

GSERM - St. Gallen 2023

Analyzing Panel Data

June 14, 2023

Generalized Least Squares Models

Start with a focus on residuals... For:

$$Y_{it} = \mathbf{X}_{it}\beta + u_{it}$$

i.i.d. OLS u_{it} s require:

$$\begin{aligned} \mathbf{u}\mathbf{u}' \equiv \mathbf{\Omega} &= \sigma^2 \mathbf{I} \\ &= \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} \end{aligned}$$

This means that within units:

- $\text{Var}(u_{it}) = \text{Var}(u_{is}) \forall t \neq s$ (temporal homoscedasticity)
- $\text{Cov}(u_{it}, u_{is}) = 0 \forall t \neq s$ (no within-unit autocorrelation)

and between units:

- $\text{Var}(u_{it}) = \text{Var}(u_{jt}) \forall i \neq j$ (cross-unit homoscedasticity)
- $\text{Cov}(u_{it}, u_{jt}) = 0 \forall i \neq j$ (no between-unit / spatial correlation)

Estimator:

$$\hat{\beta}_{GLS} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{Y}$$

with:

$$\widehat{V(\beta_{GLS})} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}$$

Two approaches:

- Use OLS \hat{u}_{it} s to get $\hat{\Omega}$ (“feasible GLS” / “weighted least squares”)
- Use substantive knowledge about the data to structure Ω

The variance-covariance matrix is:

$$\begin{aligned}\text{Var}(\hat{\beta}_{WLS}) &= \sigma^2(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1} \\ &\equiv (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\end{aligned}$$

A common case is:

$$\text{Var}(u_i) = \sigma^2 \frac{1}{N_i}$$

where N_i is the number of observations upon which (aggregate) observation i is based.

“Robust” Variance Estimators

Recall that, if $\sigma_i^2 \neq \sigma_j^2 \forall i \neq j$,

$$\begin{aligned}\text{Var}(\beta_{\text{Het.}}) &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{Q} (\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

where $\mathbf{Q} = (\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})$ and $\mathbf{W} = \sigma^2\mathbf{\Omega}$.

We can rewrite \mathbf{Q} as

$$\begin{aligned}\mathbf{Q} &= \sigma^2(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X}) \\ &= \sum_{i=1}^N \sigma_i^2 \mathbf{x}_i \mathbf{x}_i'\end{aligned}$$

Estimate $\hat{\mathbf{Q}}$ as:

$$\hat{\mathbf{Q}} = \sum_{i=1}^N \hat{u}_i^2 \mathbf{x}_i \mathbf{x}_i'$$

Yields:

$$\begin{aligned} \widehat{\text{Var}}(\hat{\beta})_{\text{Robust}} &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\hat{\mathbf{Q}}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1} \left[\mathbf{X}' \left(\sum_{i=1}^N \hat{u}_i^2 \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \mathbf{X} \right] (\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

“Robust” VCV estimates:

- are heteroscedasticity-consistent, but
- are biased in small samples, and
- are less efficient than “naive” estimates when $\text{Var}(u) = \sigma^2 \mathbf{I}$.

“Clustering”

Huber / White

?????????

WLS / GLS

I know very little
about my error
variances...

I know a great
deal about my
error variances...

A common case:

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

with

$$\sigma_{ij}^2 = \sigma_{ik}^2.$$

“Robust, clustered” estimator:

$$\widehat{\text{Var}}(\beta)_{\text{Clustered}} = (\mathbf{X}'\mathbf{X})^{-1} \left\{ \mathbf{X}' \left[\sum_{i=1}^N \left(\sum_{j=1}^{n_j} \hat{u}_{ij}^2 \mathbf{X}_{ij} \mathbf{X}_{ij}' \right) \right]^{-1} \mathbf{X} \right\} (\mathbf{X}'\mathbf{X})^{-1}$$

Robust / Clustered SEs: A Simulation

```
url_robust <- "https://raw.githubusercontent.com/IsidoreBeautrelet/economictheoryblog/master/robust_summary.R"
eval(parse(text = getURL(url_robust, ssl.verifypeer = FALSE)),
      envir=.GlobalEnv)
```

```
> set.seed(3844469)
> X <- rnorm(10)
> Y <- 1 + X + rnorm(10)
> df10 <- data.frame(ID=seq(1:10),X=X,Y=Y)
> fit10 <- lm(Y~X,data=df10)
> summary(fit10)
```

```
Call:
lm(formula = Y ~ X, data = df10)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-1.318 -0.766  0.195  0.378  1.590
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.954      0.311    3.06   0.016 *
X             0.589      0.291    2.03   0.077 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
```

```
Residual standard error: 0.985 on 8 degrees of freedom
Multiple R-squared:  0.339, Adjusted R-squared:  0.257
F-statistic: 4.11 on 1 and 8 DF, p-value: 0.0772
```

```
> rob10 <- vcovHC(fit10,type="HC1")
> sqrt(diag(rob10))
(Intercept)          X
    0.315         0.285
```

Robust / Clustered SEs: A Simulation (continued)

```
> # "Clone" each observation 100 times:
>
> df1K <- df10[rep(seq_len(nrow(df10)),each=100),]
> df1K <- pdata.frame(df1K, index="ID")
> fit1K <- lm(Y~X,data=df1K)

> summary(fit1K)

Call:
lm(formula = Y ~ X, data = df1K)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.9536     0.0279   34.2   <2e-16 ***
X              0.5893     0.0260   22.6   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.882 on 998 degrees of freedom
Multiple R-squared:  0.339, Adjusted R-squared:  0.339
F-statistic: 513 on 1 and 998 DF, p-value: <2e-16

> summary(fit1K, cluster="ID")

Call:
lm(formula = Y ~ X, data = df1K)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.954     0.297   3.21  0.0014 **
X              0.589     0.269   2.19  0.0286 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.882 on 998 degrees of freedom
Multiple R-squared:  0.339, Adjusted R-squared:  0.339
F-statistic: 4.8 on 1 and 9 DF, p-value: 0.0561
```

Example:

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

$$u_t = \rho u_{t-1} + e_t$$

with $e_t \sim i.i.d. N(0, \sigma_u^2)$ and $\rho \in [-1, 1]$ (typically).

→ “First-order autoregressive” (“AR(1)”) errors.

Serially Correlated Errors and OLS

Detection

- *Plot* of residuals vs. lagged residuals
- *Runs* test (Geary test)
- Durbin-Watson d
 - Calculated as:

$$d = \frac{\sum_{t=2}^N (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^N \hat{u}_t^2}$$

- Non-standard distribution ($d \in [0, 4]$)
- Null: No autocorrelation
- Only detects first-order autocorrelation

Serially Correlated Errors and OLS

What to do about it?

- GLS, incorporating ρ / $\hat{\rho}$ into the equation
- First-difference equations (regressing changes of Y on changes of \mathbf{X})
- Cochrane-Orcutt / Prais-Winsten:
 1. Estimate the basic equation via OLS, and obtain residuals
 2. Use the residuals to consistently estimate $\hat{\rho}$ (i.e. the empirical correlation between u_t and u_{t-1})
 3. Use this estimate of $\hat{\rho}$ to estimate the *difference equation*:

$$(Y_t - \rho Y_{t-1}) = \beta_0(1 - \rho) + \beta_1(X_t - \rho X_{t-1}) + (u_t - \rho u_{t-1})$$

4. Save the residuals, and use them to estimate $\hat{\rho}$ again
5. Repeat this process until successive estimates of $\hat{\rho}$ differ by a very small amount

Running Example Redux

The World Development Indicators:

- Cross-national country-level time series data
- $N = 215$ countries, $T = 72$ years (1960-2021) + missingness
- Full descriptions are listed in the Github repo [here](#)

Regression model:

$$\text{WBLI}_{it} = \beta_0 + \beta_1 \text{Population Growth}_{it} + \beta_2 \text{Urban Population}_{it}^2 + \beta_3 \text{Fertility Rate}_{it} + \beta_4 \ln(\text{GDP Per Capita})_{it} + \beta_5 \text{Natural Resource Rents}_{it} + \beta_6 \text{Cold War}_t + u_{it}$$

Descriptive Statistics:

	vars	n	mean	sd	min	max	range	se
WomenBusLawIndex	1	8100	60.69	18.95	17.50	100.00	82.50	0.21
PopGrowth	2	8100	1.65	1.54	-16.88	19.36	36.24	0.02
UrbanPopulation	3	8100	51.56	23.82	2.85	100.00	97.16	0.26
FertilityRate	4	8100	3.61	1.90	0.77	8.61	7.83	0.02
NaturalResourceRents	5	8100	7.04	10.77	0.00	88.59	88.59	0.12
ColdWar	6	8100	0.30	0.46	0.00	1.00	1.00	0.01
lnGDPPerCap	7	8100	8.29	1.44	5.04	11.64	6.60	0.02

How Much Autocorrelation in **X**?

Note that:

$$d = 2(1 - \rho)$$

which means that we can calculate:

$$\rho = 1 - \frac{d}{2}.$$

Autocorrelation in the Predictors

	Variable	Rho
1	Population Growth	0.852
2	Urban Population	0.974
3	Fertility Rate	0.966
4	GDP Per Capita	0.977
5	Natural Resource Rents	0.911
6	Cold War	0.916

Baseline Model: OLS (+ D-W Test)

```
> OLS<-plm(WomenBusLawIndex~PopGrowth+UrbanPopulation+FertilityRate+log(GDPPerCapita)+NaturalResourceRents+ColdWar,
+ data=WDI,model="pooling")
> summary(OLS)
.
.
.
Coefficients:
                Estimate Std. Error t-value    Pr(>|t|)
(Intercept)      60.4325    1.6861    35.8    < 2e-16 ***
PopGrowth        -2.3630    0.1306   -18.1    < 2e-16 ***
UrbanPopulation  -0.0587    0.0105    -5.6 0.000000022 ***
FertilityRate     -2.5215    0.1592   -15.8    < 2e-16 ***
log(GDPPerCapita)  2.6533    0.1936    13.7    < 2e-16 ***
NaturalResourceRents -0.3398    0.0155   -21.9    < 2e-16 ***
ColdWar          -10.9584    0.3715   -29.5    < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    2910000
Residual Sum of Squares: 1450000
R-Squared:                0.501
Adj. R-Squared: 0.501
F-statistic: 1354.19 on 6 and 8093 DF, p-value: <2e-16

> pdwtest(OLS)

Durbin-Watson test for serial correlation in panel models

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate + ...
DW = 0.14, p-value <2e-16
alternative hypothesis: serial correlation in idiosyncratic errors
```

Example: Prais-Winsten

```
> PraisWinsten<-panelAR(WomenBusLawIndex~PopGrowth+UrbanPopulation+
+      FertilityRate+log(GDPPerCapita)+NaturalResourceRents+
+      ColdWar, data=WDI,panelVar="ISO3",timeVar="YearNumeric",
+      autoCorr="ar1",panelCorrMethod="none",
+      rho.na.rm=TRUE)

> summary(PraisWinsten)

Panel Regression with AR(1) Prais-Winsten correction and homoskedastic variance

Unbalanced Panel Design:
Total obs.:      8100 Avg obs. per panel 43.316
Number of panels: 187 Max obs. per panel 52
Number of times: 52  Min obs. per panel 1

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   70.366002   2.977666   23.63 < 2e-16 ***
PopGrowth     -0.038660   0.039708   -0.97  0.33028
UrbanPopulation -0.000807   0.024026   -0.03  0.97321
FertilityRate  -5.250651   0.233565  -22.48 < 2e-16 ***
log(GDPPerCapita) 1.237712   0.345964    3.58  0.00035 ***
NaturalResourceRents -0.008431  0.007950   -1.06  0.28892
ColdWar        -0.965160   0.216238   -4.46  0.0000082 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-squared: 0.3266
Wald statistic: 1094.4684, Pr(>Chisq(6)): 0

> PraisWinsten$panelStructure$rho
[1] 0.9521
```

Better in a Table

	OLS	Prais-Winsten
Intercept	60.43* (1.69)	70.36* (2.98)
Population Growth	-2.36* (0.13)	-0.04 (0.04)
Urban Population	-0.06* (0.01)	-0.0008 (0.02)
Fertility Rate	-2.52* (0.16)	-5.25* (0.24)
ln(GDP Per Capita)	2.65* (0.19)	1.24* (0.35)
Natural Resource Rents	-0.34* (0.02)	-0.01 (0.008)
Cold War	-10.96* (0.37)	-0.97* (0.22)
$\hat{\rho}$		0.95
R ²	0.501	0.33
Adj. R ²	0.501	
<i>NT</i>	8100	8100
<i>N</i> panels		187

* $p < 0.05$

Some Panel Data Challenges

Consider the error terms in the model:

$$Y_{it} = \mathbf{X}_{it}\beta + u_{it}$$

Issues:

<u>In Words:</u>	<u>In a Formula:</u>
<u>Variances:</u>	
Unit-Wise Heteroscedasticity	$\text{Var}(u_{it}) \neq \text{Var}(u_{jt})$
Temporal Heteroscedasticity	$\text{Var}(u_{it}) \neq \text{Var}(u_{is})$
<u>Covariances:</u>	
Contemporary Cross-Unit Correlation	$\text{Cov}(u_{it}, u_{jt}) \neq 0$
Within-Unit Serial Correlation	$\text{Cov}(u_{it}, u_{is}) \neq 0$
Non-Contemporaneous Cross-Unit Correlation	$\text{Cov}(u_{it}, u_{js}) \neq 0$

Parks' (1967) Approach

Assume:

- $\text{Var}(u_{it}, u_{jt}) = \sigma^2$ or σ_i^2 (Common or unit-specific error variances)
- $\text{Var}(u_{it}) = \text{Var}(u_{is}) \forall t \neq s$ (Temporal homoscedasticity)
- $\text{Cov}(u_{it}, u_{jt}) = \sigma_{ij} \forall i \neq j$ (Pairwise contemporaneous cross-unit correlation)
- $\text{Cov}(u_{it}, u_{is}) = \rho$ or ρ_i (Common or unit-specific temporal correlation)
- $\text{Cov}(u_{it}, u_{js}) = 0 \forall i \neq j, t \neq s$ (No non-contemporaneous cross-unit correlation)

(B&K: “panel error assumptions”).

Then:

1. Use OLS to generate $\hat{u}s \rightarrow \hat{\rho} (\rightarrow \hat{\Omega})$,
2. Use $\hat{\rho}$ for Prais-Winsten.

This method was widely used prior to Beck & Katz (1995)

$$\mathbf{\Omega} = \begin{pmatrix} \Sigma & 0 & \cdots & 0 \\ 0 & \Sigma & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma \end{pmatrix} = \Sigma \otimes \mathbf{I}_T$$

where

$$\Sigma_{N \times N} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1N} & \sigma_{2N} & \cdots & \sigma_N^2 \end{pmatrix}$$

Means:

- $\frac{N(N-1)}{2}$ distinct contemporaneous covariances σ_{ij} ,
- NT observations,
- $\rightarrow 2T/(N+1)$ observations per $\hat{\sigma}$

From PROC PANEL in SAS:

Standard Corrections

For the PARKS option, the first-order autocorrelation coefficient must be estimated for each cross section. Let ρ be the $N \times 1$ vector of true parameters and $R = (r_1, \dots, r_N)'$ be the corresponding vector of estimates. Then, to ensure that only range-preserving estimates are used in PROC PANEL, the following modification for R is made:

$$r_i = \begin{cases} r_i & \text{if } |r_i| < 1 \\ \max(.95, r_{\max}) & \text{if } r_i \geq 1 \\ \min(-.95, r_{\min}) & \text{if } r_i \leq -1 \end{cases}$$

where

$$r_{\max} = \begin{cases} 0 & \text{if } r_i < 0 \text{ or } r_i \geq 1 \quad \forall i \\ \max_j[r_j : 0 \leq r_j < 1] & \text{otherwise} \end{cases}$$

and

$$r_{\min} = \begin{cases} 0 & \text{if } r_i > 0 \text{ or } r_i \leq -1 \quad \forall i \\ \max_j[r_j : -1 < r_j \leq 0] & \text{otherwise} \end{cases}$$

Whenever this correction is made, a warning message is printed.

Panel-Corrected Standard Errors

Key to PCSEs:

$$\hat{\sigma}_{ij} = \frac{\sum_{t=1}^T \hat{u}_{it} \hat{u}_{jt}}{T}$$

Define:

$$\mathbf{U}_{T \times N} = \begin{pmatrix} \hat{u}_{11} & \hat{u}_{21} & \cdots & \hat{u}_{N1} \\ \hat{u}_{12} & \hat{u}_{22} & \cdots & \hat{u}_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{u}_{1T} & \hat{u}_{2T} & \cdots & \hat{u}_{NT} \end{pmatrix}$$

$$\hat{\Sigma} = \frac{(\mathbf{U}'\mathbf{U})}{T}$$

$$\hat{\Omega}_{PCSE} = \frac{(\mathbf{U}'\mathbf{U})}{T} \otimes \mathbf{I}_T$$

Panel-Corrected Standard Errors

Correct formula:

$$\text{Cov}(\hat{\beta}_{PCSE}) = (\mathbf{X}'\mathbf{X})^{-1}[\mathbf{X}'\mathbf{\Omega}\mathbf{X}](\mathbf{X}'\mathbf{X})^{-1}$$

General Issues:

- PCSEs do not fix unit-level heterogeneity (a la “fixed” / “random” effects)
- They also do not deal with dynamics
- They depend critically on the “panel data assumptions” of Park / Beck & Katz

Panel Assumptions and Numbers of Parameters

Panel Assumptions	No AR(1)	Common $\hat{\rho}$	Separate $\hat{\rho}_i$ s
$\sigma_i^2 = \sigma^2, \text{Cov}(\sigma_{it}, \sigma_{jt}) = 0$	$k + 1$	$k + 2$	$k + N + 1$
$\sigma_i^2 \neq \sigma^2, \text{Cov}(\sigma_{it}, \sigma_{jt}) = 0$	$k + N$	$k + N + 1$	$k + 2N$
$\sigma_i^2 \neq \sigma^2, \text{Cov}(\sigma_{it}, \sigma_{jt}) \neq 0$	$\frac{N(N-1)}{2} + k + N$	$\frac{N(N-1)}{2} + k + N + 1$	$\frac{N(N-1)}{2} + k + 2N$

Example: GLS with Homoscedastic AR(1) Errors

```
> GLS<-glS(WomenBusLawIndex~PopGrowth+UrbanPopulation+FertilityRate+log(GDPPerCapita)+NaturalResourceRents+ColdWar,
+          data=WDI,correlation=corAR1(form=~1|ISO3),na.action="na.omit")
>
> summary(GLS)
Generalized least squares fit by REML
Model: WomenBusLawIndex~PopGrowth+UrbanPopulation+FertilityRate+log(GDPPerCapita)+NaturalResourceRents+ColdWar
Data: WDI
AIC    BIC logLik
38070 38133 -19026

Correlation Structure: AR(1)
Formula: ~1 | ISO3
Parameter estimate(s):
  Phi
0.9897

Coefficients:
                Value Std.Error t-value p-value
(Intercept)      50.72    3.986   12.724  0.0000
PopGrowth        -0.01    0.037   -0.228  0.8194
UrbanPopulation    0.25    0.041    6.245  0.0000
FertilityRate     -3.64    0.293  -12.422  0.0000
log(GDPPerCapita)  1.37    0.420    3.262  0.0011
NaturalResourceRents 0.02    0.007    2.500  0.0124
ColdWar          -0.46    0.204   -2.261  0.0238

.
.
.

Residual standard error: 16.92
Degrees of freedom: 8100 total; 8093 residual
```

Example: PCSEs

```
> PCSE<-panelAR(WomenBusLawIndex~PopGrowth+UrbanPopulation+FertilityRate+log(GDPPerCapita)+NaturalResourceRents+ColdWar,  
+ data=WDI,panelVar="ISO3",timeVar="YearNumeric",autoCorr="ar1",panelCorrMethod="pcse",rho.na.rm=TRUE)
```

```
> summary(PCSE)
```

Panel Regression with AR(1) Prais-Winsten correction and panel-corrected standard errors

Unbalanced Panel Design:

```
Total obs.:      8100 Avg obs. per panel 43.316  
Number of panels: 187 Max obs. per panel 52  
Number of times:  52  Min obs. per panel  1
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	70.366002	4.482539	15.70	<2e-16 ***
PopGrowth	-0.038660	0.044896	-0.86	0.3892
UrbanPopulation	-0.000807	0.025989	-0.03	0.9752
FertilityRate	-5.250651	0.365619	-14.36	<2e-16 ***
log(GDPPerCapita)	1.237712	0.464299	2.67	0.0077 **
NaturalResourceRents	-0.008431	0.012149	-0.69	0.4877
ColdWar	-0.965160	0.586562	-1.65	0.0999 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

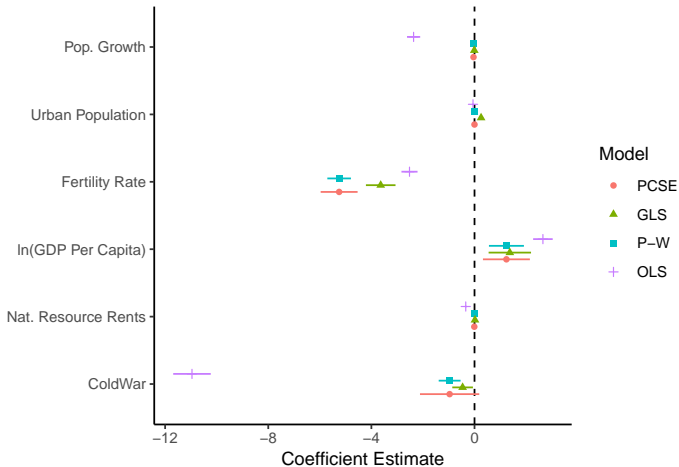
R-squared: 0.3266

Wald statistic: 400.9446, Pr(>Chisq(6)): 0

```
> PCSE$panelStructure$rho
```

```
[1] 0.9521
```

Model Comparisons



Dynamics!

Time Series: Stationarity

Stationarity: A constant d.g.p. over time.¹

Mean stationarity:

$$E(Y_t) = \mu \quad \forall t$$

Variance stationarity:

$$\text{Var}(Y_t) = E[(Y_t - \mu)^2] \equiv \sigma_Y^2 \quad \forall t$$

Covariance stationarity:

$$\text{Cov}(Y_t, Y_{t-s}) = E[(Y_t - \mu)(Y_{t-s} - \mu)] = \gamma_s \quad \forall s$$

¹A stricter form of stationarity requires that the joint probability distribution (in other words, *all* the moments) of series of observations $\{Y_1, Y_2, \dots, Y_t\}$ is the same as that for $\{Y_{1+s}, Y_{2+s}, \dots, Y_{t+s}\}$ for all t and s .

The “ARIMA” Approach

“ARIMA” = *Autoregressive Integrated Moving Average*...

A (first-order) integrated series (“random walk”) is:

$$Y_t = Y_{t-1} + u_t, \quad u_t \sim i.i.d.(0, \sigma_u^2)$$

...a/k/a a “random walk”:

$$\begin{aligned} Y_t &= Y_{t-2} + u_{t-1} + u_t \\ &= Y_{t-3} + u_{t-2} + u_{t-1} + u_t \\ &= \sum_{t=0}^T u_t \end{aligned}$$

I(1) series are not stationary.

Variance:

$$\text{Var}(Y_t) \equiv E(Y_t)^2 = t\sigma^2$$

Autocovariance:

$$\text{Cov}(Y_t, Y_{t-s}) = |t - s|\sigma^2.$$

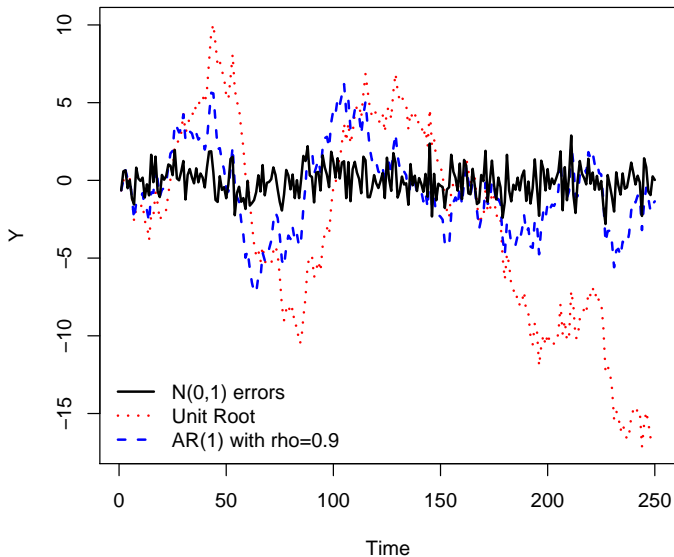
Both depend on t ...

I(1) series (continued)

More generally:

- $|\rho| > 1$
 - Series is nonstationary / *explosive*
 - Past shocks have a greater impact than current ones
 - Uncommon
- $|\rho| < 1$
 - *Stationary* series
 - Effects of shocks die out exponentially according to ρ
 - Is mean-reverting
- $|\rho| = 1$
 - Nonstationary series
 - Shocks persist at full force
 - Not mean-reverting; variance increases with t

Time Series Types, Illustrated



I(1) Series: Differencing

For an I(1) series:

$$Y_t - Y_{t-1} = u_t$$

which we often write in terms of the difference operator Δ (or sometimes ∇):

$$\Delta Y_t = u_t$$

The differenced series is just the (stationary, ergodic) white-noise process u_t .

Unit Root Tests Review: Dickey-Fuller

Two steps:

- Estimate $Y_t = \rho Y_{t-1} + u_t$,
- test the hypothesis that $\hat{\rho} = 1$, *but*
- this requires that the us are uncorrelated.

But suppose:

$$\Delta Y_t = \sum_{i=1}^p d_i \Delta Y_{t-i} + u_t$$

which yields

$$Y_t = Y_{t-1} + \sum_{i=1}^p d_i \Delta Y_{t-i} + u_t.$$

D.F. tests will be incorrect.

Augmented Dickey-Fuller Tests:

- Estimate

$$\Delta Y_t = \rho Y_{t-1} + \sum_{i=1}^p d_i \Delta Y_{t-i} + u_t$$

- Test $\hat{\rho} = 1$

Phillips-Perron Tests:

- Estimate:

$$\Delta Y_t = \alpha + \rho Y_{t-1} + u_t$$

- Calculate modified test statistics (Z_ρ and Z_t)
- Test $\hat{\rho} = 0$

Issues with Unit Roots in Panel Data

- Short series + Asymptotic tests \rightarrow “borrow strength”
- Requires uniform unit roots across cross-sectional units
- Many tests require balanced panels...
- Various alternatives:
 - Maddala and Wu (1999)
 - Hadri (2000)
 - Levin, Lin and Chu (2002)
 - Im, Pesaran, and Shin (2003)
- What to do?
 - Difference the data...
 - Error-correction models

Panel Unit Root Tests: R

```
[data wrangling...]
```

```
> purtest(WBLI.W,exo="trend",test="levinlin",pmax=2)
```

```
Levin-Lin-Chu Unit-Root Test (ex. var.: Individual Intercepts and Trend)
```

```
data: WBLI.W
```

```
z = -2.7, p-value = 0.004
```

```
alternative hypothesis: stationarity
```

```
> purtest(WBLI.W,exo="trend",test="hadri",pmax=2)
```

```
Hadri Test (ex. var.: Individual Intercepts and Trend) (Heterosked. Consistent)
```

```
data: WBLI.W
```

```
z = 197, p-value <2e-16
```

```
alternative hypothesis: at least one series has a unit root
```

```
> purtest(WBLI.W,exo="trend",test="madwu",pmax=2)
```

```
Maddala-Wu Unit-Root Test (ex. var.: Individual Intercepts and Trend)
```

```
data: WBLI.W
```

```
chisq = 364, df = 376, p-value = 0.7
```

```
alternative hypothesis: stationarity
```

```
> purtest(WBLI.W,exo="trend",test="ips",pmax=2)
```

```
Im-Pesaran-Shin Unit-Root Test (ex. var.: Individual Intercepts and Trend)
```

```
data: WBLI.W
```

```
Wtbar = 2.8, p-value = 1
```

```
alternative hypothesis: stationarity
```

A Better Table

	Test	Alternative	Statistic	Estimate	P-Value
1	Levin-Lin-Chu Unit-Root Test	stationarity	z	-2.672	0.0038
2	Hadri Test	at least one series has a unit root	z	196.75	<0.0001
3	Maddala-Wu Unit-Root Test	stationarity	χ^2	363.89	0.6635
4	Im-Pesaran-Shin Unit-Root Test	stationarity	W_t	2.77	0.9972

Note: All assume individual intercepts and trends.

“Lagged dependent variable”:

$$Y_{it} = \phi Y_{it-1} + \mathbf{X}_{it}\beta_{LDV} + \epsilon_{it}$$

If ϵ_{it} is perfect, then:

- $\hat{\beta}_{LDV}$ is biased (but consistent),
- $O(\text{bias}) = \frac{-1+3\beta_{LDV}}{T}$

If ϵ_{it} is autocorrelated...

- $\hat{\beta}_{LDV}$ is biased and inconsistent
- IV is one (bad) option...

Lagged Y s and GLS-ARMA

Can rewrite:

$$\begin{aligned}Y_{it} &= \mathbf{X}_{it}\boldsymbol{\beta}_{AR} + u_{it} \\u_{it} &= \phi u_{it-1} + \eta_{it}\end{aligned}$$

as

$$\begin{aligned}Y_{it} &= \mathbf{X}_{it}\boldsymbol{\beta}_{AR} + \phi u_{it-1} + \eta_{it} \\&= \mathbf{X}_{it}\boldsymbol{\beta}_{AR} + \phi(Y_{it-1} - \mathbf{X}_{it-1}\boldsymbol{\beta}_{AR}) + \eta_{it} \\&= \phi Y_{it-1} + \mathbf{X}_{it}\boldsymbol{\beta}_{AR} + \mathbf{X}_{it-1}\psi + \eta_{it}\end{aligned}$$

where $\psi = \phi\boldsymbol{\beta}_{AR}$ and $\psi = 0$ (by assumption).

Lagged Y s and World Domination

In:

$$Y_{it} = \phi Y_{it-1} + \mathbf{X}_{it}\beta_{LDV} + \epsilon_{it}$$

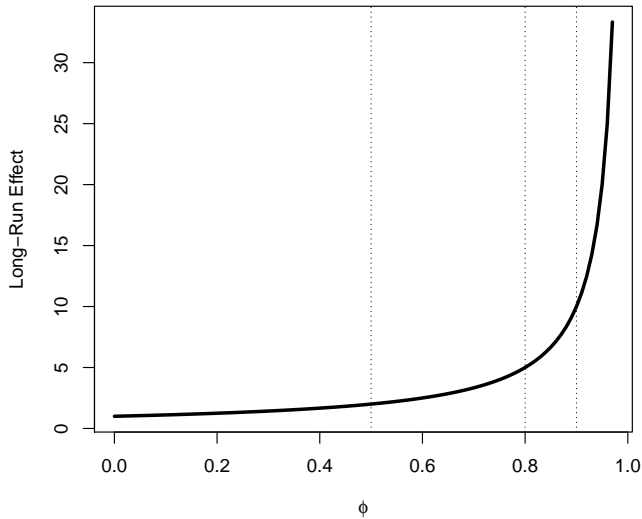
Achen: Bias “deflates” $\hat{\beta}_{LDV}$ relative to $\hat{\phi}$, “suppress” the effects of \mathbf{X} ...

Keele & Kelly (2006):

- Contingent on ϵ s having autocorrelation
- Key: In LDV, *long-run impact of a unit change in X is:*

$$\hat{\beta}_{LR} = \frac{\hat{\beta}_{LDV}}{1 - \hat{\phi}}$$

Long-Run Impact for $\hat{\beta} = 1$



Lagged Y s and Unit Effects

Consider:

$$Y_{it} = \phi Y_{it-1} + \mathbf{X}_{it}\boldsymbol{\beta} + \alpha_i + u_{it}.$$

If we omit the unit effects, we have:

$$Y_{it} = \phi Y_{it-1} + \mathbf{X}_{it}\boldsymbol{\beta} + u_{it}^*$$

with

$$u_{it}^* = \alpha_i + u_{it}$$

Lagging yields:

$$Y_{it-1} = \phi Y_{it-2} + \mathbf{X}_{it-1}\boldsymbol{\beta} + \alpha_i + u_{it-1}$$

which means

$$\text{Cov}(Y_{it-1}, u_{it}^*) \neq 0. \rightarrow \text{bias in } \hat{\phi}, \hat{\boldsymbol{\beta}}$$

Bias in $\hat{\phi}$ is

- toward zero when $\phi > 0$,
- increasing in ϕ .

Including unit effects still yields bias in $\hat{\phi}$ of $O(\frac{1}{T})$, and bias in $\hat{\beta}$.

Solutions:

- Difference/GMM estimation
- Bias correction approaches

First Difference Estimation

$$\begin{aligned}Y_{it} - Y_{it-1} &= \phi(Y_{it-1} - Y_{it-2}) + (\mathbf{X}_{it} - \mathbf{X}_{it-1})\beta + (\alpha_i - \alpha_i) + (u_{it} - u_{it-1}) \\ \Delta Y_{it} &= \phi \Delta Y_{it-1} + \Delta \mathbf{X}_{it}\beta + \Delta u_{it}\end{aligned}$$

Anderson/Hsiao: If \nexists autocorrelation, then use ΔY_{it-2} or Y_{it-2} as instruments for $\Delta Y_{it-1} \dots$

- Consistent in theory,
- in practice, the former is preferred, and both have issues if ϕ is high;
- both are inefficient.

Arellano & Bond (also Wawro): Use *all* lags of Y_{it} and \mathbf{X}_{it} from $t - 2$ and before.

- “Good” estimates, better as $T \rightarrow \infty$,
- Easy to handle higher-order lags of Y ,
- Easy software (`p1m` in R , `xtabond` in Stata).
- Model *is* fixed effects...
- \mathbf{Z}_i has $T - p - 1$ rows, $\sum_{i=p}^{T-2} i$ columns \rightarrow difficulty of estimation declines in p , grows in T .

Kiviet (1995, 1999; Bun and Kiviet 2003; Bruno 2005a,b): Derive the bias in $\hat{\phi}$ and $\hat{\beta}$, then correct it...

- More accurate than the instrumental-variables/GMM estimators of A&H/A&B...
- ...especially when T is small; but not as T gets reasonably large ($T \approx 20$)

Some Dynamic Models

	Lagged Y	First Difference	FE	Lagged Y + FE
Intercept	2.139* (0.319)	0.629* (0.039)		
Lagged WBLI	0.986* (0.002)			0.953* (0.003)
Population Growth	-0.061* (0.023)	0.003 (0.036)	-0.089 (0.095)	-0.064* (0.028)
Urban Population	0.003 (0.002)	-0.019 (0.064)	0.306* (0.020)	0.013* (0.006)
Fertility Rate	-0.073* (0.028)	-0.833* (0.335)	-2.033* (0.162)	-0.244* (0.049)
ln(GDP Per Capita)	-0.034 (0.035)	0.502 (0.445)	8.723* (0.300)	0.148 (0.095)
Natural Resource Rents	-0.009* (0.003)	0.023* (0.007)	0.065* (0.017)	-0.003 (0.005)
Cold War	-0.295* (0.069)	-0.062 (0.201)	-6.869* (0.296)	-0.408* (0.090)
R ²	0.985	0.002	0.532	0.958
Adj. R ²	0.985	0.002	0.521	0.957
Num. obs.	7996	7913	8100	7996

* $p < 0.05$

Anderson-Hsiao, Arellano-Bond, etc.

In R:

- Anderson-Hsiao can be fit using `lm` or (more easily) `plm` in the `plm` package
- Arellano-Bond is most easily fit using `pgmm` (“panel gmm”) in the `plm` package
- See Cribari and Millo (2018, Chapter 7) for statistics + code details
- [This post](#) is also useful...

Stata:

- `xtabond` / `xtdpdsys` / `xtdpd` fit both A-H and A-B / Blundell-Bond models (among others)
- [This](#) is also a good (slightly dated) reference

What if Y is *trending* over time?

- First Question: Why?
 - Organic growth (e.g., populations)
 - Temporary / short-term factors
 - Covariates...
- Second question: Should we care?
(A: Yes, usually... → “spurious regressions”)
- Third question: What to do?
 - Ignore it...
 - Include a counter / trend term...

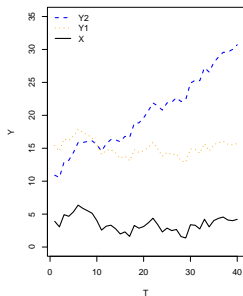
In general, adding a trend term will *decrease* the magnitudes of $\hat{\beta}$...

Trends Matter, Illustrated

Data generating processes:

$$Y_{1t} = 10 + (1 \times X_t) + u_t$$

$$Y_{2t} = 5 + (1 \times X_t) + (0.5 \times T) + u_t$$



	Y ₁	Y ₂	
		No Trend	Trend
X	0.921*** (0.245)	-0.382 (0.786)	0.874*** (0.255)
T			0.482*** (0.026)
Constant	10.300*** (0.917)	20.200*** (2.950)	5.860*** (1.200)
Observations	40	40	40
R ²	0.272	0.006	0.905
Adjusted R ²	0.253	-0.020	0.900
Residual Std. Error	1.800 (df = 38)	5.790 (df = 38)	1.810 (df = 37)

Note:

* p<0.1; ** p<0.05; *** p<0.01

Trends Matter, Part II

Table: FE Models of WBLI

	FE	FE.trend	FE.intx
Population Growth	-0.089 (0.095)	-0.266*** (0.080)	-0.227*** (0.079)
Urban Population	0.306*** (0.020)	0.019 (0.017)	0.037** (0.017)
Fertility Rate	-2.033*** (0.162)	1.276*** (0.146)	1.206*** (0.146)
ln(GDP Per Capita)	8.723*** (0.300)	2.069*** (0.274)	1.855*** (0.274)
Natural Resource Rents	0.065*** (0.017)	0.034** (0.014)	0.037*** (0.014)
Cold War	-6.869*** (0.296)	1.760*** (0.286)	8.688*** (0.927)
Trend (1950=0)		0.745*** (0.013)	0.770*** (0.013)
Cold War x Trend			-0.202*** (0.026)
Observations	8,100	8,100	8,100
R ²	0.532	0.676	0.678
Adjusted R ²	0.521	0.668	0.670
F Statistic	1,501.000*** (df = 6; 7907)	2,354.000*** (df = 7; 7906)	2,083.000*** (df = 8; 7905)

*p<0.1; **p<0.05; ***p<0.01

Another Approach: Orthogonalization

Note: We're rarely substantively interested in the fixed effects $\hat{\alpha}$...

- \rightarrow reparameterize the α s so that they are *information-orthogonal* to the other parameters in the model (including the β s and ϕ)
- Key idea: Transform the α s so that (for example):

$$E \left(\frac{\partial^2 L_i}{\partial \alpha \partial \beta} \right) = 0$$

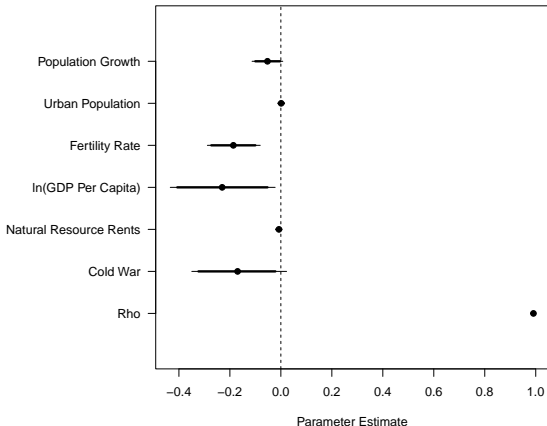
- Can do this via imposition of priors, in a Bayesian framework...
- **In general**, this approach is less assumption-laden and more efficient than the IV/GMM-based approaches discussed above.
- Provides consistent-in- N estimates for T as low as 2...

References:

- Lancaster, T. 2002. "Orthogonal Parameters and Panel Data." *Review of Economic Studies* 69:647-666.
- Pickup et al. (2017) [the "orthogonalized panel model" ("OPM")]

FE + Dynamics Using Orthogonalization

```
> library(OrthoPanels)
> set.seed(7222009)
> OPM.fit <- opm(WomenBusLawIndex~PopGrowth+UrbanPopulation+FertilityRate+
  lnGDPPerCap+NaturalResourceRents+ColdWar,data=WDI,
  index=c("ISO3","Year"),n.samp=1000)
```



OPM Results: Short- and Long-Run Effects

For $\hat{\phi} \approx 0.98$:

Parameter	Short-Run	Long-Run
Population Growth	-0.0526	-5.5623
Urban Population	0.0012	0.1468
Fertility Rate	-0.1864	-20.1788
ln(GDP Per Capita)	-0.2303	-25.5141
Natural Resource Rents	-0.0075	-0.8048
Cold War	-0.1697	-18.3407

R :

- the `plm` package (`purtest` for unit roots; `plm` for first-difference models; `pgmm` for Arellano-Bond)
- the `panelAR` package (GLS-ARMA models)
- the `glS` package (GLS)
- the `dynpanel` package (A&H, A&B; minimal...)

Stata :

- `xtglS` (GLS)
- `xtpcse` (PCSEs)
- `xtabond` / `xtdpd` (A&H A&B dynamic models)

Final Thoughts: Dynamic Panel Models

Things to consider:

- N vs. T ...
- Are dynamics nuisance or substance?
- What problem(s) do you *really* care about?