GSERM 2021
Analyzing Panel Data

June 16, 2021

For:

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

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

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

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)

Estimator:

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

with:

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

Two approaches:

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

The variance-covariance matrix is:

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

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 $\mathbf{W} = \sigma^2\boldsymbol{\Omega}$.

We can rewrite $\mathbf{Q}$ as

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

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 $\mathrm{Var}(u) = \sigma^2 \mathbf{I}$.

“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

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

"Robust, clustered" estimator:

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

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(7222009)
> 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)

Residuals:
     Min       1Q   Median       3Q      Max
-1.12328 -0.65321 -0.05073  0.43937  1.81661

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.8438     0.3020   2.794   0.0234 *
X             0.3834     0.3938   0.974   0.3588
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.9313 on 8 degrees of freedom
Multiple R-squared:  0.1059,Adjusted R-squared:  -0.005832
F-statistic: 0.9478 on 1 and 8 DF,  p-value: 0.3588

> rob10 <- vcovHC(fit10,type="HC1")
> sqrt(diag(rob10))
(Intercept)           X
  0.2932735   0.2859552
```

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)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.84383    0.02704   31.20   <2e-16 ***
X            0.38341    0.03526   10.87   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.8338 on 998 degrees of freedom
Multiple R-squared:  0.1059,Adjusted R-squared:  0.105
F-statistic: 118.2 on 1 and 998 DF,  p-value: < 2.2e-16

> summary(fit1K, cluster="ID")

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.8438     0.2766   3.050  0.00235 **
X             0.3834     0.2697   1.421  0.15551
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.8338 on 998 degrees of freedom
Multiple R-squared:  0.1059,Adjusted R-squared:  0.105
F-statistic:  2.02 on 1 and 9 DF,  p-value: 0.1889
```

Parks' Approach

Assume:

- $E(u_{it}^2) = E(u_{is}^2) \; \forall \; t \neq s$
- $E(u_{it}, u_{jt}) = \sigma_{ij} \; \forall \; i \neq j,$
- $E(u_{it}, u_{js}) = 0 \; \forall \; i \neq j, t \neq s$
- $E(u_{it}, u_{is}) = \rho$ or ρ_i

(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)

Parks' Problems

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

where

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

Means:

- $\frac{N(N-1)}{2}$ distinct contemporaneous correlations,
- 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

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

and

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

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:

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

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

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

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 to be Estimated

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 Data: Demonstrations, 1945-2014

Data:

- Data are (a subset of) Banks (2019)
- $N = 180$ countries, $T = 70$ years [1945-2014]
- Variables:
  - `Demonstrations`: Number of social/political demonstrations in that country in that year
  - `POLITY`: The country's POLITY IV score that year (-10 = fully autocratic; 10 = fully democratic)
  - `POLITY`^{2}: POLITY IV squared (expected curvilinear relationship)
  - `GDP`: The per capita GDP (PPP, in constant \$US) for that country / year
  - `Monarch`: Whether (=1) or not (=0) that country was a monarchy in that year
  - `ColdWar`: Indicator variable, coded 1 for the period 1945-1989, 0 otherwise

Regression model:

$$\begin{aligned} ln(\text{Demonstrations} + 1)_{it} &= \beta_0 + \beta_1 \text{POLITY}_{it} + \beta_2 \text{POLITY}^2_{it} + \beta_3 \ln(\text{GDP})_{it} + \\ &\quad \beta_4 \text{Monarch}_{it} + \beta_5 \text{Cold War}_{it} + u_{it} \end{aligned}$$

Data Summary

```
> summary(Demos)
    ccode          Year          POLITY           GDP        
 Min.   :  2   Min.   :1945   Min.   :-10.00   Min.   :   185  
 1st Qu.:235   1st Qu.:1969   1st Qu.: -7.00   1st Qu.:  1580  
 Median :451   Median :1985   Median :  0.00   Median :  4002  
 Mean   :456   Mean   :1984   Mean   :  0.63   Mean   :  8120  
 3rd Qu.:663   3rd Qu.:2000   3rd Qu.:  8.00   3rd Qu.: 10365  
 Max.   :950   Max.   :2014   Max.   : 10.00   Max.   :134040  
                              NA's   :111      NA's   :2348    
   Monarch        lnDemons       ColdWar          lnGDP     
 Min.   :0.0    Min.   :0.0    Min.   :0.000   Min.   : 5.2  
 1st Qu.:0.0    1st Qu.:0.0    1st Qu.:0.000   1st Qu.: 7.4  
 Median :0.0    Median :0.0    Median :1.000   Median : 8.3  
 Mean   :0.1    Mean   :0.3    Mean   :0.563   Mean   : 8.3  
 3rd Qu.:0.0    3rd Qu.:0.0    3rd Qu.:1.000   3rd Qu.: 9.2  
 Max.   :1.0    Max.   :4.3    Max.   :1.000   Max.   :11.8  
 NA's   :1198   NA's   :1149                   NA's   :2348  
```

Example: OLS

```
> OLS<-plm(lnDemons~POLITY+I(POLITY^2)+lnGDP+Monarch+ColdWar,
+          data=PDF,model="pooling")
>
> summary(OLS)
Pooling Model

Call:
plm(formula = lnDemons ~ POLITY + I(POLITY^2) + lnGDP + Monarch +
    ColdWar, data = PDF, model = "pooling")

Unbalanced Panel: n = 145, T = 1-62, N = 6505

Residuals:
   Min. 1st Qu.  Median 3rd Qu.    Max.
-0.4501 -0.2930 -0.2176 -0.0754  4.1073

Coefficients:
             Estimate Std. Error t-value Pr(>|t|)
(Intercept) -0.124639   0.058208   -2.14    0.032 *
POLITY       0.006296   0.001179    5.34  9.5e-08 ***
I(POLITY^2) -0.002267   0.000255   -8.90  < 2e-16 ***
lnGDP        0.057679   0.007513    7.68  1.9e-14 ***
Monarch     -0.046393   0.028572   -1.62    0.104
ColdWar      0.027883   0.013961    2.00    0.046 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Total Sum of Squares:    1850
Residual Sum of Squares: 1800
R-Squared:      0.0261
Adj. R-Squared: 0.0253
F-statistic: 34.8228 on 5 and 6499 DF, p-value: <2e-16
```

Example: Prais-Winsten

```
> PraisWinsten<-panelAR(lnDemons~POLITY+I(POLITY^2)+lnGDP+Monarch+ColdWar,
+                      data=Demos,panelVar="ccode",timeVar="Year",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.:       6505 Avg obs. per panel 44.862
 Number of panels: 145  Max obs. per panel 62
 Number of times:  62   Min obs. per panel 1

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.150924   0.083548   -1.81   0.0709 .
POLITY       0.005223   0.001632    3.20   0.0014 **
I(POLITY^2) -0.002332   0.000346   -6.74  1.7e-11 ***
lnGDP        0.063335   0.010705    5.92  3.5e-09 ***
Monarch     -0.042863   0.040289   -1.06   0.2874
ColdWar      0.004206   0.019559    0.22   0.8297
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

R-squared: 0.0131
Wald statistic: 91.5428, Pr(>Chisq(5)): 0

> PraisWinsten$panelStructure$rho
[1] 0.3897
```

Example: GLS with Homoscedastic AR(1) Errors

```
> GLS<- gls(lnDemons~POLITY+I(POLITY^2)+lnGDP+Monarch+ColdWar,
+                Demos,correlation=corAR1(form=~1|ccode),
+                na.action="na.omit")
> summary(GLS)
Generalized least squares fit by REML
  Model: lnDemons ~ POLITY + I(POLITY^2) + lnGDP + Monarch + ColdWar

Correlation Structure: AR(1)
 Formula: ~1 | ccode
 Parameter estimate(s):
   Phi
0.4391

Coefficients:
               Value Std.Error t-value p-value
(Intercept) -0.15206   0.09001  -1.689  0.0912
POLITY       0.00495   0.00174   2.851  0.0044
I(POLITY^2) -0.00233   0.00037  -6.361  0.0000
lnGDP        0.06405   0.01151   5.565  0.0000
Monarch     -0.04368   0.04319  -1.011  0.3119
ColdWar     -0.00296   0.02090  -0.142  0.8874

 Correlation:
            (Intr) POLITY I(POLI lnGDP  Monrch
POLITY       0.370
I(POLITY^2)  0.262 -0.199
lnGDP       -0.969 -0.368 -0.428
Monarch      0.161  0.447 -0.189 -0.164
ColdWar     -0.235  0.204 -0.225  0.146  0.077

Standardized residuals:
    Min      Q1     Med      Q3     Max
-0.8982 -0.5537 -0.4148 -0.1423  7.6882

Residual standard error: 0.5299
Degrees of freedom: 6505 total; 6499 residual
```

Example: PCSEs

```
> PCSE<-panelAR(lnDemons~POLITY+I(POLITY^2)+lnGDP+Monarch+ColdWar,
+                      data=Demos,panelVar="ccode",timeVar="Year",
+                      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.:        6505 Avg obs. per panel 44.862
 Number of panels: 145  Max obs. per panel 62
 Number of times:  62   Min obs. per panel 1

Coefficients:
            Estimate Std. Error t value     Pr(>|t|)
(Intercept) -0.15092    0.09430   -1.60        0.110
POLITY       0.00522    0.00212    2.47        0.014 *
I(POLITY^2) -0.00233    0.00040   -5.83 0.0000000058 ***
lnGDP        0.06334    0.01336    4.74 0.0000021559 ***
Monarch     -0.04286    0.03590   -1.19        0.232
ColdWar      0.00421    0.04000    0.11        0.916
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

R-squared: 0.0131
Wald statistic: 63.66, Pr(>Chisq(5)): 0

> PCSE$panelStructure$rho
[1] 0.3897
```

Dynamics!

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.

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
    - Only detects first-order autocorrelation

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

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

Mean stationarity:

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

Variance stationarity:

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

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:

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

Autocovariance:

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

Both depend on t...

I(1) series (continued)

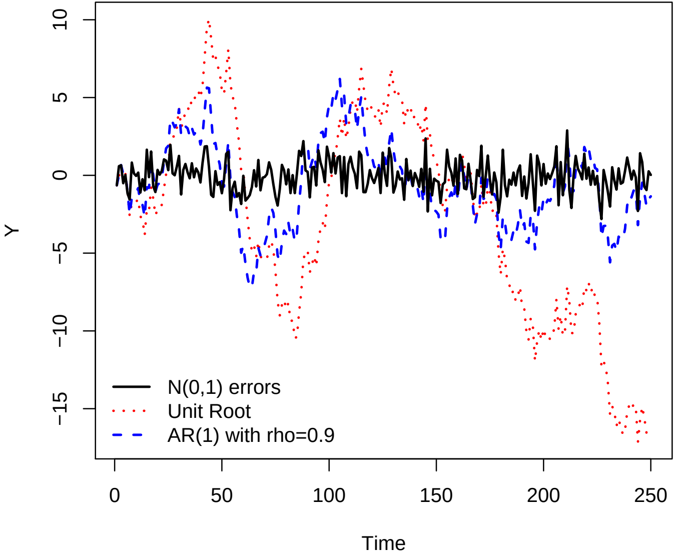
More generally:

- $|\rho| > 1$
  - 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 $\hat{\rho} = 0$, *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

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

- Test $\hat{\rho} = 0$

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)
- What to do?
  - Difference the data...
  - Error-correction models

Panel Unit Root Tests: R

```
> lnDemos<-cbind(Demos$ccode,Demos$Year,Demos$lnDemons)
> lnDemos<-na.omit(lnDemos)


> purtest(lnDemos,exo="trend",test=c("levinlin"))

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

data:  lnDemos
z = -3.2, p-value = 0.0007
alternative hypothesis: stationarity


> purtest(lnDemos,exo="trend",test=c("hadri"))

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

data:  lnDemos
z = 671, p-value <2e-16
alternative hypothesis: at least one series has a unit root


> purtest(lnDemos,exo="trend",test=c("ips"))

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

data:  lnDemos
Wtbar = -24, p-value <2e-16
alternative hypothesis: stationarity
```

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

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 inconsistent
- IV is one (bad) option...

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

Lagged Ys and World Domination

In:

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

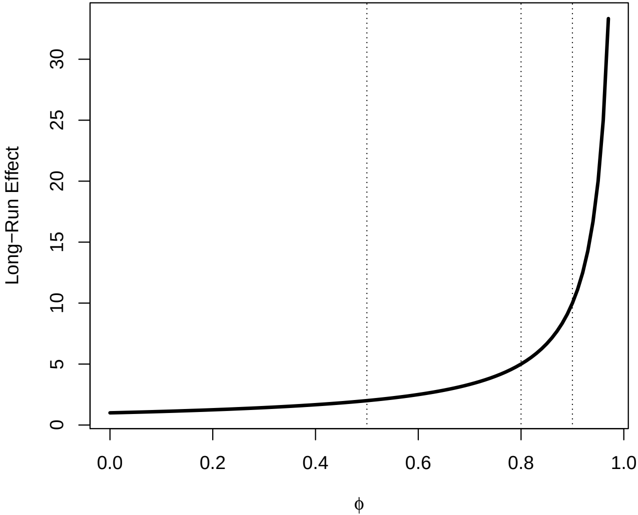
Achen: Bias "deflates" $\hat{\boldsymbol{\beta}}_{LDV}$ relative to $\hat{\phi}$, "suppress" the effects of $\mathbf{X}$...

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

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

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 ($T \approx 20$)

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.

<u>References</u>:

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

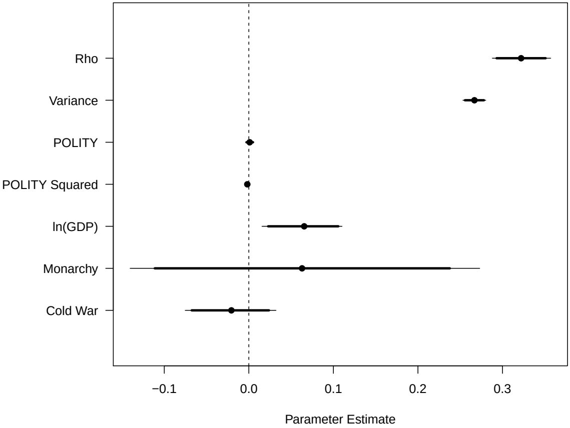
Some Dynamic Models

	LDV	First Difference	FE	LDV + FE
Intercept	-0.104	0.010		
	(0.053)	(0.007)		
Lagged ln(Demonstrations)	0.440^*			0.267^*
	(0.012)			(0.013)
POLITY	0.003^*	0.001	0.002	< 0.001
	(0.001)	(0.004)	(0.002)	(0.002)
POLITY Squared	-0.001^*	-0.003^*	-0.002^*	-0.002^*
	(< 0.001)	(0.001)	(< 0.001)	(< 0.001)
ln(GDP)	0.038^*	-0.108	0.055^*	0.049^*
	(0.007)	(0.079)	(0.015)	(0.015)
Monarch	-0.017	-0.004	0.048	0.070
	(0.026)	(0.139)	(0.068)	(0.067)
Cold War	0.011	-0.134^*	-0.035^*	-0.029
	(0.013)	(0.052)	(0.016)	(0.016)
R^2	0.200	0.004	0.013	0.077
Adj. R^2	0.199	0.003	-0.010	0.055
Num. obs.	6419	6360	6505	6419

$^* p < 0.05$

FE + Dynamics Using Orthogonalization

```
> library(OrthoPanels)
> set.seed(7222009)
> OPM.fit <- opm(lnDemons~POLITY+POLITYSQ+lnGDP+Monarch+ColdWar,
               data=PDF,index=c("ccode","Year"),n.samp=1000)
```



OPM Results: Short- and Long-Run Effects

For $\hat{\phi} \approx 0.32$:

Parameter	Short-Run	Long-Run
POLITY	0.0010	0.0015
POLITY Squared	-0.0018	-0.0027
ln(GDP)	0.0655	0.0956
Monarch	0.0629	0.0913
Cold War	-0.0206	-0.0310

Unit Effects Models: Software

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?