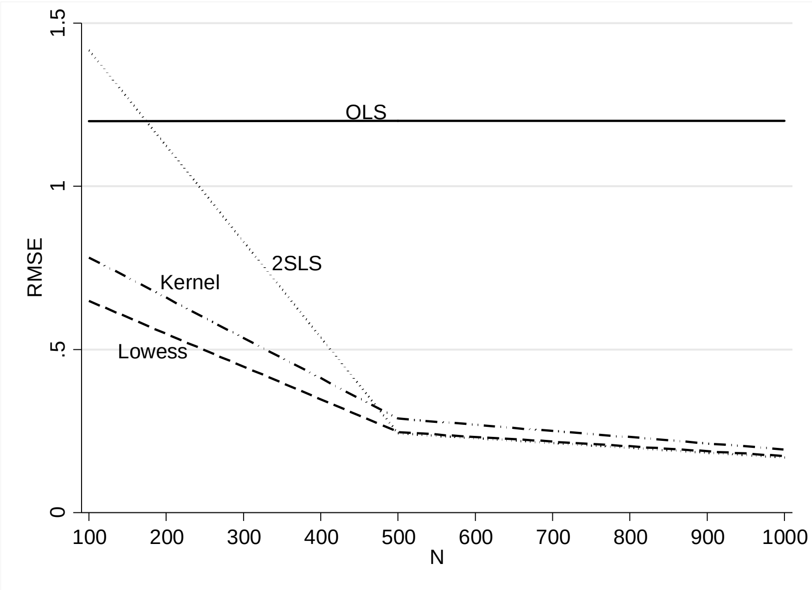
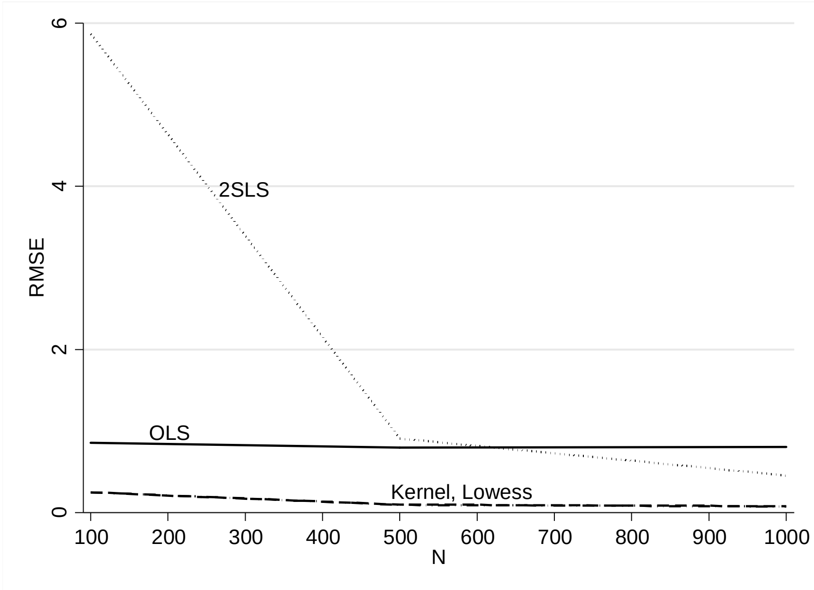
MC Results

Single Excluded Instrument

weak endogeneity and strong instrument

<table>
<tr><td>linear relationship between x and z</td><td>quadratic relationship between x and z</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td colspan="2">Endogeneity: corr(x,e)=0.3; Strength of instrument: corr(x,z)=0.6; valid instrument: corr(z,e)=0; bw=0.5</td></tr>
</table>

strong endogeneity and weak instrument

linear relationship between x and z	quadratic relationship between x and z
	
Endogeneity: corr(x,e)=0.6; Strength of instrument: corr(x,z)=0.3; valid instrument: corr(z,e)=0; bw=0.5	

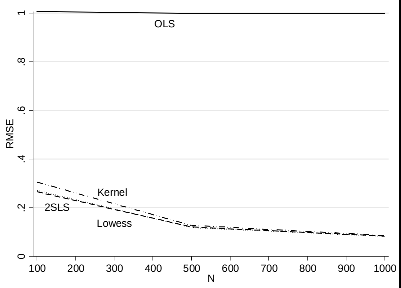
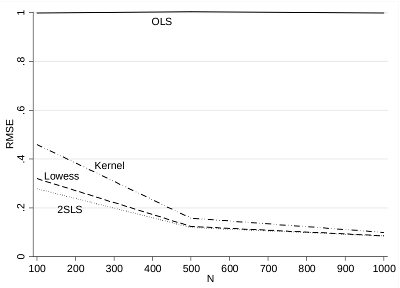
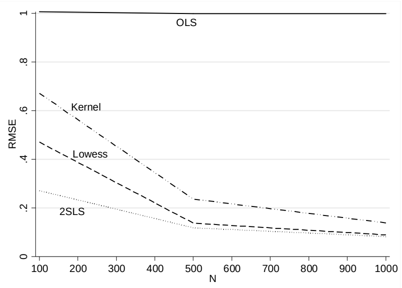
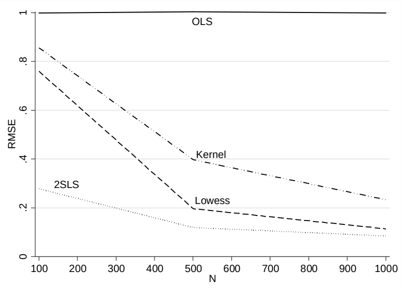
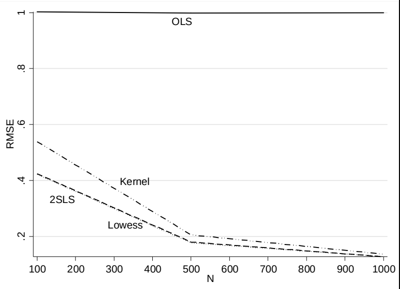
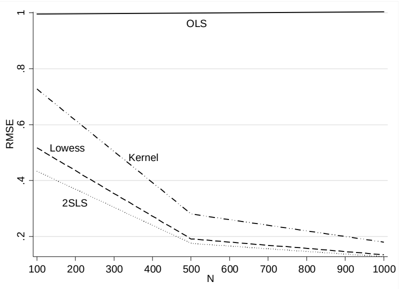
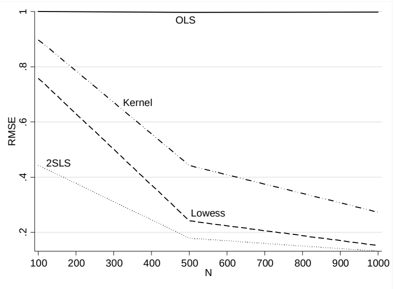
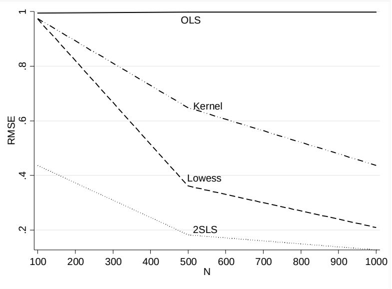
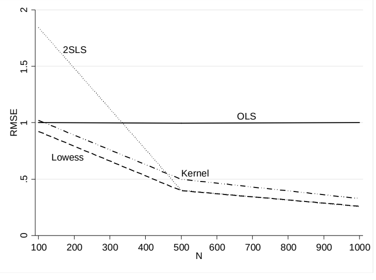
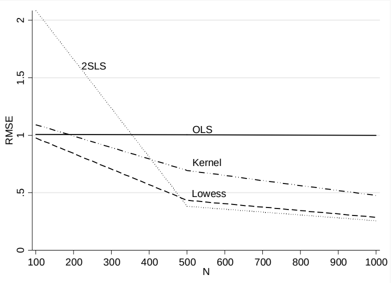
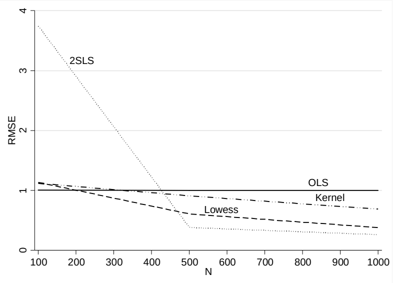
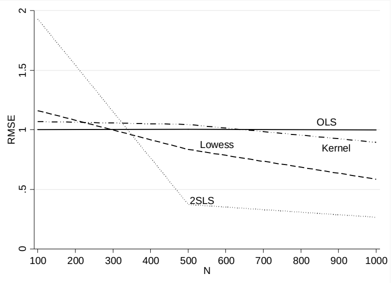
A Note on Overfitting

In this set of MCs we vary the following features of the DGP:

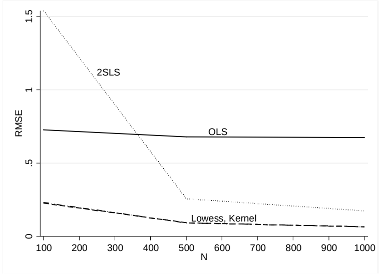
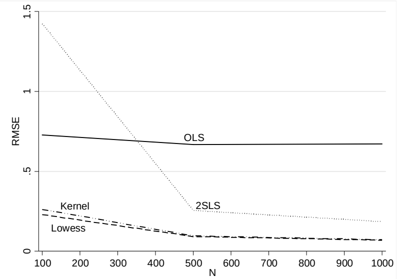
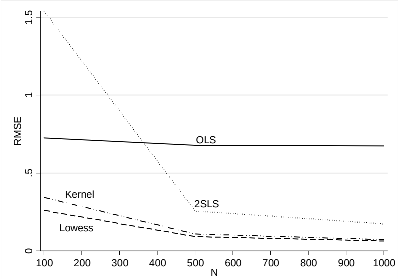
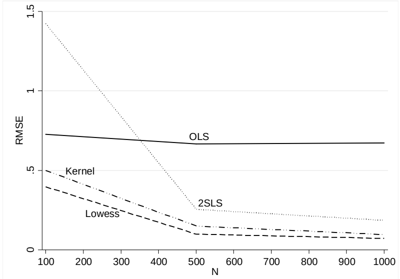
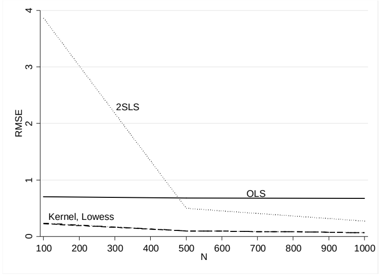
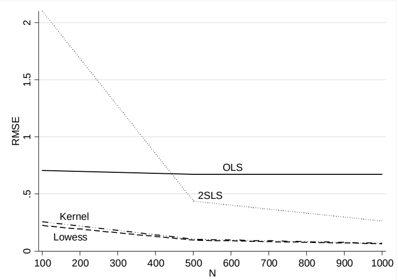
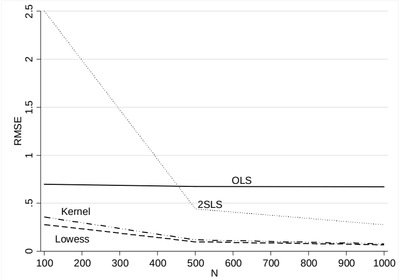
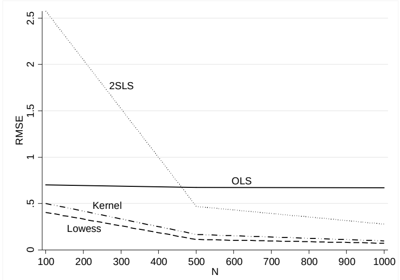
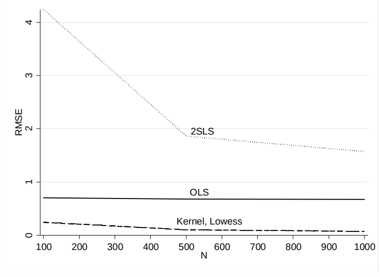
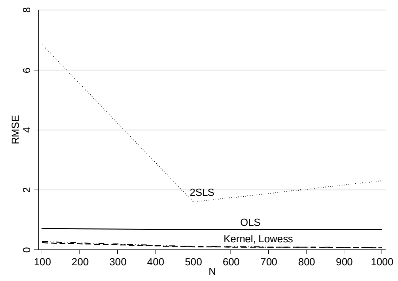
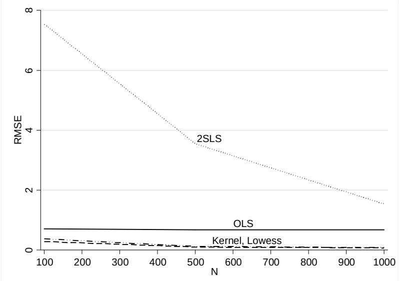
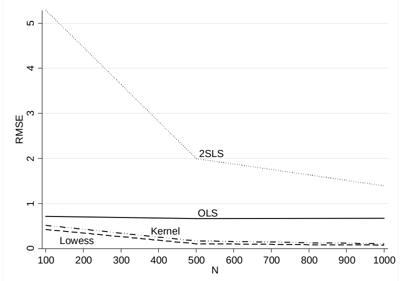
1. Number of observations: N = [100, 500, 1000]
2. Strength of the instrument Z: corr(x,z) = [0.2, 0.4, 0.6] → the problem of overfitting should be larger for weaker instruments
3. Linearity of the relationship between X and Z: gamma2=0, gamma2>0 → overfitting should be a bigger problem when Z exerts a linear effect on X
4. Bandwidth of the lowess smoothing and kernel regression in the first stage (the α parameter): bw = [0.05, 0.1, 0.2, 0.4] → smaller bandwidth potentially lead to overfitting

In addition we assume the instrument Z to be valid (i.e. corr(z,e)=0) and we hold the endogeneity of X constant at a medium level (i.e. corr(x,e)=0.5).

Effect of bandwidth and instrument strength on estimation of the endogenous RHS variable

<table>
<tr><td>Corr(x,z)</td><td>Bandwidth = 0.4</td><td>Bandwidth = 0.2</td><td>Bandwidth = 0.1</td><td>Bandwidth = 0.05</td></tr>
<tr><td>0.6</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
<tr><td>0.4</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
<tr><td>0.2</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
<tr><td colspan="5">Linear relationship between x and y; corr(x,e)=0.5; corr(z;e)=0</td></tr>
</table>

Effect of bandwidth and instrument strength on estimation of the endogenous RHS variable

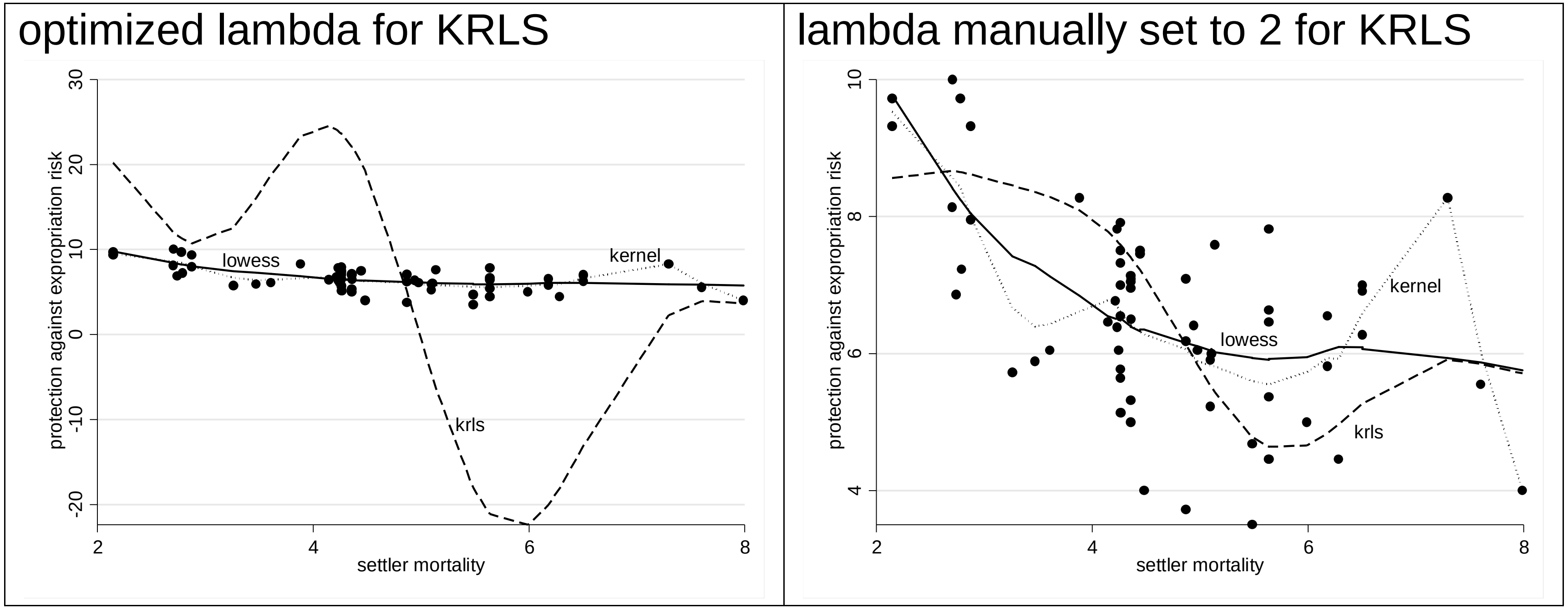
Corr(x,z)	Bandwidth = 0.4	Bandwidth = 0.2	Bandwidth = 0.1	Bandwidth = 0.05
0.6				
0.4				
0.2				
Quadratic relationship between x and y; corr(x,e)=0.5; corr(z;e)=0				

Empirical Examples

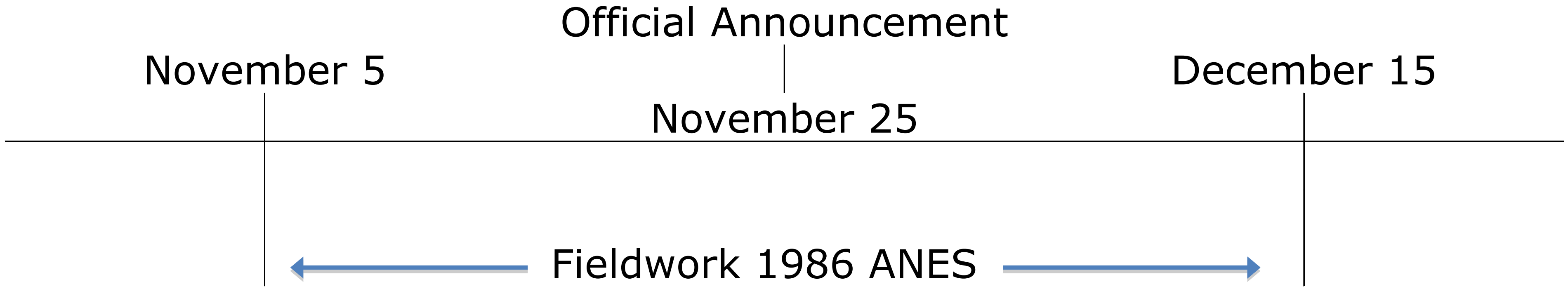
Settler Mortality as Instrument for Institutional Quality (Acemoglu, Johnson and Robinson, AER 2001)

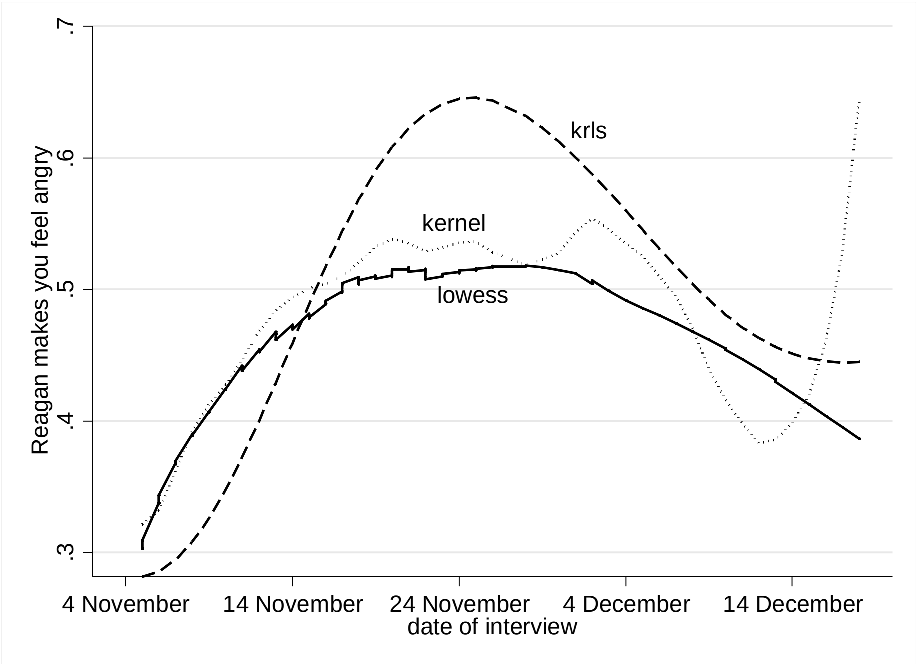
DV 2nd stage: Log GDP 1995	no exogenous covariates						with exogenous covariates					
	OLS	2SLS	Lowess	Kernel	KRLS	KRLS	OLS	2SLS	Lowess	Kernel	KRLS	KRLS
Instrument: Log European Settler Mortality												
average protection against expropriation risk	0.522***	0.944***					0.401***	1.107**				
	(0.061)	(0.157)					(0.059)	(0.464)				
Protection against expropriation (predicted)			0.749***	0.694***	0.031***	0.480***			1.078***	0.784***	5.109***	0.741***
			(0.112)	(0.097)	(0.007)	(0.073)			(0.227)	(0.169)	(1.073)	(0.154)
Lambda (KRLS)					0.055	2					39.13	2
Bandwidth (Lowess, Kernel(optimized))			0.8	0.5					0.8	0.8		
Exogenous covariates	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Intercept	4.660***	1.910*	3.109***	3.511***	7.836***	4.814***	5.737***	1.44	1.022	2.985**	-25.12***	3.374***
	(0.409)	(1.027)	(0.749)	(0.644)	(0.122)	(0.502)	(0.398)	(2.840)	(1.516)	(1.137)	(6.998)	(1.014)
R^2	0.54	0.187	0.418	0.451	0.257	0.413	0.714	0.011	0.631	0.626	0.631	0.634
N	64	64	64	64	64	64	64	64	64	64	64	64
F	72.816	36.394	44.556	50.990	21.49	43.62	28.946	6.847	19.82	19.376	19.82	20.07
2SLS First Stage, DV: Average protection against expropriation risk												
Log European settler mortality		-0.607***			-0.682**	-0.474***		-0.340*			-0.090	-0.324
		(0.126)			(0.239)	(0.109)		(0.183)			(0.062)	(0.200)
Partial R^2		0.270			0.494	0.395		0.056			0.019	0.080

Note: columns 2 (OLS) and 3(2SLS) exactly replicate results presented in Table 4/column 1; and columns 8 (OLS) and 9 (2SLS) exactly replicate results presented in table 4/ column 8 in Acemoglu, Johnson and Robinson (1990, p. 1386)


optimized lambda for KRLS
protection against expropriation risk
30
20
10
0
-10
-20
2
4
6
8
settler mortality
lowess
kernel
krls
lambda manually set to 2 for KRLS
protection against expropriation risk
10
8
6
4
2
4
6
8
settler mortality
kernel
lowess
krls

Timing of Scandals and Presidential Approval (Krosnick and Kinder, APSR 1990)

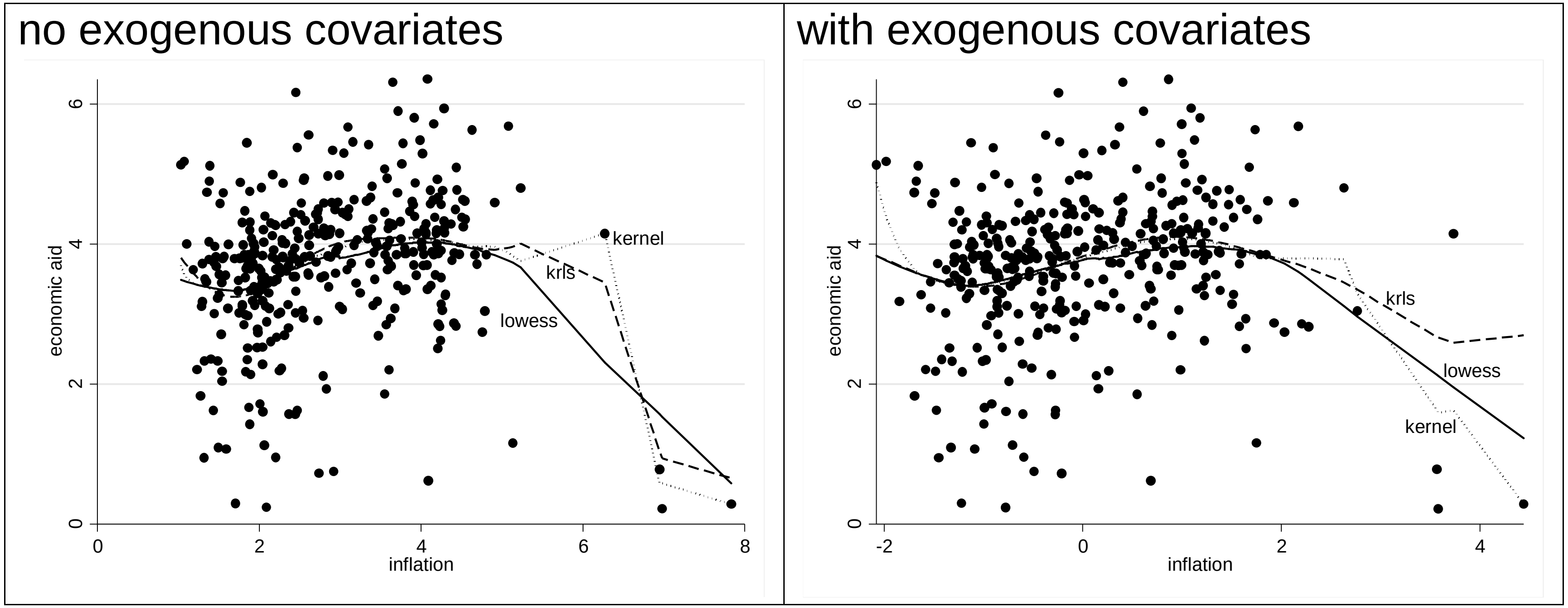




Estimating the effect of emotions on affective Presidential evaluations, Reagan 1986

DV 2nd stage: Regan feeling thermometer	OLS	2SLS: before/after 25/11	2SLS: trend (date)	Lowess	Kernel	KRLS
Instrument: date of interview						
Reagan makes	-25.542***					
you feel angry	(1.544)					
Angry (predicted)		-389.049	-68.41	-33.101**	-29.935**	-12.175**
		(1603.386)	(42.221)	(14.992)	(13.714)	(5.141)
Intercept	76.502***	246.848	96.591***	79.603***	78.195***	69.316***
	(1.057)	(751.401)	(19.813)	(6.881)	(5.952)	(2.507)
Lambda (KRLS)						95.97
Bandwidth (Lowess, Kernel)				0.8	0.8	
R^2	0.223	.	.	0.005	0.005	0.003
N	956	956	956	956	956	1924
F	273.516	0.059	2.625	4.875	4.76	5.61
2SLS First Stage, DV: Reagan makes you feel angry						
Dummy – after 25/11		0.009				
		(0.037)				
Date of interview			0.002			0.004***
			(0.001)			(0.001)
Partial R^2		0.0001	0.002			0.014

Foreign Aid and Democratic Development in Africa (Dietrich and Wright, JOP 2014)



The Effect of Economic Aid on Move to Multiparty elections

DV 2nd stage: Multiparty politics	without exogenous covariates					with exogenous covariates					
	OLS	2SLS	Lowess	Kernel	KRLS	OLS	2SLS	2SLS_Lewbel	Lowess	Kernel	KRLS
Instrument: inflation											
Economic aid	0.023					0.058**					
	(0.017)					(0.026)					
Economic aid (predicted)		0.169	0.065	0.048	0.041		0.266	0.108**	0.192***	0.134**	0.170***
		(0.125)	(0.051)	(0.043)	(0.043)		(0.18)	(0.043)	(0.069)	(0.054)	(0.066)
Exogenous covariates	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Intercept	0.048	-0.493	-0.107	-0.045	-0.020	-0.135	-1.459	-0.452	-0.607	-0.387	-0.475
	(0.066)	(0.465)	(0.189)	(0.161)	(0.159)	(0.290)	(1.174)	(0.361)	(0.385)	(0.344)	(0.362)
Lambda (KRLS)					0.264						3.526
Bandwidth (Lowess, Kernel)			0.8	0.8					0.8	0.8	
R^2	0.005	.	0.004	0.003	0.003	0.039	.	0.029	0.046	0.042	0.044
N	370	370	370	370	370	370	370	370	370	370	370
F	1.777	1.811	1.616	1.242	0.931	2.099	1.49	2.264	2.503	2.286	2.357
2SLS First Stage, DV: economic aid											
inflation		0.141**			0.166**		0.100***	0.156***			0.130**
		(0.049)			(0.078)		(0.033)	(0.028)			(0.060)
Partial R^2		0.022			0.181		0.024				0.108

Note: columns 7 (multivariate OLS) and 9 (Lewbel 2SLS) exactly replicate the empirical results in table 1, columns 1 and 2 in Dietrich, Wright (2014).

Conclusions

"Some Times You Do Not Get What You Want."

Even if exclusion can be credibly satisfied, first stage may not.

What is left to do then?

1. Try to find other instruments – Exclusion more difficult to satisfy
2. Abandon the IV altogether – throwing out the baby with the bathwater
3. The non-parametric IV presents a promising alternative

Non parametric first stage does not borrow identification at the expense of untestable functional form assumptions.

2. Fixed Effects and Dynamic Miss-specifications

First difference and fixed effects estimators are used widely to get a grip on identification by abstracting from individual characteristics

Since FE only uses variation over time, dynamic miss-specification and omitted time variant information can affect performance and outweigh gains from controlling for omitted time-invariant effects

What is more common – omitting time invariant or time varying factors?

Dynamics are complex and hard to model esp. in pooled data set ups: discussion of dynamic specification in applied research is rare at best and usually quick fixes are applied widely (LDV, dynamic panel, period FE, ADL, ECM etc.)

Sources of Dynamic Misspecification

Applied researchers often perceive serially correlated errors as noise rather than information

Yet, serially correlated errors clearly indicates a potentially severe model misspecification, which can result from various sources:

- incompletely or incorrectly modelled persistency in the dependent variable
- time-varying omitted variables
- changes in the effect strengths of time-invariant variables
- misspecified lagged effects of explanatory variables
- conditionality of treatment effects on unobserved time varying variables
- spatial dependence

What do applied researchers do?

Dynamics should be directly modelled, but dynamic misspecifications are manifold and complex…

Researchers usually use readily available econometric patches to treat serial correlation:

- do nothing (inter alia Moro et al. 2013; Humphreys/Weinstein 2006; Ross 2008)
- LDV (e.g. Lupu and Pontusson 2011; Kogan et al. 2016; Acemoglu et al. 2008, 2009; Guisinger and Singer 2010)
- Period dummies (Besley and Reynal-Querol 2011, Menaldo 2012, Egorov 2009 among many others)
- GLS type - Prais-Winsten (Mukherjee 2009, Lupu and Pontusson 2011)
- ADL (Gerber et al. 2011)

Econometric fixes are not correct per se because they are usually not modelling the true dynamic process in the underlying data generating process

Cleaning the residuals does not necessarily eliminate potential bias
Using different patches can produce vastly different estimates and effects

Bias of Dynamic Misspecification (Example)

DGP: $y_{it} = \beta x_{it-1} + u_i + \varepsilon_{it}$

If we ignore the lagged effect of x_{it} , the probability limit (plim) of the OLS estimator of *beta* in the regression $y_{it} = \beta x_{it} + \varepsilon_{it}$ is given by:

$$\frac{Cov\left(y_{it}, x_{it}\right)}{Var\left(x_{it}\right)} = \beta \frac{Cov\left(x_{it-1}, x_{it}\right)}{Var\left(x_{it}\right)} + \frac{Cov\left(u_i, x_{it}\right)}{Var\left(x_{it}\right)}$$

The probability limit of the fixed effects estimator equals:

$$\frac{Cov\left(y_{it}, \ddot{x}_{it}\right)}{Var\left(x_{it}\right)} = \beta \frac{Cov\left(x_{it-1}, \ddot{x}_{it}\right)}{Var\left(x_{it}\right)} + \frac{Cov\left(u_i, \ddot{x}_{it}\right)}{Var\left(x_{it}\right)} =$$

$$\frac{Cov\left(y_{it}, \ddot{x}_{it}\right)}{Var\left(x_{it}\right)} = \beta \frac{Cov\left(x_{it-1}, x_{it}\right)}{Var\left(x_{it}\right)} - \beta \frac{Cov\left(x_{it-1}, x_i\right)}{Var\left(x_{it}\right)}$$

$$\Leftrightarrow \frac{Cov\left(y_{it}, \ddot{x}_{it}\right)}{Var\left(x_{it}\right)} = \beta \frac{Cov\left(x_{it-1}, x_{it}\right)}{Var\left(x_{it}\right)} - \beta \frac{Var\left(x_i\right)}{Var\left(x_{it}\right)}$$

MC Set-up

We run five sets of experiments that examine different dynamic misspecifications:

i) omitted time-varying variable,
ii) omitted common trend,
iii) unit specific trend,
iv) misspecified common lag-structure, and
v) misspecified unit-specific lag-structure.

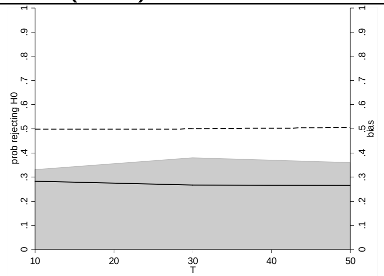
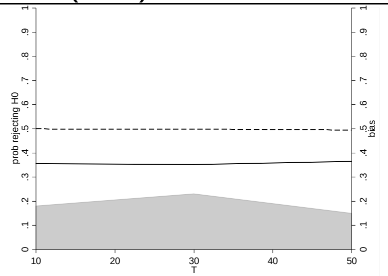
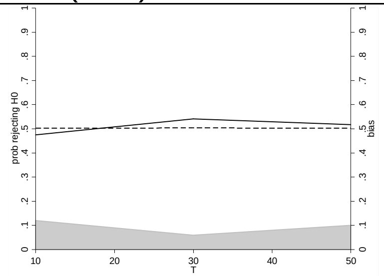
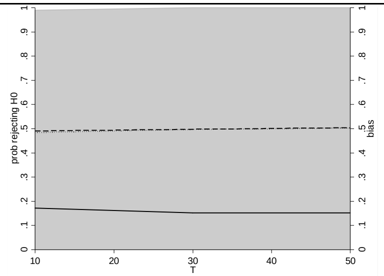
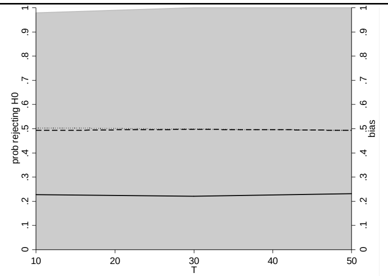
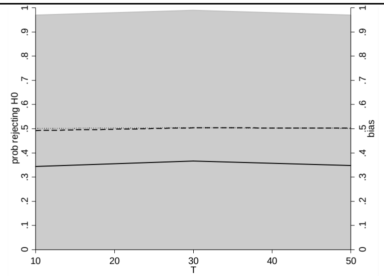
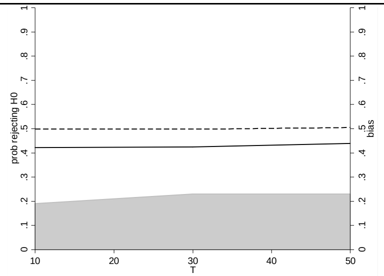
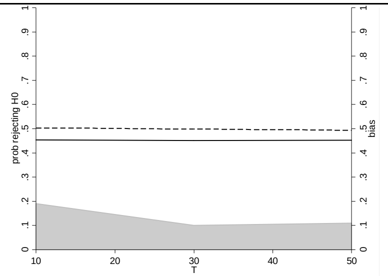
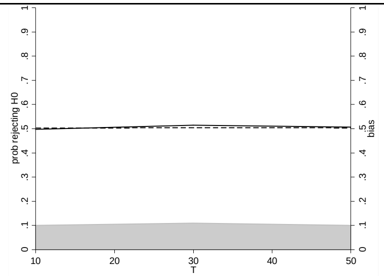
For each of the misspecifications we estimate six different fixed effects and pooled OLS models with different dynamic specification: no dynamics, lagged dependent variable (LDV, or Arellano-Bond (A-B) model), Prais-Winsten GLS transformation, period fixed effects, a combination of LDV/A-B and period fixed effects, and an autoregressive distributed lag (ADL) model.

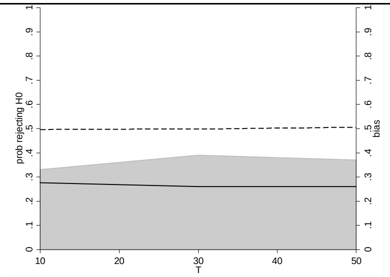
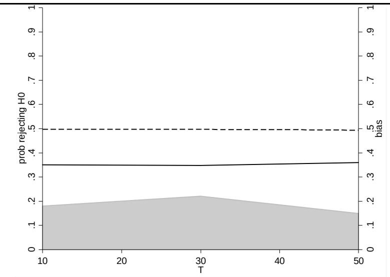
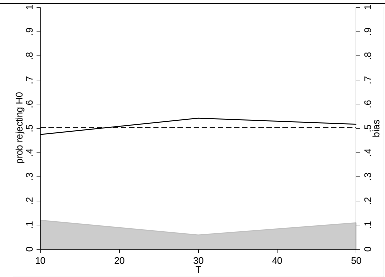
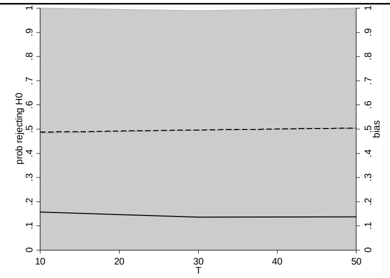
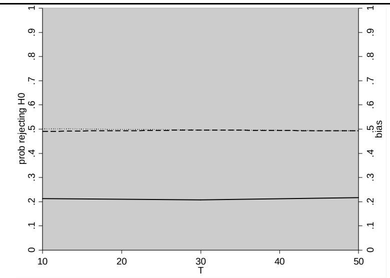
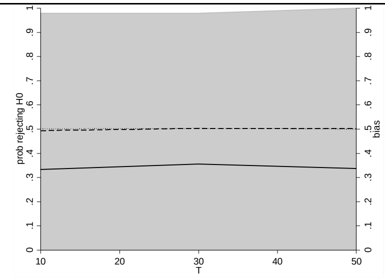
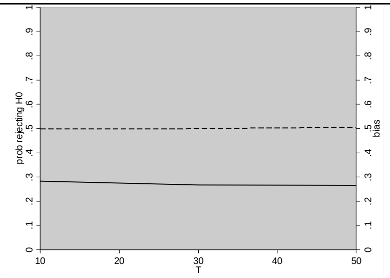
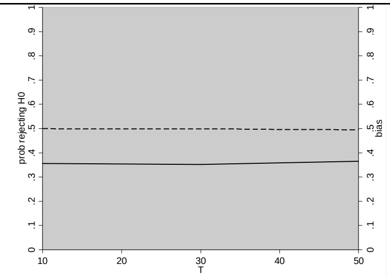
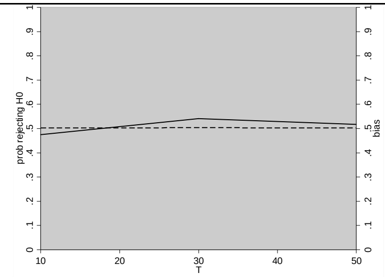
Finally we vary the correlations between the unit specific effects and the interesting RHS variable, as well as the number of periods.

Bias over all Experiments

Econometric Specification	*Bias: pooled OLS*			*Bias: FE*		
	mean	min	Max	mean	min	max
No Dynamics	0.377	0.035	0.665	0.620	0.060	1.118
LDV	0.315	0.009	0.753	0.580	0.048	1.133
Arellano-Bond (A-B)				0.597	0.000	1.393
Prais-Winsten GLS	0.547	0.028	1.321	0.612	0.042	1.182
Period Fixed Effects	0.335	0.007	0.662	0.563	0.001	1.116
LDV+ Period Fixed Effects	0.316	0.001	0.749	0.546	0.000	1.131
A-B + Period Fixed Effects				0.569	0.000	1.386
ADL	0.295	0.044	0.487	0.473	0.002	1.006

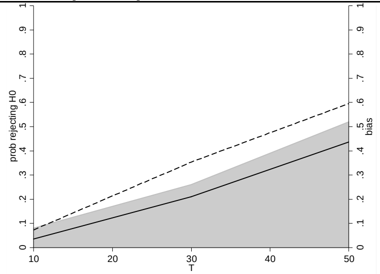
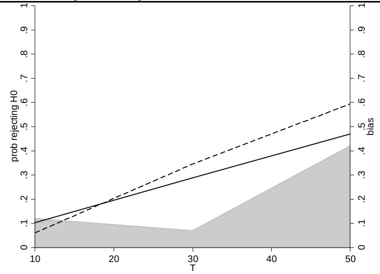
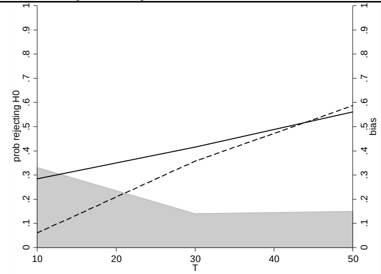
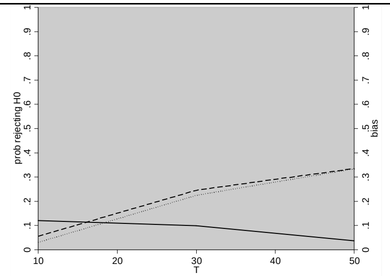
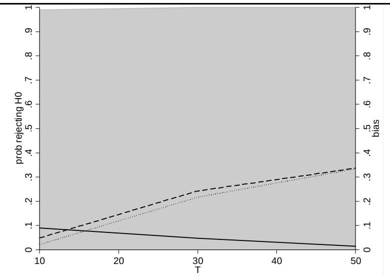
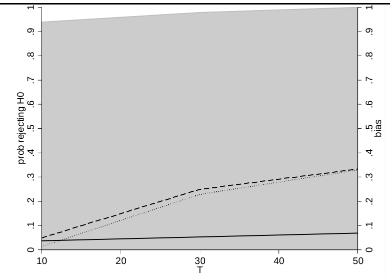
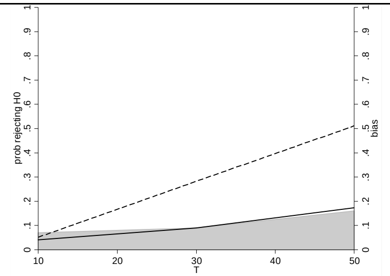
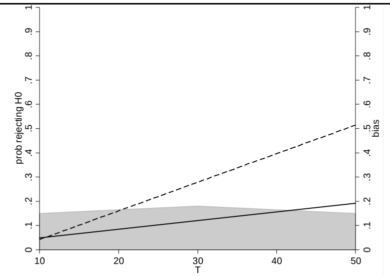
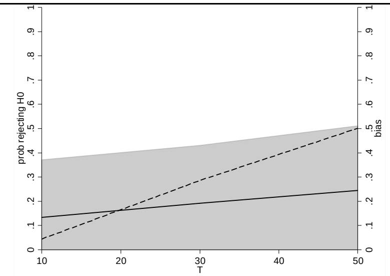
Experiment 1: Omitted Time-varying Variable

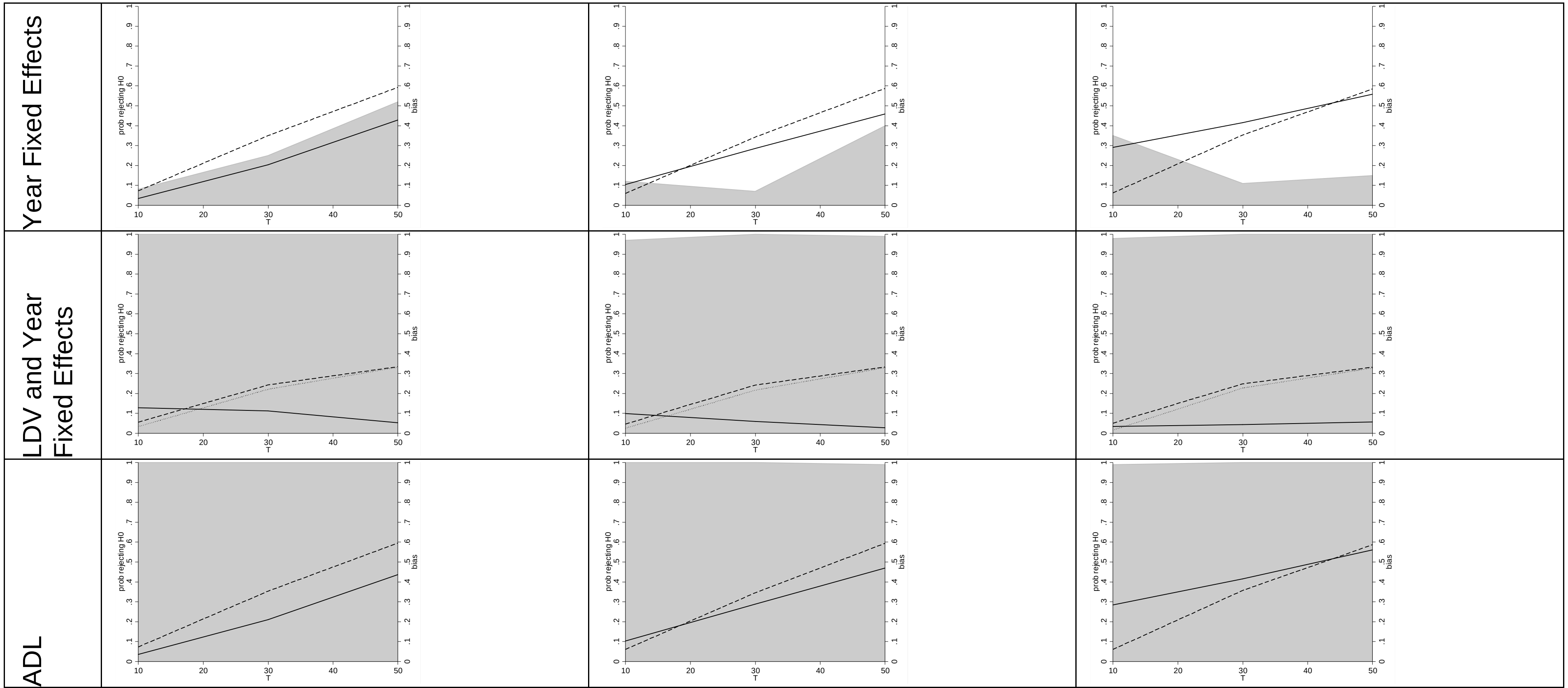
	Corr(x1,u)=0.0	Corr(x1,u)=0.2	Corr(x1,u)=0.5
No dynamics			
LDV/ Arellano-Bond			
Prais-Winsten			

Year Fixed Effects			
LDV and Year Fixed Effects			
ADL			

Right Axis – Absolute Bias: ——— OLS, - - - - - FE, · · · · · A-B; Left Axis - Probability of rejecting the H0 on the 5% level and thus suggesting FE: grey shaded area = Hausman Test

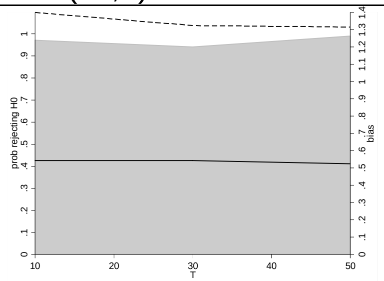
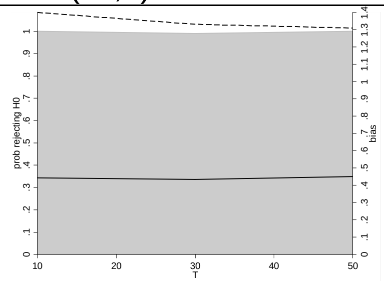
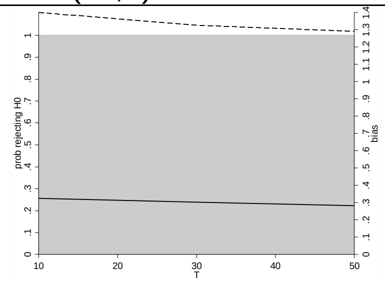
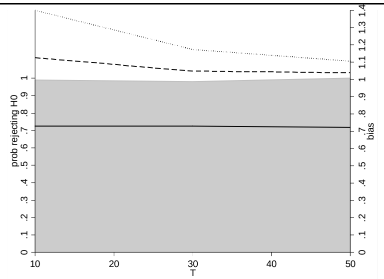
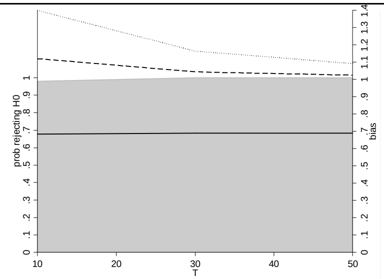
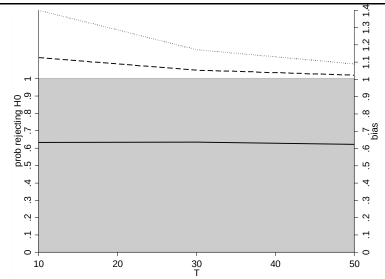
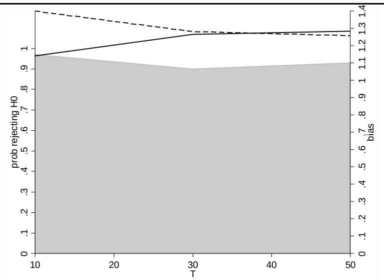
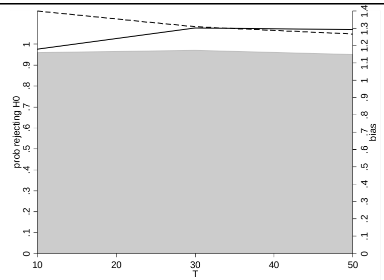
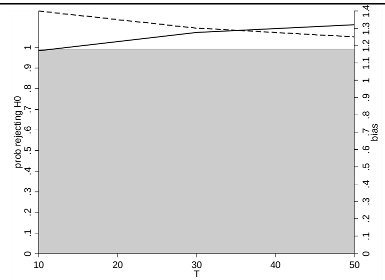
Experiment 3: Correlated Unit-specific Trends

	Corr(x1,u)=0.0	Corr(x1,u)=0.2	Corr(x1,u)=0.5
No dynamics			
LDV/ Arellano-Bond			
Prais-Winsten			

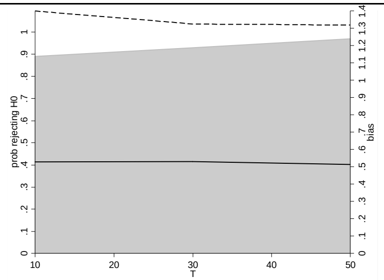


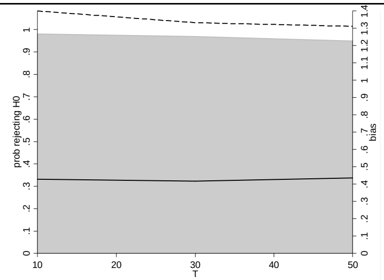
Right Axis – Absolute Bias: ——— OLS, - - - - - FE, · · · · · A-B;
Left Axis - Probability of rejecting the H0 on the 5% level and thus suggesting FE: grey shaded area = Hausman Test

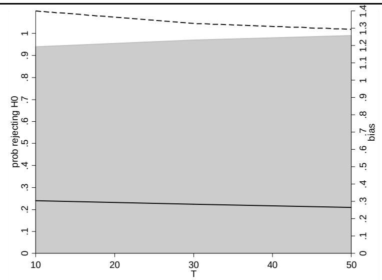
Experiment 4: Miss-specified Lag-structure of RHS variable

	Corr(x1,u)=0.0	Corr(x1,u)=0.2	Corr(x1,u)=0.5
No dynamics			
LDV/ Arellano-Bond			
Prais-Winsten			

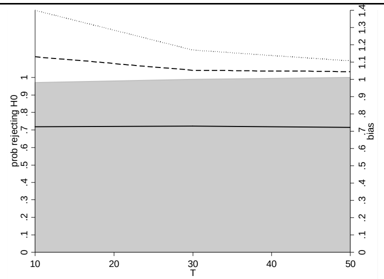
Year Fixed Effects

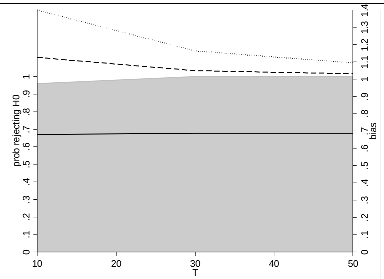


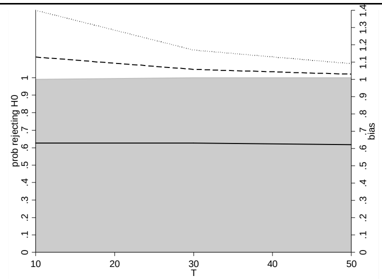




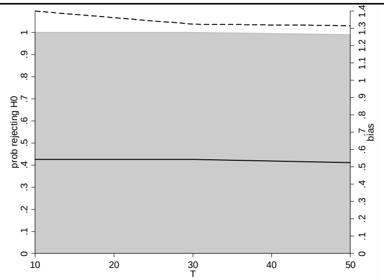
LDV and Year Fixed Effects

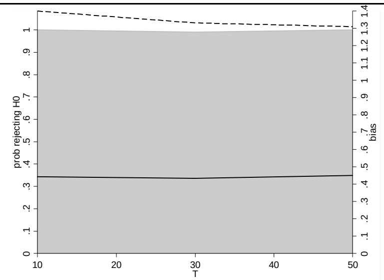


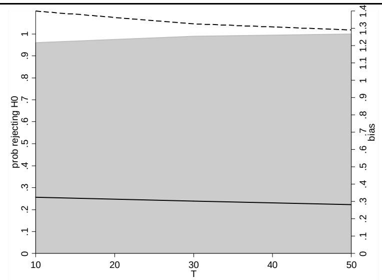




ADL







Right Axis – Absolute Bias: ——— OLS, - - - - - FE, · · · · · A-B;
Left Axis - Probability of rejecting the H0 on the 5% level and thus suggesting FE: grey shaded area = Hausman Test

What do we learn?

The fixed effects estimator is biased in the presence of dynamic misspecification and omitted

Fixed effects estimates amplify the bias from dynamic misspecification as compared to the ‘naïve’ OLS model

The Hausman-test does not reliably identify the least biased estimator when both time-invariant and time-varying omitted variables or other dynamic misspecifications exist

Applied researchers are ill-advised to rely on the Hausman-test for model selection or use the fixed effects model as default unless they can convincingly justify the assumption of correctly specified dynamics.

The findings caution applied researchers to not overlook the flipside of the fixed effects estimator

Methodologists need to study the properties of estimators in the presence of multiple model misspecifications

3. Institutional persistency: time invariant and slowly moving variables

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

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

Complication(s)

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 by far

- government spending
- per capita income
- household income
- trade openess

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

Hausman-Taylor, Amemiya-MaCurdy

The Hausman-Taylor and Amemiya-MaCurdy estimators are developed for data where the unit specific effects are correlated with RHS variables – which would require a FE specification, but some of the theoretically interesting explanatory variables are not changing over time.

In a FE model the coefficient of time-constant explanatory variables are not identified.

The models are based on a correlated random effects model (see Mundlak 1978 and Chamberlain 1982, 1984) and use instrumental variables for the endogenous RHS variables.

Excursus: Chamberlain, Mundlak: the strict exogeneity assumption still holds but these models allow arbitrary correlation between u_i and x_{it}. The Chamberlain approach is to replace the unobserved unit specific effect u_i with its linear projection onto the explanatory variables in all time periods (plus the projection error).
Assumptions: x_{it} has to vary substantially over time, $E(u_i \mid x_i)$ has to be linear.

The underlying assumption is that only some of the time-varying (x_{it}) and time-invariant (z_i) variables are correlated with the unit specific effects u_i.
The uncorrelated x_{it} and z_i therefore can be used as instruments for the correlated RHS variables.
The within transformed x_{it} serve as instruments for the correlated x_{it} (these are estimated by FE) and the unit means of the uncorrelated x_{it} ($xbar_i$) as well as the uncorrelated z_i serve as instruments for the correlated z_i.

Hausman-Taylor

$$y_{it} = \mathbf{x1}'_{it}\boldsymbol{\beta}_1 + \mathbf{x2}'_{it}\boldsymbol{\beta}_2 + \mathbf{z1}'_{i}\boldsymbol{\alpha}_1 + \mathbf{z2}'_{i}\boldsymbol{\alpha}_2 + \varepsilon_{it} + u_i$$

$$E[u_i \mid \mathbf{x1}_{it}, \mathbf{z1}_i] = 0$$

$$E[u_i \mid \mathbf{x2}_{it}, \mathbf{z2}_i] \neq 0 \;\Rightarrow\; \textbf{OLS and GLS are inconsistent}$$

$$Var[u_i \mid \mathbf{x1}_{it}, \mathbf{x2}_{it}, \mathbf{z1}_i, \mathbf{z2}_i] = \sigma_u^2$$

$$E[\varepsilon_{\mathbf{it}} \mid \mathbf{x1}_{it}, \mathbf{x2}_{it}, \mathbf{z1}_i, \mathbf{z2}_i] = 0$$

$$Var[\varepsilon_{\mathbf{it}} \mid \mathbf{x1}_{it}, \mathbf{x2}_{it}, \mathbf{z1}_i, \mathbf{z2}_i] = \sigma_\varepsilon^2$$

$$Cov[\varepsilon_{\mathbf{it}}, u_i \mid \mathbf{x1}_{it}, \mathbf{x2}_{it}, \mathbf{z1}_i, \mathbf{z2}_i] = 0$$

$$Var[\varepsilon_{\mathbf{it}} + u_i \mid \mathbf{x1}_{it}, \mathbf{x2}_{it}, \mathbf{z1}_i, \mathbf{z2}_i] = \sigma_\varepsilon^2 + \sigma_u^2$$

$$Cov[\varepsilon_{\mathbf{it}} + u_i, \varepsilon_{\mathbf{is}} + u_i \mid \mathbf{x1}_{it}, \mathbf{x2}_{it}, \mathbf{z1}_i, \mathbf{z2}_i] = \sigma_u^2$$

Hausman-Taylor

$$y_{it} = \mathbf{x1}'_{it}\boldsymbol{\beta}_1 + \mathbf{x2}'_{it}\boldsymbol{\beta}_2 + \mathbf{z1}'_{i}\boldsymbol{\alpha}_1 + \mathbf{z2}'_{i}\boldsymbol{\alpha}_2 + \varepsilon_{it} + u_i$$

Deviations from group means removes all time invariant variables

$$y_{it} - \bar{y}_i = (\mathbf{x1}_{it} - \overline{\mathbf{x1}}_i)'\boldsymbol{\beta}_1 + (\mathbf{x2}_{it} - \overline{\mathbf{x2}}_i)'\boldsymbol{\beta}_2 + \varepsilon_{it}$$

Implication: $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ are consistently estimated by LSDV.

$(\mathbf{x1}_{it} - \overline{\mathbf{x1}}_i) = K_1$ instrumental variables

$(\mathbf{x2}_{it} - \overline{\mathbf{x2}}_i) = K_2$ instrumental variables

$\mathbf{z1}_i$ $= L_1$ instrumental variables (uncorrelated with u)

? $= L_2$ instrumental variables (where do we get them?)

H&T: $\overline{\mathbf{x1}}_i = K_1$ additional instrumental variables. Needs $K_1 \geq L_2$.

Hausman-Taylor: FGLS estimator

(1) LSDV estimates of $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma_\varepsilon^2$

(2) $(\mathbf{e^*})' = (\bar{e}_1, \bar{e}_1, \ldots, \bar{e}_1), (\bar{e}_2, \bar{e}_2, \ldots, \bar{e}_2), \ldots, (\bar{e}_N, \bar{e}_N, \ldots, \bar{e}_N)$

($\Sigma_{i=1}^{N} T_i$ observations).

$$\mathbf{Z_i}^* = \begin{bmatrix} \mathbf{z}'_{i1} & \mathbf{z}'_{i2} \\ \mathbf{z}'_{i1} & \mathbf{z}'_{i2} \\ \vdots & \vdots \\ \mathbf{z}'_{i1} & \mathbf{z}'_{i2} \end{bmatrix} = \begin{array}{l} T_i \text{ rows, repeat invariant variables} \\ L_1 + L_2 \text{ columns} \end{array}$$

$$\mathbf{W_i} = \begin{bmatrix} \mathbf{z}'_{i1} & \mathbf{x}'_{i1,1} \\ \mathbf{z}'_{i1} & \mathbf{x}'_{i1,2} \\ \vdots & \vdots \\ \mathbf{z}'_{i1} & \mathbf{x}'_{i1,T_i} \end{bmatrix} = \begin{array}{l} T_i \text{ rows, repeat } \mathbf{z}'_{i1}\text{, time varying } \mathbf{x}'_{i1,t} \\ L_1 + K_1 \text{ columns} \end{array}$$

Hausman-Taylor: FGLS estimator

(2 cont.) IV regression of $\mathbf{e}^*$ on $\mathbf{Z}^*$ with instruments $\mathbf{W_i}$ consistently estimates $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$.

(3) With fixed T, residual variance in (2) estimates $\sigma_u^2 + \sigma_\varepsilon^2 / T$

With unbalanced panel, it estimates $\sigma_u^2 + \sigma_\varepsilon^2 \overline{(1/T)}$ or something resembling this. (1) provided an estimate of σ_ε^2 so use the two to obtain estimates of σ_u^2 and σ_ε^2. For each group, compute

$\hat{\theta}_i = 1 - \sqrt{\hat{\sigma}_\varepsilon^2 / (\hat{\sigma}_\varepsilon^2 + T_i \hat{\sigma}_u^2)}$

(4) Transform $[\mathbf{x_{it1}}, \mathbf{x_{it2}}, \mathbf{z_{i1}}, \mathbf{z_{i2}}]$ to

$$\mathbf{W_i}^* = [\mathbf{x_{it1}}, \mathbf{x_{it2}}, \mathbf{z_{i1}}, \mathbf{z_{i2}}] - \hat{\theta}_i [\bar{\mathbf{x}}_{\mathbf{i1}}, \bar{\mathbf{x}}_{\mathbf{i2}}, \mathbf{z_{i1}}, \mathbf{z_{i2}}]$$

and y_{it} to $y_{it}^* = y_{it} - \hat{\theta}_i \bar{y}_i$.

Unfortunately, these ‘solutions’ are pretty bad…

Hausman-Taylor: inefficient, biased if instruments are poor.

The researcher is left with the decision which of the x_{it} and z_i are uncorrelated with the unit specific effects. For the x_{it} a regression addition test (Mundlak formulation of the Hausman specification test) helps to separate the correlated from the uncorrelated x_{it}. For the time-constant z_i such a test does not exist.

There have to be at least as many uncorrelated RHS variables as there are correlated RHS variables – that is often not the case.

random effects: biased

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

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:

$$V_{FEVD}\left(\beta,\gamma\right)=\left(H'W\right)^{-1}H'\Omega H\left(W'H\right)^{-1}$$

$$H=\left[\ddot{X},Z\right]$$

$$W=\left[X,Z\right]$$

$$\Omega=\sigma_{\varepsilon}^{2}I_{NT}+\sigma_{\eta}^{2}I_{N}\otimes\iota_{T}\iota_{T}'$$

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

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:

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

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

Demeaning (Stage 1 estimation):

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

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

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

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

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:

$$RMSE = \sqrt{\frac{\sum\left(\hat{\beta} - \beta_{true}\right)^2}{N}} = \sqrt{Var\left(\hat{\beta}\right) + \left[Bias\left(\hat{\beta}, \beta_{true}\right)^2\right]}$$

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)

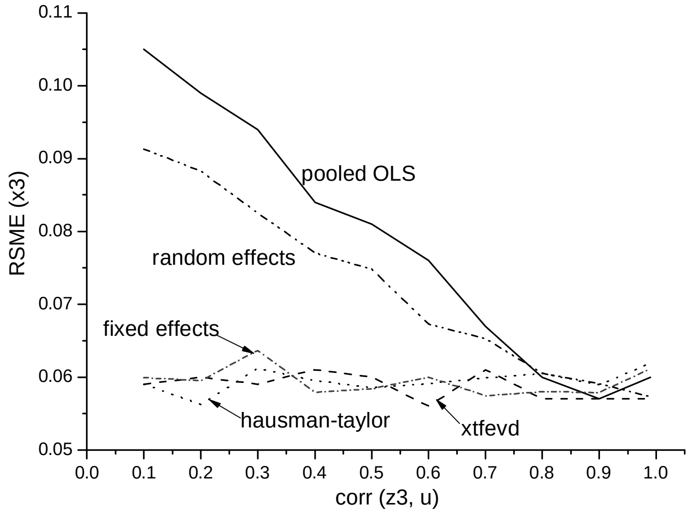
<table>
<tr><td></td><td>Change in the correlation between the time-invariant variable and the unit effects corr(z3, u_i) affects…</td><td>Change in the correlation between the time-varying variable and the unit effects corr(x3, u_i) affects…</td></tr>
<tr><td>…the RMSE of x3</td><td>

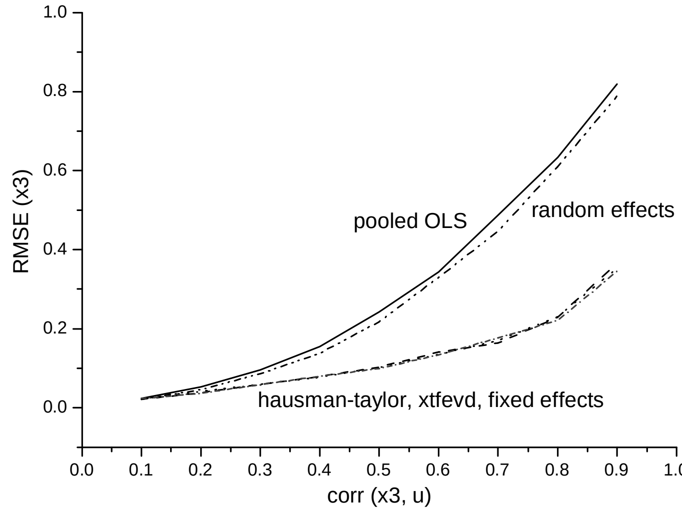
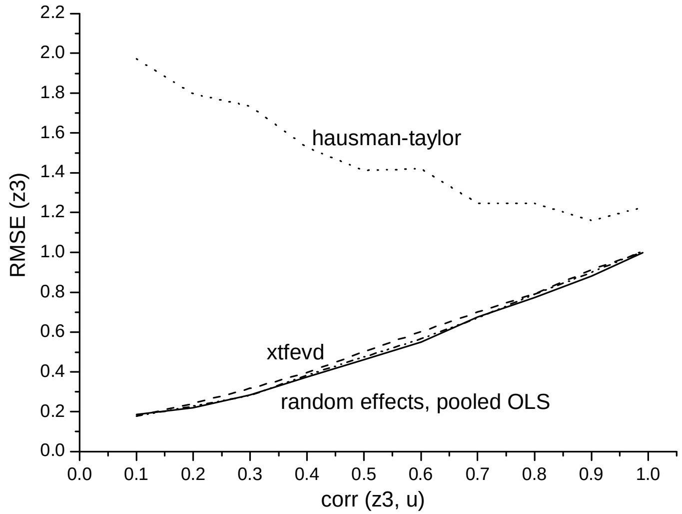
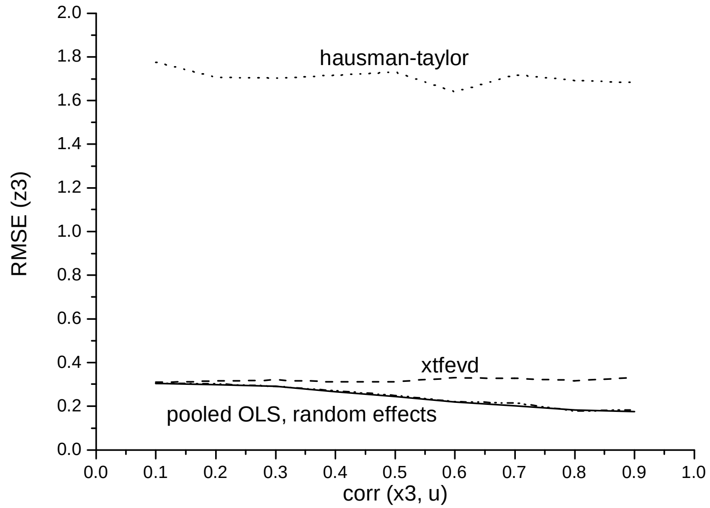
Figure 1: corr(z3, u_i) on RSME(x3)
Parameter settings: N=30, T=20, rho(u_i,x3)=0.3</td><td>

Figure 3: corr(x3, u_i) on RSME(x3)
Parameter settings: N=30, T=20, rho(u_i,z3)=0.3</td></tr>
</table>

Changes in corr(x3, u_i)

	Change in the correlation between the time-invariant variable and the unit effects corr(z3, u_i) affects…	Change in the correlation between the time-varying variable and the unit effects corr(x3, u_i) affects…
…the RMSE of z3	 Figure 2: corr(z3, u_i) on RSME(z3) Parameter settings: N=30, T=20, rho(u_i,x3)=0.3	 Figure 4: corr(x3, u_i) on RSME(z3) Parameter settings: N=30, T=20, rho(u_i,z3)=0.3

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 undoubtedly 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

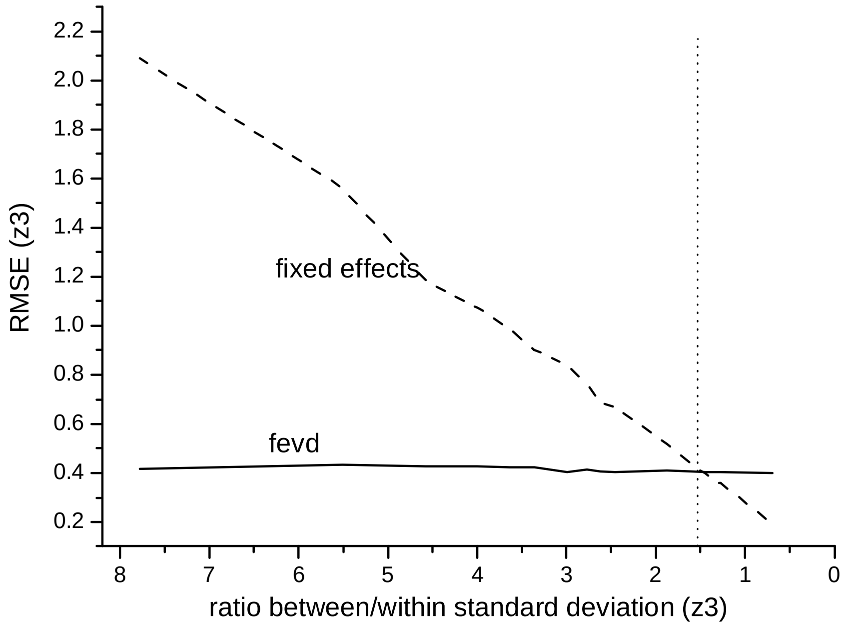
	p-ols		*fevd*		RE		FE	
	RMSE	average bias	RMSE	average bias	RMSE	average bias	RMSE	average bias
time-varying variable x3	0.265	-0.265	0.069	0.001	0.230	-0.230	0.069	0.000
Rarely changing variable z3	0.133	0.028	0.131	0.001	0.133	0.027	0.858	0.008

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)

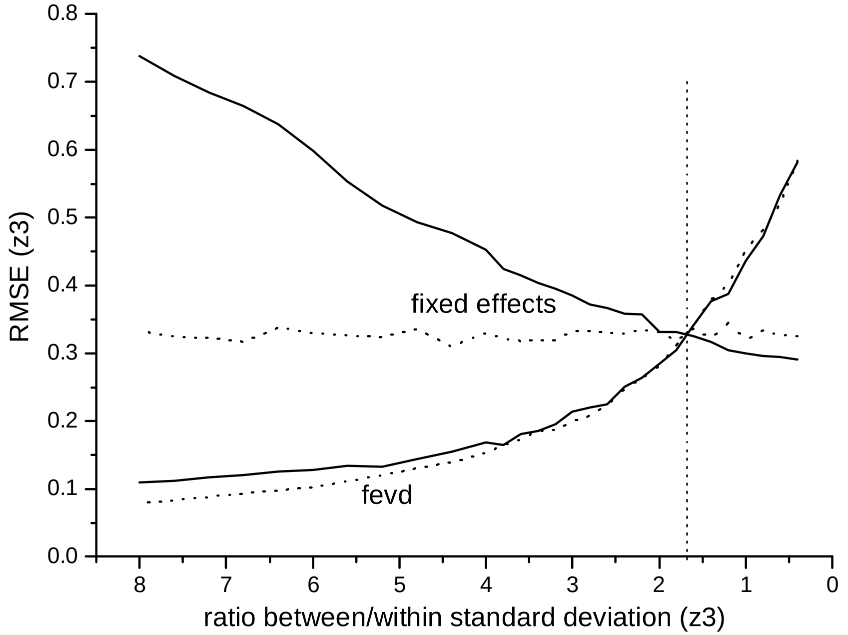
varying the within variance



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

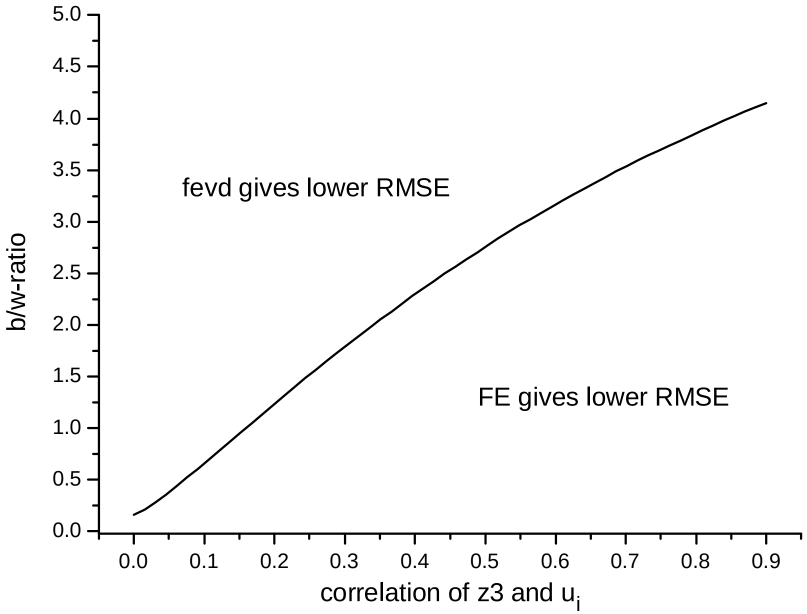
Ratio Between/Within Standard Deviation (z3)

varying the between variance



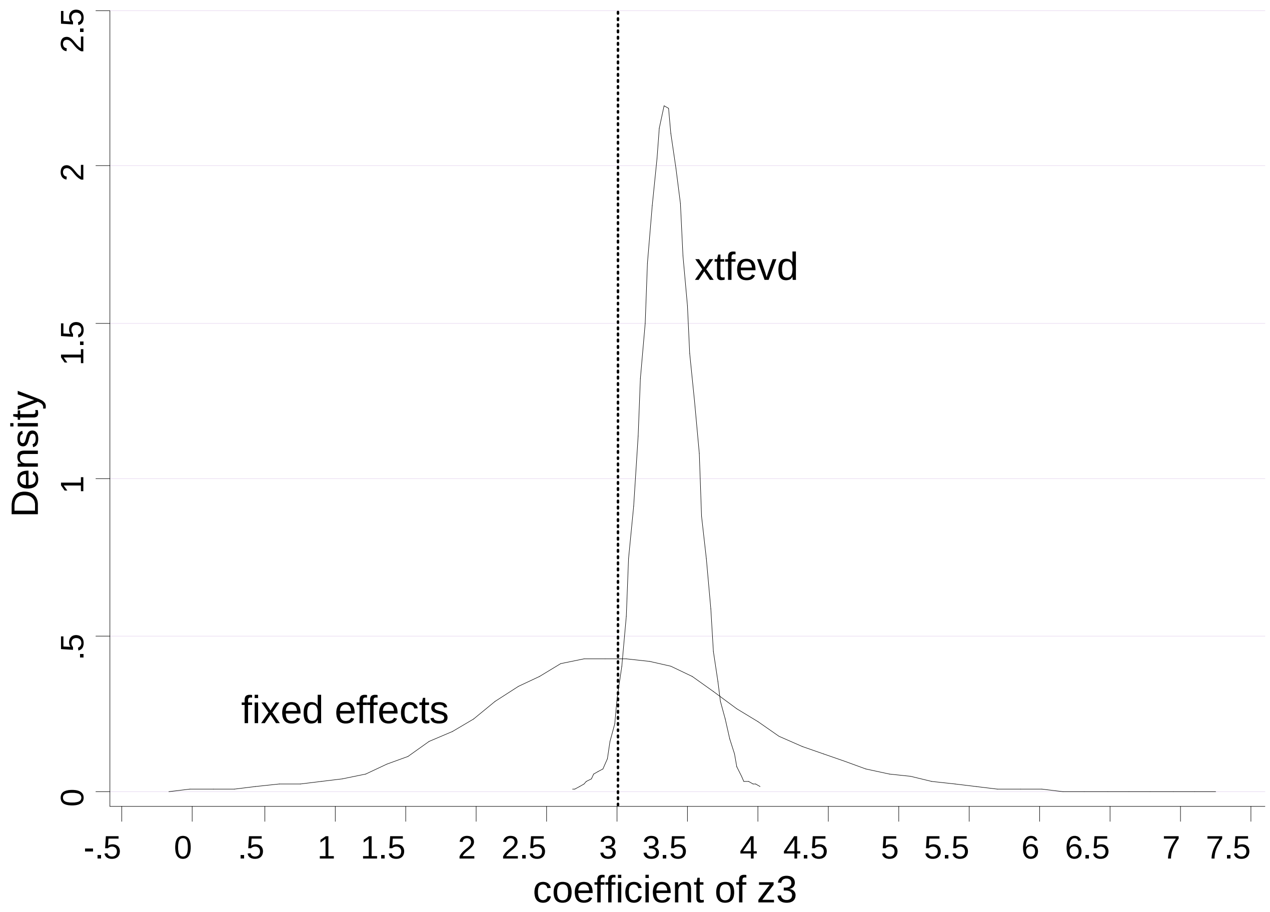
Parameter settings:
N=30
T=20,
rho(u,x3)=0
rho(u,z3)=0.3
between SD (z3): 0.4…8
within SD (z3): 1

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.

Conclusion

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.

Some More Stuff from the 2011 PA Symposium on FEVD

Breusch, Trevor; Ward, Michael B; Nguyen, Hoa and Kompas, Tom: „On the Fixed Effects Vector Decomposition". Political Analysis.

Greene, William: Fixed Effects Vector Decomposition: A Magical Solution to the Problem of Time Invariant Variables in Fixed Effects Models? Political Analysis.

Criticism:

1. Standard Errors – too small (the 2007 PA paper doesn't even talk about SEs; they use the SE computation of the fevd.ado beta version

2. Greene: efficiency gains are illusory

3. BWNK: propose 2 estimators in order to improve over FEVD, a pre-test estimator and a shrinkage estimator

The IV option of FEVD

FEVD has an option that allows instrumenting correlated time invariant variables in the second stage.
If only internal instruments are available, FEVD-IV equals HT and outperforms HT with valid external instruments.

Bias and Efficiency of HT and IV-FEVD with Valid Internal and External Instruments

bias and SD beta of z2	internal instruments		external instruments	
	IV-FEVD	HT	IV-FEVD	HT
T=20 N=30	0.886	0.886	0.990	0.858
	(1.081)	(1.081)	(0.316)	(0.743)
T=20 N=100	0.971	0.971	1.001	0.976
	(0.241)	(0.241)	(0.154)	(0.233)
T=100 N=30	0.923	0.923	1.008	0.875
	(0.755)	(0.755)	(0.294)	(0.674)

Bias: mean β (true beta=1.0) (values closer to 1.0 indicate less bias)
Efficiency: in parentheses: standard deviation of the β (smaller values indicate higher efficiency)
MC setup: DGP: y=x1+x2+z1+z2+u+eps, corr(z2,u)=0.5, corr(z2,x1)=0.5, corr(z2,ext. instr.)=0.5, all other correlations are set to zero, all variables are drawn from a standard normal distribution

BWNK Pre-test Estimator

The pre-test estimator suggested by BWNK utilizes a Durbin-Wu-Hausman test (DWH-test) for the time-invariant variable

$$y_{it} = \beta_1 x_{it}^1 + \beta_2 x_{it}^2 + \gamma_1 z_i^1 + \gamma_2 z_i^2 + u_i + \varepsilon_{it}$$

1. step:

$$z_i^2 = \delta_1 x_{it}^1 + \delta_2 z_i^1 + zresid_i$$

2. step:

$$y_{it} = \beta_1 x_{it}^1 + \beta_2 x_{it}^2 + \gamma_1 z_i^1 + \gamma_2 z_i^2 + \phi zresid_i + \xi_{it} .$$

if ϕ is significant z^2 is indeed endogenous to the time invariant part of the error term u.

this test is reliable only if the instruments are perfectly valid.

The test rapidly loses power if the instruments are slightly correlated with the unobserved unit effects.

Whether the instruments are correlated with the unobserved unit effects is as unknown as whether the time-invariant variable of interest is correlated with the unit effects.

The Ratio of DWH Pre-Tests Suggesting Endogeneity of z2 with 95-percent Confidence

validity of instrument $corr(u,\bar{x}_1)$	strength of instrument $corr(z_2,\bar{x}_1)$	endogeneity $corr(z_2,u)$ 0	0.1	0.3	0.5	0.7	0.9
0	0	0.443	0.428	0.417	0.405	0.361	0.211
	0.1	0.413	0.459	0.433	0.400	0.400	0.386
	0.3	0.440	0.445	0.488	0.574	0.730	0.950
	0.5	0.478	0.470	0.596	0.796	0.978	..
	0.7	0.461	0.523	0.813	0.991	1.000	..
0.1	0	0.492	0.457	0.518	0.472	0.444	0.384
	0.1	0.475	0.478	0.466	0.417	0.360	0.207
	0.3	0.500	0.480	0.423	0.454	0.487	0.715
	0.5	0.537	0.460	0.441	0.612	0.867	1.000
	0.7	0.581	0.505	0.591	0.902	1.000	..
0.3	0	0.776	0.774	0.760	0.802	0.873	0.976
	0.1	0.756	0.756	0.740	0.720	0.763	0.832
	0.3	0.789	0.724	0.646	0.555	0.461	0.247
	0.5	0.866	0.755	0.593	0.432	0.406	0.684
	0.7	0.915	0.842	0.539	0.463	0.848	..
0.5	0	0.963	0.975	0.965	0.980	0.999	..
	0.1	0.977	0.973	0.972	0.970	0.987	1.000
	0.3	0.986	0.966	0.946	0.908	0.863	0.892
	0.5	0.987	0.981	0.924	0.814	0.619	0.310
	0.7	1.000	0.996	0.930	0.712	0.390	0.766
0.7	0	0.997	0.999	1.000	1.000	1.000	..
	0.1	0.999	1.000	0.999	0.999	1.000	..
	0.3	0.999	0.999	0.999	1.000	0.996	..
	0.5	1.000	1.000	0.999	0.987	0.983	0.964
	0.7	1.000	1.000	1.000	0.984	0.898	0.457

DGP: y=x1+x2+z1+z2+u+eps: HT - endogenous(z2), instruments(x1), z1-random, x2-random; N/T=20/30

grey-shaded cell describe conditions under which DWH works reliably

BWNK Shrinkage Estimator

Shrinkage estimator: aims at combining FEVD's efficiency and HT's assumed unbiasedness.

Weighted average of FEVD and HT: $Shrinkage = FEVD + w\left(HT - FEVD\right)$

BWNK use the following weight:

$$w = \frac{\mu_{fevd}^2 + \sigma_{fevd}^2 - \sigma_{fevd,ht}}{\mu_{fevd}^2 + \sigma_{fevd}^2 + \sigma_{ht}^2 - 2\sigma_{fevd,ht}}$$; μ - bias, σ - variance

BWNK use empirical estimates for the variance and covariance which are readily available from the IV variance equation

Bias of FEVD is computed as the difference between the HT and FEVD point estimates

If the correlation between the instruments and the unobserved unit effects exceeds the correlation between the time-invariant variable and the unobserved unit effects, the Hausman-test detects a significant difference in the estimation result and the weighting formula will put a larger weight on the HT model, despite the fact that it is less efficient *and* more biased.

FEVD vs. Shrinkage

MC experiments – DGP: $y_{it} = \beta_1 x_{it}^1 + \beta_2 x_{it}^2 + \gamma_1 z_i^1 + \gamma_2 z_i^2 + u_i + \varepsilon_{it}$

Three correlations (and the number of observations) influence the relative performance of FEVD, HT, the pre-test-model and the shrinkage estimator:

1. The correlation between the time invariant variable and the unit effects, corr(z,u). This correlation cannot be observed.

2. The correlation between the instruments and the unit effects, corr(m,u). This correlation cannot be observed.

3. The correlation between the instruments and the time-invariant variable, corr(m,z). This correlation can be observed.

Difference in the RMSE between FEVD and Shrinkage Estimator whith perfect validity

		strengths				
		0	0.1	0.3	0.5	0.7
endogeneity	0.00	-0.159	-0.152	-0.137	-0.112	-0.041
	0.10	-0.152	-0.159	-0.130	-0.082	-0.030
	0.30	-0.089	-0.073	-0.034	0.040	0.081
	0.50	-0.037	-0.007	0.113	0.207	0.289
	0.70	-0.002	0.054	0.283	0.410	0.488
	0.90	-0.002	0.072	0.459	..	..

corr(m,u)=0.3

		strengths				
		0	0.1	0.3	0.5	0.7
endogeneity	0.00	-0.475	-0.446	-0.439	-0.346	-0.240
	0.10	-0.435	-0.466	-0.403	-0.322	-0.213
	0.30	-0.361	-0.393	-0.359	-0.231	-0.094
	0.50	-0.237	-0.288	-0.244	-0.074	0.049
	0.70	-0.135	-0.233	-0.159	0.075	0.238
	0.90	-0.038	-0.182	-0.054	..	..

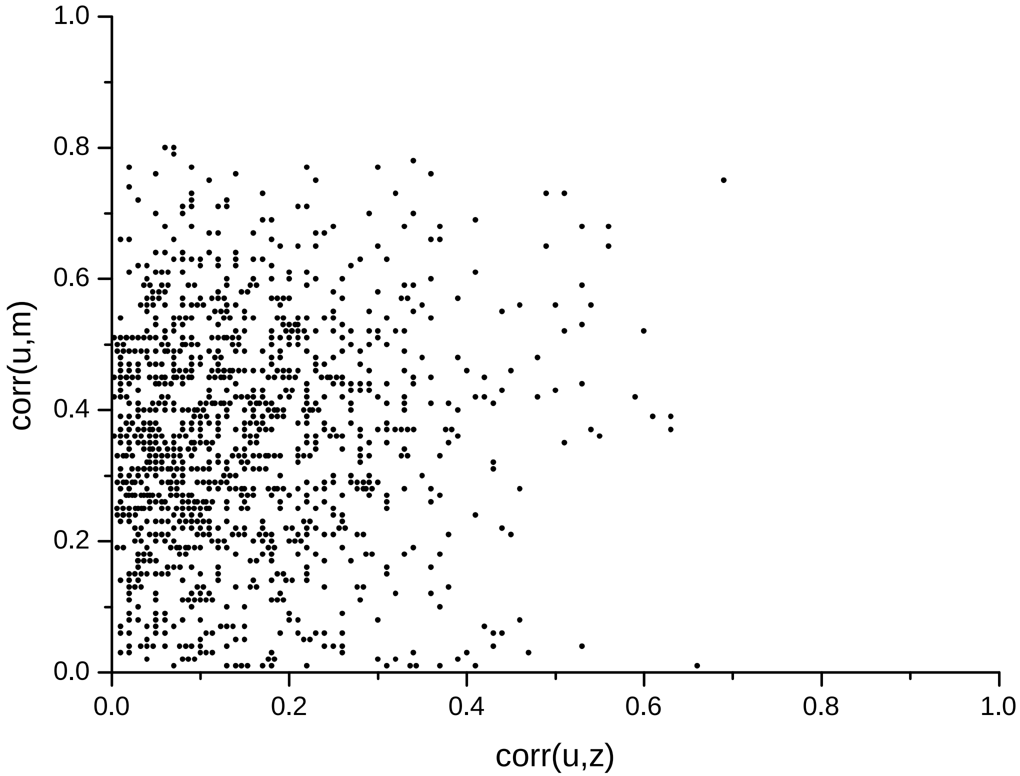
corr(m,u)=0.7

		strengths				
		0	0.1	0.3	0.5	0.7
endogeneity	0.00	-1.299	-1.331	-1.295	-1.086	-0.776
	0.10	-1.258	-1.295	-1.308	-1.060	-0.761
	0.30	-1.165	-1.216	-1.227	-0.927	-0.633
	0.50	-0.971	-1.084	-1.080	-0.767	-0.452
	0.70	-0.852	-0.925	-0.954	-0.619	-0.278
	0.90	..	..	..	..	..

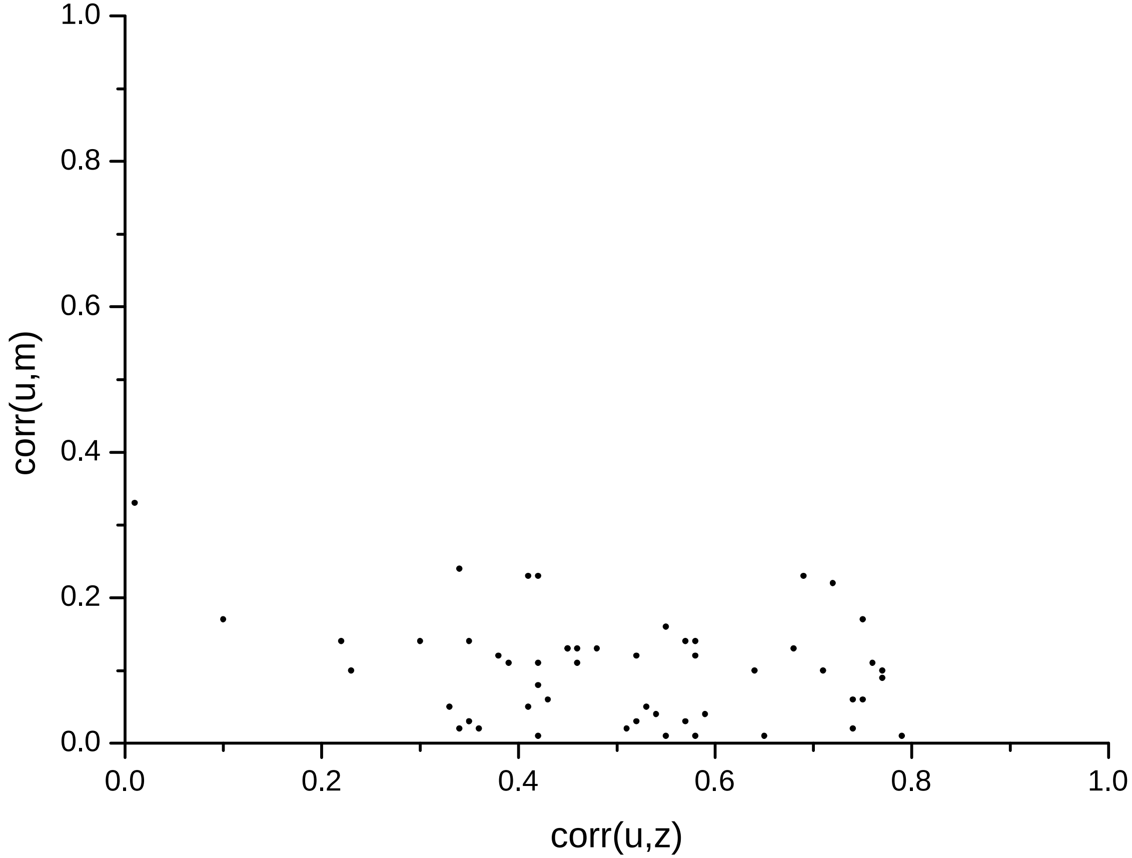
MC Experiment: Random Draw of Correlations

MC Experiment 2: Random Draw of Correlations from an Approximation of the Probability Density Function of Correlations of Random Variables

Constellations in which FEVD significantly outperforms BWNK’s shrinkage model (N=30, T=20)



Constellations in which BWNK's shrinkage model significantly outperforms FEVD (N=30, T=20)



5000 models: N=30 and T=20, FEVD significantly better in 741 cases, shrinkage significantly better in 37 cases. N=100, T=20: 773: 21; T=100, N=20: 710:38. FEVD is more reliable in between 95.2 (N20, T=30), 97.3 (N=100, T=20), 94.9 (N=20, T=100), 94.4 (N=20, T=500) percent.

Standard Errors

FEVD variance formula:

$$V_{FEVD4.0}(\beta,\gamma) = (H'W)^{-1} H'\Omega H (W'H)^{-1}$$

$$H = [\ddot{X}, Z]; \quad W = [X, Z]$$

$$\Omega = \sigma_{\varepsilon}^2 I_{NT} + \sigma_{\hat{\eta}}^2 I_N \otimes \iota_T \iota_T'$$

where I_N is an N x N identity matrix and ι_T is a T x 1 vector of ones.

BWNK:

$$\Omega_{BWNK} = \sigma_{\varepsilon}^2 I_{NT} + \sigma_{\hat{u}}^2 I_N \otimes \iota_T \iota_T'$$

Greene:

$$V_{GREENE}(\gamma) = (Z'Z)^{-1} Z'\Omega Z (Z'Z)^{-1}$$

$$\Omega_{GREENE} = \sigma_{\eta}^2 + \sigma_{\varepsilon}^2 \left\{ \frac{1}{T} + \bar{x}_i' \left[X'\ddot{X} \right]^{-1} \bar{x}_i \right\}$$

where $\ddot{X} = x_{it} - \frac{1}{T}\sum_{t=1}^{T} x_{it}$

MC Experiments

Decision criterion:

$$underconfidence = 100\frac{\sqrt{\sum_{k=1}^{K}\left(s.e.\left(\hat{\beta}^{k}\right)\right)^{2}}}{\sqrt{\sum_{k=1}^{K}\left(\hat{\beta}^{k}-\overline{\hat{\beta}}\right)^{2}}}$$

Under-/Overconfidence as a Function of N and T

SD(u)=1; corr(x,z,u)=0 number of obs		mean[se(z)] / SD[beta(z)] FEVD4.0	BWNK	GREENE	pooled ols
N	T				
	10	91	121	102	56
10	30	86	124	92	37
	50	83	124	87	29
	70	82	121	85	25
	100	83	126	85	22
	10	108	137	119	59
30	30	98	133	103	37
	50	100	139	103	31
	70	103	146	106	28
	100	96	138	98	22
	10	106	135	117	58
50	30	99	134	104	37
	50	99	137	102	30
	70	100	140	103	26
	100	99	140	101	22
	10	111	139	122	59
70	30	107	145	112	40
	50	105	145	109	31
	70	100	139	102	26
	100	98	138	100	22
	10	112	140	123	60
100	30	104	140	109	38
	50	99	137	103	30
	70	104	144	107	27
	100	100	140	102	22

Under-/Overconfidence as a Function of var(u) with constant var(x): var(z)

		mean[se(z)] / SD[beta(z)] FEVD	BWNK	GREENE	Pooled OLS
SD(u)					
0.5	N=20	104	183	118	58
1	T=30	98	137	104	38
1.5	corr(x,u)=0	92	113	94	28
2	corr(z,u)=0	86	101	88	23
3		92	102	93	21
4		88	96	89	19
5		89	95	89	18
0.5	N=20	103	204	115	55
1	T=30	94	160	99	35
1.5	corr(x,u)=0.5	93	143	95	27
2	corr(z,u)=0.5	92	134	93	23
3		87	119	88	19
4		94	124	94	19
5		93	120	93	19

Under-/Overconfidence as function of corr(x,u) and corr(z,u)

N;T=20;30; SD(u)=1		mean[se(z)] / SD[beta(z)]			
corr(x,u)	corr(z,u)	FEVD	BWNK	GREENE	Pooled-OLS
0	0.1	96	138	103	41
	0.3	94	155	105	42
	0.5	97	176	107	45
	0.7	103	206	107	48
	0.9	112	318	131	69
0.1	0.1	94	130	99	37
	0.3	99	147	101	38
	0.5	98	167	101	40
	0.7	96	215	110	48
	0.9	112	320	127	66
0.3	0.1	97	125	96	33
	0.3	98	136	95	33
	0.5	96	161	99	37
	0.7	104	202	103	42
	0.9	110	317	122	61
0.5	0.1	95	121	96	31
	0.3	92	136	97	32
	0.5	102	164	102	36
	0.7	100	208	107	41
	0.9	101	304	114	55
0.7	0.1	93	115	92	28
	0.3	93	128	92	29
	0.5	97	153	96	32
	0.7	98	195	101	36
	0.9	106	302	111	51
0.9	0.1	98	115	92	28
	0.3	98	128	92	29
	0.5	102	153	96	32
	0.7	99	195	101	36
	0.9	108	302	111	51

What do we learn?

Statistically testing for endogeneity is a non-starter because of the infinite regress

Careful theoretical justification of exogeneity of employed variation remains indispensable

Thinking about design and identification strategies is necessary but not sufficient for making valid inferences

Using exclusively variation over time but ignoring dynamic specification can exacerbate the problem → quick fixes do not work

Statistical inference is plagued by trade-offs: validity-strength, unbiasedness-efficiency, internal-external validity etc. that cannot be ignored when only limited data is available (i.e. we don’t live in asymptotia)

Study design and model specification are equally important