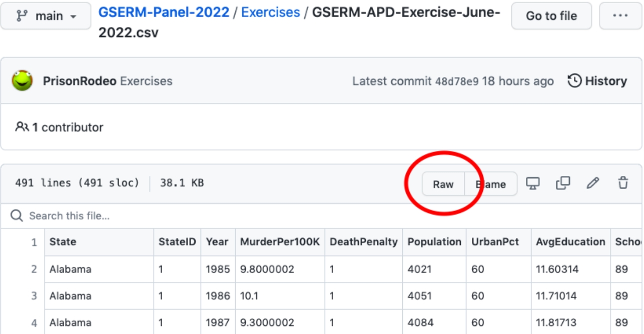
GSERM - St. Gallen (2022)
Analyzing Panel Data

June 8, 2022

Data on Github



Can also use (e.g.) `read_csv` (in `readr`):

```
> library(readr)
> Data<-read_csv("https://github.com/PrisonRodeo/GSERM-Panel-2022/raw/main/Exercises/GSERM-APD-Exercise-June-2022.csv")
```

Generalized Least Squares Models

For:

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

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

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

That is, 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)

The Key: $\mathbf{\Omega}$

Estimator:

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

with:

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

Two approaches:

- Use OLS $\hat{u}_{it}$s to get $\hat{\mathbf{\Omega}}$ ("feasible GLS")
- Use substantive knowledge about the data to structure $\mathbf{\Omega}$

Getting to Know WLS

The variance-covariance matrix is:

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

A common case is:

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

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

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

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

Huber's Insight

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

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

Yields:

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

"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:

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

with

$$\sigma^2_{ij} = \sigma^2_{ik}.$$

"Robust, clustered" estimator:

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

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:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_t + u_t \\ u_t &= \rho u_{t-1} + e_t \end{aligned}$$

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

$\rightarrow$ "First-order autoregressive" ("AR(1)") errors.

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:

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

Descriptive Statistics:

```
> describe(subset,fast=TRUE)
                     vars    n  mean    sd   min    max range   se
WomenBusLawIndex        1 7566 59.99 18.79 17.50 100.00 82.50 0.22
PopGrowth               2 7566  1.70  1.46 -6.77  17.51 24.28 0.02
UrbanPopulation         3 7566 51.19 23.90  2.85 100.00 97.16 0.27
FertilityRate           4 7566  3.67  1.91  0.90   8.61  7.70 0.02
NaturalResourceRents    5 7566  6.82 10.45  0.00  87.51 87.51 0.12
ColdWar                 6 7566  0.32  0.47  0.00   1.00  1.00 0.01
lnGDPPerCap             7 7566  8.28  1.44  5.32  11.63  6.31 0.02
```

How Much Autocorrelation in $\mathbf{X}$?

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	0.942
2	Urban Population	0.973
3	Fertility Rate	0.968
4	GDP Per Capita	0.976
5	Natural Resource Rents	0.917
6	Cold War	0.913

Baseline Model: OLS (+ D-W Test)

```
> OLS<-plm(WomenBusLawIndex~PopGrowth+UrbanPopulation+FertilityRate+
+          log(GDPPerCapita)+NaturalResourceRents+ColdWar,
+          data=WDI,model="pooling")

> summary(OLS)
Pooling Model

Call:
plm(formula = WomenBusLawIndex ~ PopGrowth + UrbanPopulation +
    FertilityRate + log(GDPPerCapita) + NaturalResourceRents +
    ColdWar, data = WDI, model = "pooling")

Unbalanced Panel: n = 186, T = 1-50, N = 7566

Coefficients:
                     Estimate Std. Error t-value    Pr(>|t|)
(Intercept)           54.8395     1.7261   31.77     < 2e-16 ***
PopGrowth             -3.1926     0.1437  -22.21     < 2e-16 ***
UrbanPopulation       -0.0584     0.0109   -5.37 0.000000083 ***
FertilityRate         -1.7928     0.1652  -10.85     < 2e-16 ***
log(GDPPerCapita)      3.1544     0.1993   15.83     < 2e-16 ***
NaturalResourceRents  -0.3486     0.0162  -21.54     < 2e-16 ***
ColdWar              -11.3437     0.3716  -30.53     < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Total Sum of Squares:    2670000
Residual Sum of Squares: 1280000
R-Squared:      0.519
Adj. R-Squared: 0.519
F-statistic: 1361.01 on 6 and 7559 DF, p-value: <2e-16

> # Durbin-Watson test:
> 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.:       7566 Avg obs. per panel 40.677
 Number of panels: 186  Max obs. per panel 50
 Number of times:  50   Min obs. per panel 1

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          71.81494    3.16208   22.71  < 2e-16 ***
PopGrowth            -0.03691    0.08543   -0.43   0.6657
UrbanPopulation      -0.03721    0.02468   -1.51   0.1317
FertilityRate        -5.64038    0.25230  -22.36  < 2e-16 ***
log(GDPPerCapita)     1.41254    0.36363    3.88   0.0001 ***
NaturalResourceRents -0.01931    0.00868   -2.23   0.0261 *
ColdWar              -0.90520    0.22070   -4.10 0.000041 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

R-squared: 0.3621
Wald statistic: 1075.5886, Pr(>Chisq(6)): 0

> PraisWinsten$panelStructure$rho
[1] 0.9523
```

Better in a Table

	OLS	Prais-Winsten
Intercept	54.84^*	71.81^*
	(1.73)	(3.16)
Population Growth	-3.19^*	-0.04
	(0.14)	(0.08)
Urban Population	-0.06^*	-0.04
	(0.01)	(0.02)
Fertility Rate	-1.79^*	-5.64^*
	(0.17)	(0.25)
ln(GDP Per Capita)	3.15^*	1.41^*
	(0.20)	(0.36)
Natural Resource Rents	-0.35^*	-0.02^*
	(0.02)	(0.008)
Cold War	-11.34^*	-0.91^*
	(0.37)	(0.22)
$\hat{\rho}$		0.95
R^2	0.52	0.36
Adj. R^2	0.52	
NT	7566	7566
N panels		186

$^* p < 0.05$

Some Panel Data Challenges

Consider the error terms in the model:

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

Issues:

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

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 $\rightarrow \hat{\rho}$ ($\rightarrow \hat{\mathbf{\Omega}}$),
2. Use $\hat{\rho}$ for Prais-Winsten.

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

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

where

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

Means:

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

More Parks Problems

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 $R = (r_1, \ldots, r_N)'$ 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:

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

where

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

and

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

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:

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

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

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

Panel-Corrected Standard Errors

Correct formula:

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

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
$\sigma_i^2 = \sigma^2$, $Cov(\sigma_{it}, \sigma_{jt}) = 0$	$k+1$	$k+2$	$k+N+1$
$\sigma_i^2 \neq \sigma^2$, $Cov(\sigma_{it}, \sigma_{jt}) = 0$	$k+N$	$k+N+1$	$k+2N$
$\sigma_i^2 \neq \sigma^2$, $Cov(\sigma_{it}, \sigma_{jt}) \neq 0$	$\frac{N(N-1)}{2} + k + N$	$\frac{N(N-1)}{2} + k + N + 1$	$\frac{N(N-1)}{2} + k + 2N$

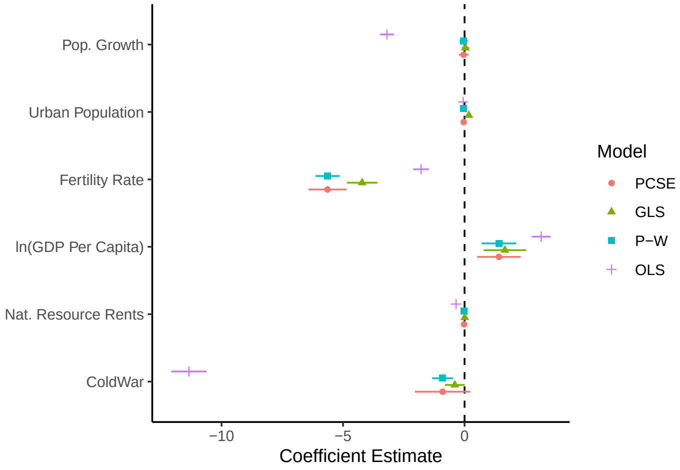
Example: GLS with Homoscedastic AR(1) Errors

```
> 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 + ColdW
  Data: WDI
    AIC   BIC logLik
  35750 35813 -17866

Correlation Structure: AR(1)
 Formula: ~1 | ISO3
 Parameter estimate(s):
   Phi
0.9888

Coefficients:
                     Value Std.Error t-value p-value
(Intercept)          53.93     4.195  12.855  0.0000
PopGrowth             0.04     0.079   0.469  0.6393
UrbanPopulation       0.18     0.041   4.318  0.0000
FertilityRate        -4.22     0.319 -13.222  0.0000
log(GDPPerCapita)     1.66     0.446   3.720  0.0002
NaturalResourceRents  0.01     0.008   1.080  0.2804
ColdWar              -0.40     0.207  -1.929  0.0538
```

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.:       7566 Avg obs. per panel 40.677
 Number of panels: 186  Max obs. per panel 50
 Number of times:  50   Min obs. per panel 1

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           71.8149     4.6453   15.46   <2e-16 ***
PopGrowth             -0.0369     0.0978   -0.38    0.706
UrbanPopulation       -0.0372     0.0250   -1.49    0.136
FertilityRate         -5.6404     0.3983  -14.16   <2e-16 ***
log(GDPPerCapita)      1.4125     0.4559    3.10    0.002 **
NaturalResourceRents -0.0193     0.0129   -1.49    0.135
ColdWar               -0.9052     0.5804   -1.56    0.119
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

R-squared: 0.3621
Wald statistic: 394.4807, Pr(>Chisq(6)): 0

> PCSE$panelStructure$rho
[1] 0.9523
```

Model Comparisons



Dynamics!

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

Mean stationarity:

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

Variance stationarity:

$$Var(Y_t) = E[(Y_t - \mu)^2] \equiv \sigma_Y^2 \; \forall \; t$$

Covariance stationarity:

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

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

The "ARIMA" Approach

"ARIMA" $=$ *Autoregressive Integrated Moving Average*...

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

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

...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 Properties

I(1) series are not stationary.

Variance:

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

Autocovariance:

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

Both depend on t...

I(1) series (continued)

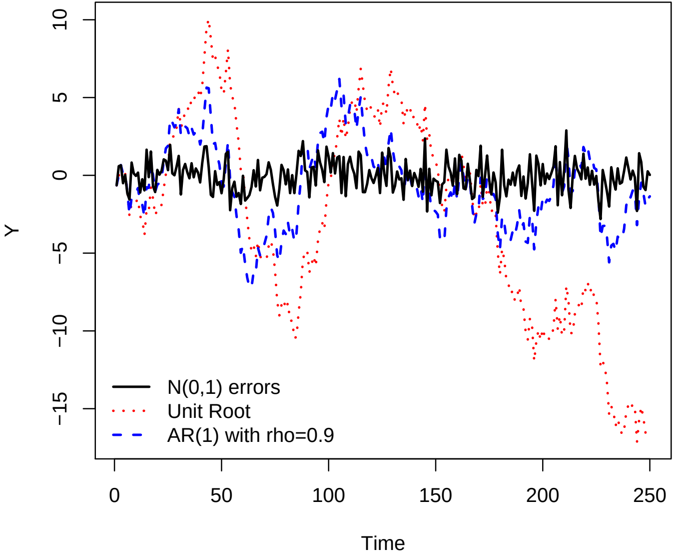
More generally:

- $|\rho| > 1$
  - 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.

Unit Root Alternatives

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 = 0.003
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 = 189, 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 = 332, df = 376, p-value = 0.9
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 = 3.6, p-value = 1
alternative hypothesis: stationarity
```

A Better Table

Table: Panel Unit Root Tests: WBRI

Test	Alternative	Statistic	Estimate	P-Value
Levin-Lin-Chu Test	stationarity	z	-2.286	0.0111
Hadri Test (Heterosked. Consistent)	$\geq$ one series has a unit root	z	192.036	< 0.001
Maddala-Wu Test	stationarity	χ^2	782.604	< 0.001
Im-Pesaran-Shin Test	stationarity	$\bar{W}_t$	3.342	0.9996

Note: All assume individual intercepts and trends.

Lagged: Y?

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

If ϵ_{it} is perfect...

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

If ϵ_{it} is autocorrelated...

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

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 $\psi = \phi\beta_{AR}$ and $\psi = 0$ (by assumption).

Lagged Ys and World Domination

In:

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

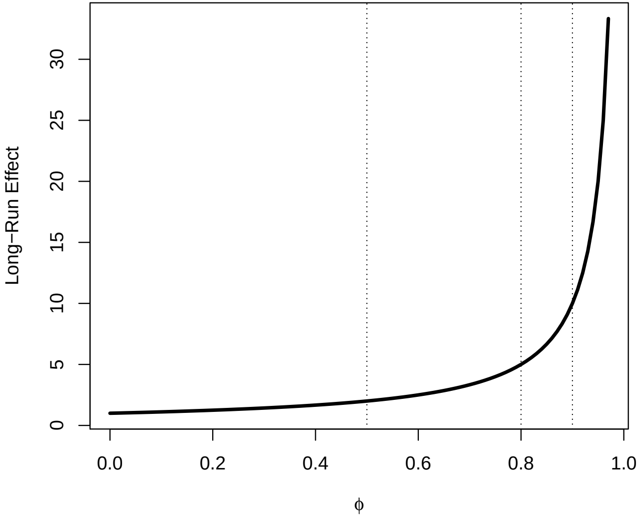
Achen: Bias "deflates" $\hat{\boldsymbol{\beta}}_{LDV}$ relative to $\hat{\phi}$, "suppress" the effects of $\mathbf{X}$...

Keele & Kelly (2006):

- Contingent on ϵs 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 Ys 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

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

“Nickell” Bias

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 $\hat{\boldsymbol{\beta}}$.

Solutions:

- Difference/GMM estimation
- Bias correction approaches

First Difference Estimation

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

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

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

A&H $\rightarrow$ A&B

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 (`plm` 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, <u>grows in T</u>.

Bias-Correction Models

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	2.260^{*}	0.641^{*}		
	(0.335)	(0.040)		
Lagged WBLI	0.986^{*}			0.948^{*}
	(0.002)			(0.004)
Population Growth	-0.051	0.035	0.073	0.011
	(0.027)	(0.077)	(0.119)	(0.037)
Urban Population	0.002	-0.040	0.248^{*}	0.009
	(0.002)	(0.062)	(0.021)	(0.007)
Fertility Rate	-0.085^{*}	-1.023^{*}	-2.066^{*}	-0.292^{*}
	(0.030)	(0.373)	(0.166)	(0.052)
ln(GDP Per Capita)	-0.036	0.780	9.161^{*}	0.276^{*}
	(0.037)	(0.476)	(0.310)	(0.102)
Natural Resource Rents	-0.010^{*}	0.020^{*}	0.035	-0.003
	(0.003)	(0.008)	(0.018)	(0.006)
Cold War	-0.298^{*}	-0.021	-7.192^{*}	-0.445^{*}
	(0.072)	(0.204)	(0.295)	(0.094)
R^2	0.984	0.003	0.535	0.956
Adj. R^2	0.984	0.002	0.523	0.954
Num. obs.	7463	7380	7566	7463

$^{*}p < 0.05$

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... $\rightarrow$ "spurious regressions")
- Third question: What to do?
  - Ignore it...
  - Include a counter / trend term...

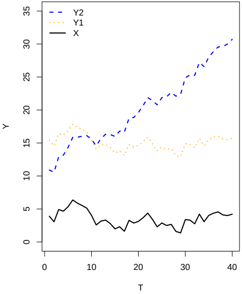
In general, adding a trend term will *decrease* the magnitudes of $\hat{\boldsymbol{\beta}}$...

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***	−0.382	0.874***
	(0.245)	(0.786)	(0.255)
T			0.482***
			(0.026)
Constant	10.300***	20.200***	5.860***
	(0.917)	(2.950)	(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	0.073	-0.287^{***}	-0.242^{**}
	(0.119)	(0.100)	(0.100)
Urban Population	0.248^{***}	-0.024	-0.003
	(0.021)	(0.018)	(0.018)
Fertility Rate	-2.066^{***}	1.080^{***}	1.018^{***}
	(0.166)	(0.150)	(0.149)
ln(GDP Per Capita)	9.161^{***}	2.867^{***}	2.585^{***}
	(0.310)	(0.283)	(0.283)
Natural Resource Rents	0.035^{*}	0.009	0.008
	(0.018)	(0.015)	(0.015)
Cold War	-7.192^{***}	1.660^{***}	9.300^{***}
	(0.295)	(0.293)	(0.944)
Trend (1950=0)		0.749^{***}	0.783^{***}
		(0.013)	(0.014)
Cold War x Trend			-0.220^{***}
			(0.026)
Observations	7,566	7,566	7,566
R^2	0.535	0.674	0.678
Adjusted R^2	0.523	0.666	0.669
F Statistic	$1,414.000^{***}$ (df = 6; 7374)	$2,182.000^{***}$ (df = 7; 7373)	$1,937.000^{***}$ (df = 8; 7372)

*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 αs so that they are *information-orthogonal* to the other parameters in the model (including the βs and ϕ)
- Key idea: Transform the αs so that (for example):

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

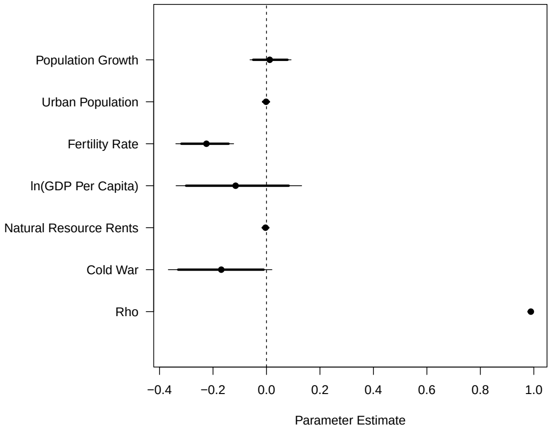
- Can do this via imposition of priors, in a Bayesian framework...
- **In general**, this approach is less assumption-laden and more efficient than the IV/GMM-based approaches discussed above.
- Provides consistent-in-N estimates for T as low as 2...

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	0.0122	0.9148
Urban Population	-0.0016	-0.1420
Fertility Rate	-0.2247	-19.0090
ln(GDP Per Capita)	-0.1155	-9.9996
Natural Resource Rents	-0.0037	-0.3086
Cold War	-0.1691	-14.3630

R :

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

Stata :

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

Final Thoughts: Dynamic Panel Models

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