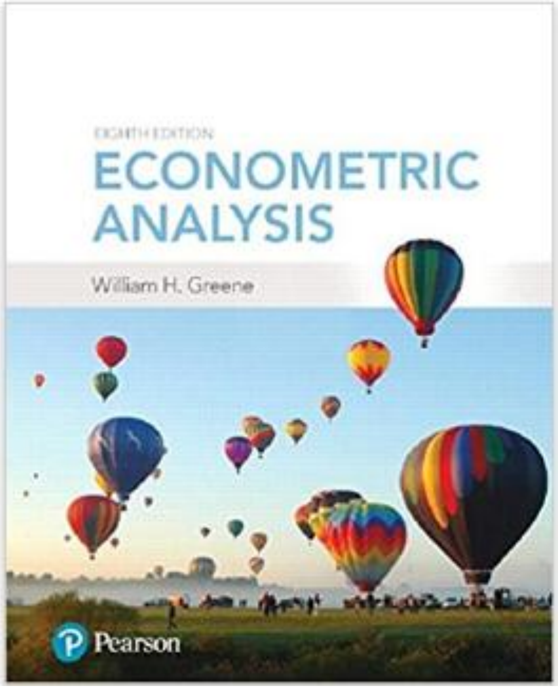
Econometrics I

Professor William Greene

Stern School of Business

Department of Economics



Econometrics I

Part 16 – Panel Data-2

The Random Effects Model

- The random effects model

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + c_i + \varepsilon_{it}, \text{ observation for person i at time t}$$

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + c_i\mathbf{i} + \boldsymbol{\varepsilon}_i, \; T_i \text{ observations in group i}$$

$$= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{c}_i + \boldsymbol{\varepsilon}_i, \text{ note } \mathbf{c}_i = (c_i, c_i, ..., c_i)'$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{c} + \boldsymbol{\varepsilon}, \; \Sigma_{i=1}^{N} T_i \text{ observations in the sample}$$

$$c = (\mathbf{c}_1', \mathbf{c}_2', ...\mathbf{c}_N')', \Sigma_{i=1}^{N} T_i \text{ by 1 vector}$$

- c_i is uncorrelated with $\mathbf{x}_{it}$ for all t;

  $E[c_i \mid \mathbf{X}_i] = 0$

  $E[\varepsilon_{it} \mid \mathbf{X}_i, c_i] = 0$

Notation

$$\begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \\ \vdots \\ \mathbf{y_N} \end{bmatrix} = \begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \\ \vdots \\ \mathbf{X_N} \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{bmatrix} + \begin{bmatrix} u_1\mathbf{i_1} \\ u_2\mathbf{i_2} \\ \vdots \\ u_N\mathbf{i_N} \end{bmatrix} \quad \begin{matrix} T_1 \text{ observations} \\ T_2 \text{ observations} \\ \vdots \\ T_N \text{ observations} \end{matrix}$$

$$= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} + \mathbf{u} \quad \Sigma_{\mathbf{i=1}}^{\mathbf{N}} T_i \text{ observations}$$

$$= \mathbf{X}\boldsymbol{\beta} + \mathbf{w}$$

In all that follows, except where explicitly noted, X, $\mathbf{X}_i$ and $\mathbf{x}_{it}'$ contain a constant term as the first element.

To avoid notational clutter, in those cases, $\mathbf{x}_{it}'$ etc. will simply denote the counterpart without the constant term.

Use of the symbol K for the number of variables will thus be context specific but will usually include the constant term.

Error Components Model

A Generalized Regression Model

$$y_{it} = \mathbf{x}_{it}'\mathbf{b} + \varepsilon_{it} + u_i$$

$$E[\varepsilon_{it} \mid \mathbf{X}_i] = 0$$

$$E[\varepsilon_{it}^2 \mid \mathbf{X}_i] = \sigma_\varepsilon^2$$

$$E[u_i \mid \mathbf{X}_i] = 0$$

$$E[u_i^2 \mid \mathbf{X}_i] = \sigma_u^2$$

$$\text{Var}[\boldsymbol{\varepsilon}_i + u_i\mathbf{i}] = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_u^2 & \sigma_u^2 & \ldots & \sigma_u^2 \\ \sigma_u^2 & \sigma_\varepsilon^2 + \sigma_u^2 & \ldots & \sigma_u^2 \\ \ldots & \ldots & \ddots & \ldots \\ \sigma_u^2 & \sigma_u^2 & \ldots & \sigma_\varepsilon^2 + \sigma_u^2 \end{bmatrix} = \boldsymbol{\Omega}_\mathbf{i}$$

$\mathbf{y}_i = \mathbf{X}_\mathbf{i}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i + u_i\mathbf{i}$ for T_i observations

Notation

$$\text{Var}[\boldsymbol{\varepsilon}_i + u_i\mathbf{i}\,] = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_\varepsilon^2 + \sigma_u^2 & \cdots & \sigma_u^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_\varepsilon^2 + \sigma_u^2 \end{bmatrix}$$

$$= \sigma_\varepsilon^2 \mathbf{I}_{\mathbf{T_i}} + \sigma_u^2 \mathbf{ii}' \quad T_i \times T_i$$

$$= \sigma_\varepsilon^2 \mathbf{I}_{\mathbf{T_i}} + \sigma_u^2 \mathbf{ii}'$$

$$= \boldsymbol{\Omega}_i$$

$$\text{Var}[\mathbf{w} \mid \mathbf{X}] = \begin{bmatrix} \boldsymbol{\Omega}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Omega}_N \end{bmatrix}$$

(Note these differ only in the dimension T_i)

Convergence of Moments

$$\frac{\mathbf{X}'\mathbf{X}}{\Sigma_{i=1}^{N} T} = \Sigma_{i=1}^{N} f_i \frac{\mathbf{X}_i'\mathbf{X}_i}{T_i} = \text{a weighted sum of individual moment matrices}$$

$$\frac{\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}}{\Sigma_{i=1}^{N} T} = \Sigma_{i=1}^{N} f_i \frac{\mathbf{X}_i'\boldsymbol{\Omega}_i\mathbf{X}_i}{T_i} = \text{a weighted sum of individual moment matrices}$$

$$= \sigma_\varepsilon^2 \Sigma_{i=1}^{N} f_i \frac{\mathbf{X}_i'\mathbf{X}_i}{T_i} + \sigma_u^2 \Sigma_{i=1}^{N} f_i \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i'$$

Note asymptotics are with respect to N. Each matrix $\frac{\mathbf{X}_i'\mathbf{X}_i}{T_i}$ is the moments for the T_i observations. Should be 'well behaved' in micro level data. The average of N such matrices should be likewise. T or T_i is assumed to be fixed (and small).

Random vs. Fixed Effects

- Random Effects
  - Small number of parameters
  - Efficient estimation
  - Objectionable orthogonality assumption ($c_i \perp \mathbf{X}_i$)
- Fixed Effects
  - Robust – generally consistent
  - Large number of parameters

Ordinary Least Squares

- Standard results for OLS in a GR model
  - Consistent
  - Unbiased
  - Inefficient
- True variance of the least squares estimator

$$\text{Var}[\mathbf{b} \mid \mathbf{X}] = \frac{\mathbf{1}}{\Sigma_{i=1}^{N} T_i} \left[\frac{\mathbf{X}'\mathbf{X}}{\Sigma_{i=1}^{N} T_i} \right]^{-1} \frac{\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}}{\Sigma_{i=1}^{N} T_i} \left[\frac{\mathbf{X}'\mathbf{X}}{\Sigma_{i=1}^{N} T_i} \right]^{-1}$$

$$\rightarrow \mathbf{0} \times \rightarrow \mathbf{Q}^{-1} \times \rightarrow \mathbf{Q}^{*} \times \rightarrow \mathbf{Q}^{-1}$$

$$\rightarrow \mathbf{0} \text{ as } N \rightarrow \infty$$

Estimating the Variance for OLS

$$\text{Var}[\mathbf{b} \mid \mathbf{X}] = \frac{\mathbf{1}}{\Sigma_{i=1}^{N} T_i} \left[\frac{\mathbf{X}'\mathbf{X}}{\Sigma_{i=1}^{N} T_i} \right]^{-1} \left(\frac{\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}}{\Sigma_{i=1}^{N} T_i} \right) \left[\frac{\mathbf{X}'\mathbf{X}}{\Sigma_{i=1}^{N} T_i} \right]^{-1}$$

In the spirit of the White estimator, use

$$\frac{\mathbf{X}'\hat{\boldsymbol{\Omega}}\mathbf{X}}{\Sigma_{i=1}^{N} T_i} = \Sigma_{i=1}^{N} f_i \frac{\mathbf{X}_i'\hat{\mathbf{w}}_i\hat{\mathbf{w}}_i'\mathbf{X}_i}{T_i}, \ \hat{\mathbf{w}}_i = \mathbf{y}_i - \mathbf{X}_i\mathbf{b}, \ f_i = \frac{T_i}{\Sigma_{i=1}^{N} T_i}$$

Hypothesis tests are then based on Wald statistics.

THIS IS THE 'CLUSTER' ESTIMATOR

OLS Results for Cornwell and Rupert

```
+----------------------------------------------------+
| Residuals    Sum of squares        =   522.2008     |
|              Standard error of e   =   .3544712     |
| Fit          R-squared             =   .4112099     |
|              Adjusted R-squared    =   .4100766     |
+----------------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	5.40159723	.04838934	111.628	.0000	
EXP	.04084968	.00218534	18.693	.0000	19.8537815
EXPSQ	-.00068788	.480428D-04	-14.318	.0000	514.405042
OCC	-.13830480	.01480107	-9.344	.0000	.51116447
SMSA	.14856267	.01206772	12.311	.0000	.65378151
MS	.06798358	.02074599	3.277	.0010	.81440576
FEM	-.40020215	.02526118	-15.843	.0000	.11260504
UNION	.09409925	.01253203	7.509	.0000	.36398559
ED	.05812166	.00260039	22.351	.0000	12.8453782

Alternative Variance Estimators

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]
Constant	5.40159723	.04838934	111.628	.0000
EXP	.04084968	.00218534	18.693	.0000
EXPSQ	-.00068788	.480428D-04	-14.318	.0000
OCC	-.13830480	.01480107	-9.344	.0000
SMSA	.14856267	.01206772	12.311	.0000
MS	.06798358	.02074599	3.277	.0010
FEM	-.40020215	.02526118	-15.843	.0000
UNION	.09409925	.01253203	7.509	.0000
ED	.05812166	.00260039	22.351	.0000
Robust - Cluster				
Constant	5.40159723	.10156038	53.186	.0000
EXP	.04084968	.00432272	9.450	.0000
EXPSQ	-.00068788	.983981D-04	-6.991	.0000
OCC	-.13830480	.02772631	-4.988	.0000
SMSA	.14856267	.02423668	6.130	.0000
MS	.06798358	.04382220	1.551	.1208
FEM	-.40020215	.04961926	-8.065	.0000
UNION	.09409925	.02422669	3.884	.0001
ED	.05812166	.00555697	10.459	.0000

Generalized Least Squares

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}]^{-1}[\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}]$$

$$= [\Sigma_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\Omega}_i^{-1}\mathbf{X}_i]^{-1}[\Sigma_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\Omega}_i^{-1}\mathbf{y}_i]$$

$$\boldsymbol{\Omega}_i^{-1} = \frac{1}{\sigma_\varepsilon^2}\left[I_{T_i} - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T_i\sigma_u^2}\mathbf{ii}' \right]$$

(note, depends on i only through T_i)

Generalized Least Squares

GLS is equivalent to OLS regression of

$y_{it}* = y_{it} - \theta_i \bar{y}_{i}.$ on $\mathbf{x}_{it}* = \mathbf{x}_{it} - \theta_i \bar{\mathbf{x}}_{i}.$,

where $\theta_i = 1 - \dfrac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T_i \sigma_u^2}}$

$$\text{Asy.Var}[\hat{\boldsymbol{\beta}}] = [\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}]^{-1} = \sigma_\varepsilon^2[\mathbf{X}'*\mathbf{X}*]^{-1}$$

Estimators for the Variances

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} + u_i$$

Using the OLS estimator of $\boldsymbol{\beta}$, $\mathbf{b}_{OLS}$,

$$\frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}(y_{it} - a - \mathbf{x}_{it}'\mathbf{b})^2}{\left(\sum_{i=1}^{N} T_i\right) - 1 - K} \text{ estimates } \sigma_\varepsilon^2 + \sigma_U^2$$

With the LSDV estimates, a_i and $\mathbf{b}_{LSDV}$,

$$\frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}(y_{it} - a_{\mathbf{i}} - \mathbf{x}_{it}'\mathbf{b})^2}{\left(\sum_{i=1}^{N} T_i\right) - N - K} \text{ estimates } \sigma_\varepsilon^2$$

Using the difference of the two,

$$\left[\frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}(y_{it} - a - \mathbf{x}_{it}'\mathbf{b})^2}{\left(\sum_{i=1}^{N} T_i\right) - 1 - K}\right] - \left[\frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}(y_{it} - a_{\mathbf{i}} - \mathbf{x}_{it}'\mathbf{b})^2}{\left(\sum_{i=1}^{N} T_i\right) - N - K}\right] \text{ estimates } \sigma_U^2$$

Practical Problems with FGLS

- The preceding regularly produce negative estimates of σ_u^2.
- Estimation is made very complicated in unbalanced panels.

A bulletproof solution

From the robust LSDV estimator: $\hat{\sigma}_\varepsilon^2 = \dfrac{\Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i}(y_{it} - a_i - \mathbf{x}_{it}'\mathbf{b}_{LSDV})^2}{\Sigma_{i=1}^{N} T_i}$

From the pooled OLS estimator: $Est(\sigma_\varepsilon^2 + \sigma_u^2) = \dfrac{\Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i}(y_{it} - a_{OLS} - \mathbf{x}_{it}'\mathbf{b}_{OLS})^2}{\Sigma_{i=1}^{N} T_i} \geq \hat{\sigma}_\varepsilon^2$

$$\hat{\sigma}_u^2 = \frac{\Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i}(y_{it} - a_{OLS} - \mathbf{x}_{it}'\mathbf{b}_{OLS})^2 - \Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i}(y_{it} - a_i - \mathbf{x}_{it}'\mathbf{b}_{LSDV})^2}{\Sigma_{i=1}^{N} T_i} \geq 0$$

Stata Variance Estimators

$$\hat{\sigma}_\varepsilon^2 = \frac{\Sigma_{i=1}^N \Sigma_{t=1}^{T_i} (y_{it} - a_i - \mathbf{x}_{it}'\mathbf{b}_{LSDV})^2}{\Sigma_{i=1}^N T_i - K - N} > 0 \text{ based on FE estimates}$$

$$\hat{\sigma}_u^2 = \text{Max}\left[0, \frac{\text{SSE(group means)}}{N - A} - \frac{(N-K)\hat{\sigma}_\varepsilon^2}{(N-A)\bar{\bar{T}}}\right] \geq 0$$

where A = K or if $\hat{\sigma}_u^2$ is negative,

A=trace of a matrix that somewhat resembles $\mathbf{I}_K$.

Many other adjustments exist. None guaranteed to be positive. No optimality properties or even guaranteed consistency.

Other Variance Estimators

From the group means regression: $\sigma_\varepsilon^2 / \bar{T} + \sigma_u^2 = \dfrac{\sum_{i=1}^N (\bar{y}_{it} - \tilde{a} - \bar{\mathbf{x}}_i' \tilde{\mathbf{b}}_{MEANS})^2}{N - K - 1}$

(Wooldridge) Based on $E[w_{it}w_{is} \mid \mathbf{X}_i] = \sigma_u^2$ if $t \neq s$, $\hat{\sigma}_u^2 = \dfrac{\sum_{i=1}^N \sum_{t=1}^{T_i - 1} \sum_{s=t+1}^{T_i} \hat{w}_{it}\hat{w}_{is}}{\sum_{i=1}^N T_i - K - N}$

There are many others. Generally if the original, standard choices fail, these will also.

x´ does not contain a constant term in the preceding.

Fixed Effects Estimates

```
----------------------------------------------------------------------
Least Squares with Group Dummy Variables..........
LHS=LWAGE     Mean                 =         6.67635
Residuals     Sum of squares       =        82.34912
              Standard error of e  =          .15205
These  2 variables have no within group variation.
FEM      ED
F.E. estimates are based on a generalized inverse.
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
     EXP|     .11346***      .00247        45.982   .0000      19.8538
   EXPSQ|    -.00042***    .544864D-04     -7.789   .0000      514.405
     OCC|    -.02106         .01373        -1.534   .1251       .51116
    SMSA|    -.04209**       .01934        -2.177   .0295       .65378
      MS|    -.02915         .01897        -1.536   .1245       .81441
     FEM|      .000        ......(Fixed Parameter).......
   UNION|     .03413**       .01491         2.290   .0220       .36399
      ED|      .000        ......(Fixed Parameter).......
--------+-------------------------------------------------------------
```

Computing Variance Estimators

Using the full list of variables (FEM and ED are time invariant)

OLS sum of squares = 522.2008.

$\sigma_\varepsilon^2 + \sigma_u^2$ = 522.2008 / (4165 - 9) = 0.12565.

Using full list of variables and a generalized inverse (same as dropping FEM and ED), LSDV sum of squares = 82.34912.

$\widehat{\sigma_\varepsilon^2}$ = 82.34912 / (4165 - 8-595) = 0.023119.

$\widehat{\sigma_u^2}$ = 0.12565 - 0.023119 = 0.10253

Both estimators are positive. We stop here. If $\widehat{\sigma_u^2}$ were negative, we would use estimators without DF corrections.

Application

```
-----------------------------------------------------------------------
Random Effects Model: v(i,t)     = e(i,t) + u(i)
Estimates:  Var[e]               =        .023119
            Var[u]               =        .102531
            Corr[v(i,t),v(i,s)] =        .816006
Lagrange Multiplier Test vs. Model (3) =3713.07
( 1 degrees of freedom, prob. value =  .000000)
(High values of LM favor FEM/REM over CR model)
Fixed vs. Random Effects (Hausman)      =     .00   (Cannot be computed)
( 8 degrees of freedom, prob. value = 1.000000)
(High (low) values of H favor F.E.(R.E.) model)
            Sum of Squares            1411.241136
            R-squared                    -.591198
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
EXP	.08819204	.00224823	39.227	.0000	19.8537815
EXPSQ	-.00076604	.496074D-04	-15.442	.0000	514.405042
OCC	-.04243576	.01298466	-3.268	.0011	.51116447
SMSA	-.03404260	.01620508	-2.101	.0357	.65378151
MS	-.06708159	.01794516	-3.738	.0002	.81440576
FEM	-.34346104	.04536453	-7.571	.0000	.11260504
UNION	.05752770	.01350031	4.261	.0000	.36398559
ED	.11028379	.00510008	21.624	.0000	12.8453782
Constant	4.01913257	.07724830	52.029	.0000	

Testing for Effects: An LM Test

Breusch and Pagan Lagrange Multiplier statistic

$$y_{it} = \beta' x_{it} + u_i + \varepsilon_{it},\ u_i \text{ and } \varepsilon_{it} \sim \text{Normal}\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_u^2 & 0\\ 0 & \sigma_\varepsilon^2\end{pmatrix}\right]$$

$$H_0:\ \sigma_u^2 = 0$$

General

$$LM = \frac{(\Sigma_{i=1}^N T_i)^2}{2\Sigma_{i=1}^N T_i(T_i - 1)}\left[\frac{\Sigma_{i=1}^N (T_i\,\bar{e}_i\,)^2}{\Sigma_{i=1}^N \Sigma_{t=1}^T e_{it}^2} - 1\right]^2 \longrightarrow \chi^2[1]$$

Balanced Panel

$$LM = \frac{NT}{2(T\text{-}1)}\left[\frac{\Sigma_{i=1}^N [(T\bar{e}_i^2) - \mathbf{e}_i'\mathbf{e}_i]}{\Sigma_{i=1}^N \mathbf{e}_i'\mathbf{e}_i}\right]^2$$

Application: Cornwell-Rupert

```
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=LWAGE    Mean                 =    6.676346    |
|              Standard deviation   =    .4615122    |
| Model size   Parameters           =           7    |
|              Degrees of freedom   =        4158    |
| Residuals    Sum of squares       =    556.3030    |
|              Standard error of e  =    .3657745    |
| Fit          R-squared            =    .3727592    |
|              Adjusted R-squared   =    .3718541    |
+----------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant       5.66098218       .04685914    120.808   .0000
 FEM           -.39478212       .02603413    -15.164   .0000    .11260504
 ED             .05688005       .00267743     21.244   .0000   12.8453782
 OCC           -.11220205       .01464317     -7.662   .0000    .51116447
 SMSA           .15504405       .01233744     12.567   .0000    .65378151
 MS             .09569050       .02133490      4.485   .0000    .81440576
 EXP            .01043785       .00054206     19.256   .0000   19.8537815
+--------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)     |
| Estimates:  Var[e]              =   .235368D-01  |
|             Var[u]              =   .110254D+00  |
|             Corr[v(i,t),v(i,s)] =   .824078      |
| Lagrange Multiplier Test vs. Model (3) = 3797.07 |
| ( 1 df, prob value =  .000000)                   |
| (High values of LM favor FEM/REM over CR model.) |
+--------------------------------------------------+
 Constant       4.24669585       .07763394     54.702   .0000
 FEM           -.34715010       .04681514     -7.415   .0000    .11260504
 ED             .11120152       .00525209     21.173   .0000   12.8453782
 OCC           -.03908144       .01298962     -3.009   .0026    .51116447
 SMSA          -.03881553       .01645862     -2.358   .0184    .65378151
 MS            -.06557030       .01815465     -3.612   .0003    .81440576
 EXP            .05737298       .00088467     64.852   .0000   19.8537815
```

Testing for Effects

```
Regress;  lhs=lwage;rhs=fixedx,varyingx;res=e$
Matrix  ;  tebar=7*gxbr(e,person)$
Calc     ;  list;lm=595*7/(2*(7-1))*
                 (tebar'tebar/sumsqdev - 1)^2$

         LM     =  3797.06757
```

A Hausman Test for FE vs. RE

Estimator	Random Effects $E[c_i \mid \mathbf{X}_i] = 0$	Fixed Effects $E[c_i \mid \mathbf{X}_i] \neq 0$
FGLS (Random Effects)	Consistent and Efficient	Inconsistent
LSDV (Fixed Effects)	Consistent Inefficient	Consistent Possibly Efficient

Computing the Hausman Statistic

$$\text{Est.Var}[\hat{\boldsymbol{\beta}}_{FE}] = \hat{\sigma}^2_{\varepsilon}\left[\Sigma_{i=1}^{N}\mathbf{X}_i'\left(\mathrm{I} - \frac{1}{T_i}\mathbf{ii}'\right)\mathbf{X}_i\right]^{-1}$$

$$\text{Est.Var}[\hat{\boldsymbol{\beta}}_{RE}] = \hat{\sigma}^2_{\varepsilon}\left[\Sigma_{i=1}^{N}\mathbf{X}_i'\left(\mathrm{I} - \frac{\hat{\gamma}_i}{T_i}\mathbf{ii}'\right)\mathbf{X}_i\right]^{-1}, 0 \le \hat{\gamma}_i = \frac{T_i\hat{\sigma}^2_u}{\hat{\sigma}^2_{\varepsilon} + T_i\hat{\sigma}^2_u} \le 1$$

As long as $\hat{\sigma}^2_{\varepsilon}$ and $\hat{\sigma}^2_u$ are consistent, as N $\rightarrow \infty$, $\text{Est.Var}[\hat{\boldsymbol{\beta}}_{FE}] - \text{Est.Var}[\hat{\boldsymbol{\beta}}_{RE}]$ will be nonnegative definite. In a finite sample, to ensure this, both must be computed using the same estimate of $\hat{\sigma}^2_{\varepsilon}$. The one based on LSDV will generally be the better choice.

Note that columns of zeros will appear in $\text{Est.Var}[\hat{\boldsymbol{\beta}}_{FE}]$ if there are time invariant variables in **X**.

β does not contain the constant term in the preceding.

Hausman Test

```
+--------------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)           |
| Estimates:  Var[e]              =    .235368D-01       |
|             Var[u]              =    .110254D+00       |
|             Corr[v(i,t),v(i,s)] =    .824078           |
| Lagrange Multiplier Test vs. Model (3) = 3797.07       |
| ( 1 df, prob value =  .000000)                         |
| (High values of LM favor FEM/REM over CR model.)       |
| Fixed vs. Random Effects (Hausman)      = 2632.34      |
| ( 4 df, prob value =  .000000)                         |
| (High (low) values of H favor FEM (REM).)              |
+--------------------------------------------------------+
```

Fixed Effects

```
+----------------------------------------------------+
| Panel:Groups   Empty        0,   Valid data     595 |
|                Smallest     7,   Largest          7 |
|                Average group size              7.00 |
| There are  2 vars. with no within group variation. |
| ED        FEM                                       |
| Look for huge standard errors and fixed parameters.|
| F.E. results are based on a generalized inverse.   |
| They will be highly erratic. (Problematic model.)  |
| Unable to compute std.errors for dummy var. coeffs.|
+----------------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
WKS	.00083	.00060003	1.381	.1672	46.811525
OCC	-.02157	.01379216	-1.564	.1178	.5111645
IND	.01888	.01545450	1.221	.2219	.3954382
SOUTH	.00039	.03429053	.011	.9909	.2902761
SMSA	-.04451**	.01939659	-2.295	.0217	.6537815
UNION	.03274**	.01493217	2.192	.0283	.3639856
EXP	.11327***	.00247221	45.819	.0000	19.853782
EXPSQ	-.00042***	.546283D-04	-7.664	.0000	514.40504
ED	.000	......(Fixed Parameter).......			
FEM	.000	......(Fixed Parameter).......			

Random Effects

```
+--------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)     |
| Estimates:  Var[e]              =   .235368D-01  |
|             Var[u]              =   .110254D+00  |
|             Corr[v(i,t),v(i,s)] =   .824078      |
| Lagrange Multiplier Test vs. Model (3) = 3797.07 |
| ( 1 df, prob value =  .000000)                   |
| (High values of LM favor FEM/REM over CR model.) |
+--------------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
WKS	.00094	.00059308	1.586	.1128	46.811525
OCC	-.04367***	.01299206	-3.361	.0008	.5111645
IND	.00271	.01373256	.197	.8434	.3954382
SOUTH	-.00664	.02246416	-.295	.7677	.2902761
SMSA	-.03117*	.01615455	-1.930	.0536	.6537815
UNION	.05802***	.01349982	4.298	.0000	.3639856
EXP	.08744***	.00224705	38.913	.0000	19.853782
EXPSQ	-.00076***	.495876D-04	-15.411	.0000	514.40504
ED	.10724***	.00511463	20.967	.0000	12.845378
FEM	-.24786***	.04283536	-5.786	.0000	.1126050
Constant	3.97756***	.08178139	48.637	.0000	

The Hausman Test, by Hand

```
--> matrix;  br=b(1:8) ; vr=varb(1:8,1:8)$
--> matrix ; db = bf - br ; dv = vf - vr $
--> matrix ; list ; h =db'<dv>db$

Matrix H           has  1 rows and  1 columns.
                     1
         +--------------
        1| 2523.64910

--> calc;list;ctb(.95,8)$
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 Result  =     15.507313
```

Hello, professor greene.

I’ve taken the liberty of attaching some LIMDEP output in order to ask your view on whether my Hausman test stat is “large,” requiring the FEM, or not, allowing me to use the (much better for my research) REM.

Specifically, my test statistic, corrected for heteroscedasticity, is about 34 and significant with 6 df.

I considered this a large value until I found your “assignment 2” on the internet which shows a value of 2554 with 4 df. Now, I’d like to assert that 34/6 is a small value.

Variable Addition

A Fixed Effects Model

$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$

LSDV estimator - Deviations from group means:

To estimate $\boldsymbol{\beta}$, regress $(y_{it} - \bar{y}_i)$ on $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$

Algebraic equivalent: OLS regress y_{it} on $(\mathbf{x}_{it}, \bar{\mathbf{x}}_i)$

Mundlak interpretation: $\alpha_i = \alpha + \boldsymbol{\delta}'\bar{\mathbf{x}}_i + u_i$

Model becomes $y_{it} = \alpha + \boldsymbol{\delta}'\bar{\mathbf{x}}_i + u_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it}$

$$= \alpha + \boldsymbol{\delta}'\bar{\mathbf{x}}_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it} + u_i$$

= a random effects model with the group means.

Estimate by FGLS.

A Variable Addition Test

- Asymptotic equivalent to Hausman
- Also equivalent to Mundlak formulation
- In the random effects model, using FGLS
  - Only applies to time varying variables
  - Add expanded group means to the regression (i.e., observation i,t gets same group means for all t.
  - Use Wald test to test for coefficients on means equal to 0. Large chi-squared weighs against random effects specification.

Means Added to REM - Mundlak

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
WKS	.00083	.00060070	1.380	.1677	46.811525
OCC	-.02157	.01380769	-1.562	.1182	.5111645
IND	.01888	.01547189	1.220	.2224	.3954382
SOUTH	.00039	.03432914	.011	.9909	.2902761
SMSA	-.04451**	.01941842	-2.292	.0219	.6537815
UNION	.03274**	.01494898	2.190	.0285	.3639856
EXP	.11327***	.00247500	45.768	.0000	19.853782
EXPSQ	-.00042***	.546898D-04	-7.655	.0000	514.40504
ED	.05199***	.00552893	9.404	.0000	12.845378
FEM	-.41306***	.03732204	-11.067	.0000	.1126050
WKSB	.00863**	.00363907	2.371	.0177	46.811525
OCCB	-.14656***	.03640885	-4.025	.0001	.5111645
INDB	.04142	.02976363	1.392	.1640	.3954382
SOUTHB	-.05551	.04297816	-1.292	.1965	.2902761
SMSAB	.21607***	.03213205	6.724	.0000	.6537815
UNIONB	.08152**	.03266438	2.496	.0126	.3639856
EXPB	-.08005***	.00533603	-15.002	.0000	19.853782
EXPSQB	-.00017	.00011763	-1.416	.1567	514.40504
Constant	5.19036***	.20147201	25.762	.0000	

Wu (Variable Addition) Test

```
--> matrix ; bm=b(12:19);vm=varb(12:19,12:19)$
--> matrix ; list ; wu = bm'<vm>bm $

Matrix WU       has  1 rows and  1 columns.
                 1
        +--------------
       1|  3004.38076
```

LSDV is a Control Function Estimator

$$
\begin{aligned}
y_{it} &= \mathbf{x}_{it}'\boldsymbol{\beta} + c_i + \varepsilon_{it} \\
&= \mathbf{x}_{it}'\boldsymbol{\beta} + (c_i + \varepsilon_{it}) \\
&= \mathbf{x}_{it}'\boldsymbol{\beta} + w_{it}
\end{aligned}
$$

$$\text{Cov}[\mathbf{x}_{it}, w_{it}] = \text{Cov}[\mathbf{x}_{it}, (c_i + \varepsilon_{it})] = \mathbf{g}(\mathbf{x}_{it}) \neq \mathbf{0}$$

$\mathbf{x}_{it}$ is correlated with the FEs embedded in w_{it}.

LS regression of **y** on **X** is inconsistent because **X** is correlated with **w**. We seek a control function **h**(.) such that **X**|**h**(.) is uncorrelated with **w**. (In the presence of **h**(.), **X** is not correlated with **w**.)

Using the Frisch-Waugh theorem

$$\mathbf{b} = [\mathbf{X}'\mathbf{M_D}\mathbf{X}]^{-1}\left[\mathbf{X}'\mathbf{M_D}\mathbf{y}\right]$$

Consider regression of y on $[\mathbf{X},\bar{\mathbf{X}}]$. I.e., add group means to the regression.

LSDV is a Control Function Estimator

Consider regression of y on $[\mathbf{X},\bar{\mathbf{X}}]$. I.e., add group means to the regression.

$$[\mathbf{X},\bar{\mathbf{X}}] = \left[\begin{array}{cccc|cccc} \mathbf{x}_{11} & \mathbf{x}_{12} & \cdots & \mathbf{x}_{1K} & \bar{x}_{11}.\mathbf{i}_1 & \bar{x}_{12}.\mathbf{i}_1 & \cdots & \bar{x}_{11}\mathbf{i}_1 \\ \mathbf{x}_{21} & \mathbf{x}_{22} & \cdots & \mathbf{x}_{2K} & \bar{x}_{21}.\mathbf{i}_2 & \bar{x}_{22}.\mathbf{i}_2 & \cdots & \bar{x}_{11}\mathbf{i}_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{N1} & \mathbf{x}_{N2} & \cdots & \mathbf{x}_{NK} & \bar{x}_{N1}.\mathbf{i}_N & \bar{x}_{N2}.\mathbf{i}_N & \cdots & \bar{x}_{NK}.\mathbf{i}_N \end{array}\right]$$

$$= [\mathbf{X}, (\mathbf{I}\text{-}\mathbf{M}_{\mathbf{D}})\mathbf{X}]$$

$$= [\mathbf{X}, \mathbf{P}_{\mathbf{D}}\mathbf{X}]$$

$$= [\mathbf{X},\mathbf{F}]$$

LSDV is a Control Function Estimator

Using the Frisch-Waugh theorem

$$\mathbf{b}_{\text{ControlFunction}} = [\mathbf{X}'\mathbf{M_F X}]^{-1}\left[\mathbf{X}'\mathbf{M_F y}\right]$$

$$\mathbf{X}'\mathbf{M_F X} = \mathbf{X}'[\mathbf{I} - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}']\mathbf{X}$$

$$= \mathbf{X}'[\mathbf{I} - \mathbf{P}_D\mathbf{X}(\mathbf{X}'\mathbf{P}_D'\mathbf{P}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}_D']\mathbf{X}$$

$\mathbf{P}_D$ is symmetric and idempotent. And $\mathbf{P}_D = \mathbf{I}\text{-}\mathbf{M}_D$

$$= \mathbf{X}'[\mathbf{I} - (\mathbf{I}\text{-}\mathbf{M}_D)\mathbf{X}(\mathbf{X}'(\mathbf{I}\text{-}\mathbf{M}_D)\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}\text{-}\mathbf{M}_D)]\mathbf{X}$$

Multiply this out in full and collect some terms

$$=\mathbf{X}'\mathbf{I}\mathbf{X} - \left(\mathbf{X}'(\mathbf{I}\text{-}\mathbf{M}_D)\mathbf{X}\right)\left(\mathbf{X}'(\mathbf{I}\text{-}\mathbf{M}_D)\mathbf{X}\right)^{-1}\mathbf{X}'(\mathbf{I}\text{-}\mathbf{M}_D)\mathbf{X}$$

The two large matrices cancel. One more step

$$=\mathbf{X}'\mathbf{X} - \mathbf{X}'(\mathbf{I}\text{-}\mathbf{M}_D)\mathbf{X} = \mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{X} + \mathbf{X}'\mathbf{M}_D\mathbf{X}$$

$= \mathbf{X}'\mathbf{M}_D\mathbf{X}$. Likewise, $\left[\mathbf{X}'\mathbf{M_F y}\right] = \left[\mathbf{X}'\mathbf{M}_D\mathbf{y}\right]$. Therefore,

$$\mathbf{b}_{\text{ControlFunction}} = \mathbf{b}_{\text{LSDV}}$$

LSDV least squares with fixed effects ....

LWAGE	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
EXP	.09694***	.00119	81.53	.0000	.09461	.09927
WKS	.00114*	.00060	1.90	.0581	-.00004	.00233

***, **, * ==> Significance at 1%, 5%, 10% level.
Model was estimated on Jul 03, 2017 at 03:44:03 PM

Ordinary least squares regression ....

LWAGE	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
Constant	5.97472***	.09287	64.34	.0000	5.79271	6.15674
EXP	.09694***	.00319	30.35	.0000	.09068	.10320
WKS	.00114	.00162	.71	.4806	-.00203	.00432
EXPBAR	-.09096***	.00325	-28.00	.0000	-.09733	-.08459
WKSBAR	.01131***	.00254	4.46	.0000	.00634	.01628

Random Effects Model: v(i,t) = e(i,t) + u(i)
Estimates: Var[e] = .023564
Var[u] = .146433

LWAGE	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
EXP	.09694***	.00119	81.51	.0000	.09461	.09927
WKS	.00114*	.00060	1.89	.0582	-.00004	.00233
EXPBAR	-.09096***	.00189	-48.03	.0000	-.09467	-.08725
WKSBAR	.01131**	.00488	2.32	.0205	.00174	.02087
Constant	5.97472***	.23064	25.90	.0000	5.52267	6.42678

A Hierarchical Linear Model Interpretation of the FE Model

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + c_i + \varepsilon_{it}, \text{ (}\mathbf{x}\text{ does not contain a constant)}$$

$$E[\varepsilon_{it}|\mathbf{X}_i, c_i] = 0, \text{Var}[\varepsilon_{it}|\mathbf{X}_i, c_i] = \sigma_\varepsilon^2$$

$$c_i = \alpha + \mathbf{z}_i'\boldsymbol{\delta} + u_i,$$

$$E[u_i|\mathbf{z}_i'] = 0, \ \text{Var}[u_i|\mathbf{z}_i'] = \sigma_u^2$$

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + [\alpha + \mathbf{z}_i'\boldsymbol{\delta} + u_i] + \varepsilon_{it}$$

Hierarchical Linear Model as REM

```
+--------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)     |
| Estimates:  Var[e]              =   .235368D-01  |
|             Var[u]              =   .110254D+00  |
|             Corr[v(i,t),v(i,s)] =   .824078      |
|             Sigma(u)            =  0.3303        |
+--------------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
OCC	-.03908144	.01298962	-3.009	.0026	.51116447
SMSA	-.03881553	.01645862	-2.358	.0184	.65378151
MS	-.06557030	.01815465	-3.612	.0003	.81440576
EXP	.05737298	.00088467	64.852	.0000	19.8537815
FEM	-.34715010	.04681514	-7.415	.0000	.11260504
ED	.11120152	.00525209	21.173	.0000	12.8453782
Constant	4.24669585	.07763394	54.702	.0000	

Evolution: Correlated Random Effects

Unknown parameters

$$y_{it} = \alpha_i + \beta'\mathbf{x}_{it} + \varepsilon_{it}, \quad \Theta = [\alpha_1, \alpha_2, ..., \alpha_N, \beta, \sigma_\varepsilon^2]$$

Standard estimation based on LS (dummy variables)

Ambiguous definition of the distribution of y_{it}

Effects model, nonorthogonality, heterogeneity

$$y_{it} = \alpha_i + \beta'\mathbf{x}_{it} + \varepsilon_{it}, \quad E[\alpha_i \mid \mathbf{X}_i] = g(\mathbf{X}_i) \neq 0$$

Contrast to random effects $E[\alpha_i \mid X_i] = \alpha$

Standard estimation (still) based on LS (dummy variables)

Correlated random effects, more detailed model

$$y_{it} = \alpha_i + \beta'\mathbf{x}_{it} + \varepsilon_{it}, \quad P[\alpha_i \mid \mathbf{X}_i] = g(\mathbf{X}_i) \neq 0$$

Linear projection? $\alpha_i = \boldsymbol{\theta}'\mathbf{x}_i + u_i \quad Cor(u_i, \mathbf{x}_i) = 0$

Mundlak's Estimator

Mundlak, Y., "On the Pooling of Time Series and Cross Section Data, Econometrica, 46, 1978, pp. 69-85.

Write $c_i = \bar{\mathbf{x}}_i'\boldsymbol{\delta} + u_i$, $E[c_i \mid \mathbf{x}_{i1}, \mathbf{x}_{i1}, \ldots \mathbf{x}_{iT_i}] = \bar{\mathbf{x}}_i'\boldsymbol{\delta}$

Assume c_i contains all time invariant information

$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + c_i\mathbf{i} + \boldsymbol{\varepsilon}_i$, T_i observations in group i

$= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{i}\bar{\mathbf{x}}_i'\boldsymbol{\delta} + \boldsymbol{\varepsilon}_i + u_i\mathbf{i}$

Looks like random effects.

$\text{Var}[\boldsymbol{\varepsilon}_i + u_i\mathbf{i}] = \boldsymbol{\Omega}_i + \boldsymbol{\sigma}_u^2\mathbf{ii}'$

This is the model we used for the Wu test.

Mundlak's Approach for an FE Model with Time Invariant Variables

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_{\mathbf{i}}'\boldsymbol{\delta} + c_i + \varepsilon_{it}, \text{ (}\mathbf{x}\text{ does not contain a constant)}$$

$$E[\varepsilon_{it}|\mathbf{X}_i, c_i] = 0, \text{Var}[\varepsilon_{it}|\mathbf{X}_i, c_i] = \sigma_\varepsilon^2$$

$$c_i = \alpha + \bar{\mathbf{x}}_{\mathbf{i}}'\boldsymbol{\theta} + w_i,$$

$$E[w_i|\mathbf{X}_i, \mathbf{z}_{\mathbf{i}}] = 0, \ \text{Var}[w_i|\mathbf{X}_i, \mathbf{z}_{\mathbf{i}}] = \sigma_w^2$$

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_{\mathbf{i}}'\boldsymbol{\delta} + \alpha + \bar{\mathbf{x}}_{\mathbf{i}}'\boldsymbol{\theta} + w_i + \varepsilon_{it}$$

$=$ random effects model including group means of time varying variables.

Mundlak Form of FE Model

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
x(i,t)					
OCC	-.02021384	.01375165	-1.470	.1416	.51116447
SMSA	-.04250645	.01951727	-2.178	.0294	.65378151
MS	-.02946444	.01915264	-1.538	.1240	.81440576
EXP	.09665711	.00119262	81.046	.0000	19.8537815
z(i)					
FEM	-.34322129	.05725632	-5.994	.0000	.11260504
ED	.05099781	.00575551	8.861	.0000	12.8453782
Means of x(i,t) and constant					
Constant	5.72655261	.10300460	55.595	.0000	
OCCB	-.10850252	.03635921	-2.984	.0028	.51116447
SMSAB	.22934020	.03282197	6.987	.0000	.65378151
MSB	.20453332	.05329948	3.837	.0001	.81440576
EXPB	-.08988632	.00165025	-54.468	.0000	19.8537815
Variance Estimates					
Var[e]	.0235632				
Var[u]	.0773825				

Panel Data Extensions

- Dynamic models: lagged effects of the dependent variable
- Endogenous RHS variables
- Cross country comparisons– large T
- More general parameter heterogeneity – not only the constant term
- Nonlinear models such as binary choice

The Hausman and Taylor Model

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

Model: **x2** and **z2** are correlated with u.

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 $K_1 \geq L_2$.

H&T's 4 Step FGLS Estimator

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

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

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

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

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

H&T's 4 STEP IV Estimator

Instrumental Variables $\mathbf{V_i}$ =

$(\mathbf{x1_{it}} - \overline{\mathbf{x1_i}})$ = 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)

$\overline{\mathbf{x1_i}}$ = K_1 additional instrumental variables.

Now do 2SLS of $\mathbf{y}^*$ on $\mathbf{W}^*$ with instruments $\mathbf{V}$ to estimate all parameters. I.e.,

$$[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2] = (\hat{\mathbf{W}}^{*\prime}\hat{\mathbf{W}}^*)^{-1}\hat{\mathbf{W}}^{*\prime}\mathbf{y}^*.$$

TABLE 13.3 Estimated Log Wage Equations

	Variables	*OLS*	*GLS/RE*	*LSDV*	*HT/IV-GLS*	*HT/IV-GLS*
$\mathbf{x}_1$	Experience	0.0132	0.0133	0.0241	0.0217	
		(0.0011)[a]	(0.0017)	(0.0042)	(0.0031)	
	Bad health	−0.0843	−0.0300	−0.0388	−0.0278	−0.0388
		(0.0412)	(0.0363)	(0.0460)	(0.0307)	(0.0348)
	Unemployed	−0.0015	−0.0402	−0.0560	−0.0559	
	Last Year	(0.0267)	(0.0207)	(0.0295)	(0.0246)	
	Time	NR[b]	NR	NR	NR	NR
$\mathbf{x}_2$	Experience					0.0241
						(0.0045)
	Unemployed					−0.0560
						(0.0279)
$\mathbf{z}_1$	Race	−0.0853	−0.0878		−0.0278	−0.0175
		(0.0328)	(0.0518)		(0.0752)	(0.0764)
	Union	0.0450	0.0374		0.1227	0.2240
		(0.0191)	(0.0296)		(0.0473)	(0.2863)
	Schooling	0.0669	0.0676			
		(0.0033)	(0.0052)			
	Constant	NR	NR	NR	NR	NR
$\mathbf{z}_2$	Schooling				0.1246	0.2169
					(0.0434)	(0.0979)
	σ_ε	0.321	0.192	0.160	0.190	0.629
	$\rho = \sqrt{\sigma_u^2/(\sigma_u^2 + \sigma_\varepsilon^2)}$		0.632		0.661	0.817
	Spec. Test [3]		20.2		2.24	0.00

[a]Estimated asymptotic standard errors are given in parentheses.
[b]NR indicates that the coefficient estimate was not reported in the study.

Arellano/Bond/Bover's Formulation Builds on Hausman and Taylor

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

Instrumental variables for period t

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

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

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

$\overline{\mathbf{x1}}_i \quad = K_1$ additional instrumental variables. $K_1 \geq L_2$.

Let $v_{it} = \varepsilon_{it} + u_i$

Let $\mathbf{z}_{it}' = [(\mathbf{x1}_{it} - \overline{\mathbf{x1}}_i)',(\mathbf{x2}_{it} - \overline{\mathbf{x2}}_i)',\mathbf{z1}_i',\overline{\mathbf{x1}}']$

Then $E[\mathbf{z}_{it}v_{it}] = \mathbf{0}$

We formulate this for the T_i observations in group i.

Arellano/Bond/Bover's Formulation Adds a Lagged DV to H&T

$$y_{it} = \delta y_{i,t-1} + \mathbf{x1}_{it}'\boldsymbol{\beta}_1 + \mathbf{x2}_{it}'\boldsymbol{\beta}_2 + \mathbf{z1}_i'\boldsymbol{\alpha}_1 + \mathbf{z2}_i'\boldsymbol{\alpha}_2 + \varepsilon_{it} + u_i$$

Parameters : $\boldsymbol{\theta} = [\delta, \boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2']'$

The data

$$\mathbf{y}_i = \begin{bmatrix} y_{i,2} \\ y_{i,3} \\ \vdots \\ y_{i,T_i} \end{bmatrix}, \quad \mathbf{X}_i = \begin{bmatrix} y_{i,1} & \mathbf{x1}_{i2}' & \mathbf{x2}_{i2}' & \mathbf{z1}_i' & \mathbf{z2}_i' \\ y_{i,2} & \mathbf{x1}_{i3}' & \mathbf{x2}_{i3}' & \mathbf{z1}_i' & \mathbf{z2}_i' \\ & & & & \\ y_{i,T-1} & \mathbf{x1}_{iT_i}' & \mathbf{x2}_{iT_i}' & \mathbf{z1}_i' & \mathbf{z2}_i' \end{bmatrix}, \quad T_i\text{-1 rows}$$

1 K1 K2 L1 L2 columns

This formulation is the same as H&T with $y_{i,t-1}$ contained in $x2_{it}$.

Dynamic (Linear) Panel Data (DPD) Models

- Application
- Bias in Conventional Estimation
- Development of Consistent Estimators
- Efficient GMM Estimators

Dynamic Linear Model

Balestra-Nerlove (1966), 36 States, 11 Years

Demand for Natural Gas

Structure

New Demand: $G^*_{i,t} = G_{i,t} - (1-\delta)G_{i,t-1}$

Demand Function $G^*_{i,t} = \beta_1 + \beta_2 P_{i,t} + \beta_3 \Delta N_{i,t} + \beta_4 N_{i,t} + \beta_5 \Delta Y_{i,t} + \beta_6 Y_{i,t} + \varepsilon_{i,t}$

G=gas demand

N = population

P = price

Y = per capita income

Reduced Form

$$G_{i,t} = \beta_1 + \beta_2 P_{i,t} + \beta_3 \Delta N_{i,t} + \beta_4 N_{i,t} + \beta_5 \Delta Y_{i,t} + \beta_6 Y_{i,t} + \beta_7 G_{i,t-1} + \alpha_i + \varepsilon_{i,t}$$

A General DPD model

$$y_{i,t} = \mathbf{x}'_{i,t}\boldsymbol{\beta} + \boxed{\delta y_{i,t-1}} + \boxed{c_i} + \varepsilon_{i,t}$$

$$E[\varepsilon_{i,t} \mid \mathbf{X}_i, c_i] = 0$$

$$E[\varepsilon_{i,t}^2 \mid \mathbf{X}_i, c_i] = \sigma_\varepsilon^2, \;\; E[\varepsilon_{i,t}\varepsilon_{i,s} \mid \mathbf{X}_i, c_i] = 0 \text{ if } t \neq s.$$

$$E[c_i \mid \mathbf{X}_i] = g(\mathbf{X}_i)$$

No correlation across individuals

OLS and GLS are both inconsistent.

Arellano and Bond Estimator

Base on first differences

$$y_{i,t} - y_{i,t-1} = (\mathbf{x}_{i,t} - \mathbf{x}_{i,t-1})'\boldsymbol{\beta} + \delta(y_{i,t-1} - y_{i,t-2}) + (\varepsilon_{i,t} - \varepsilon_{i,t-1})$$

Instrumental variables

$$y_{i,3} - y_{i,2} = (\mathbf{x}_{i,3} - \mathbf{x}_{i,2})'\boldsymbol{\beta} + \delta(y_{i,2} - y_{i,1}) + (\varepsilon_{i,3} - \varepsilon_{i,2})$$

Can use y_{i1}

$$y_{i,4} - y_{i,3} = (\mathbf{x}_{i,4} - \mathbf{x}_{i,3})'\boldsymbol{\beta} + \delta(y_{i,3} - y_{i,2}) + (\varepsilon_{i,4} - \varepsilon_{i,3})$$

Can use $y_{i,1}$ and y_{i2}

$$y_{i,5} - y_{i,4} = (\mathbf{x}_{i,5} - \mathbf{x}_{i,4})'\boldsymbol{\beta} + \delta(y_{i,4} - y_{i,3}) + (\varepsilon_{i,5} - \varepsilon_{i,4})$$

Can use $y_{i,1}$ and y_{i2} and $y_{i,3}$

Arellano and Bond Estimator

More instrumental variables - Predetermined X

$$y_{i,3} - y_{i,2} = (\mathbf{x}_{i,3} - \mathbf{x}_{i,2})'\boldsymbol{\beta} + \delta(y_{i,2} - y_{i,1}) + (\varepsilon_{i,3} - \varepsilon_{i,2})$$

Can use y_{i1} and $\mathbf{x}_{i,1}, \mathbf{x}_{i,2}$

$$y_{i,4} - y_{i,3} = (\mathbf{x}_{i,4} - \mathbf{x}_{i,3})'\boldsymbol{\beta} + \delta(y_{i,3} - y_{i,2}) + (\varepsilon_{i,4} - \varepsilon_{i,3})$$

Can use $y_{i,1}, y_{i2}, \mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \mathbf{x}_{i,3}$

$$y_{i,5} - y_{i,4} = (\mathbf{x}_{i,5} - \mathbf{x}_{i,4})'\boldsymbol{\beta} + \delta(y_{i,4} - y_{i,3}) + (\varepsilon_{i,5} - \varepsilon_{i,4})$$

Can use $y_{i,1}, y_{i2}, y_{i,3}, \mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \mathbf{x}_{i,3}, \mathbf{x}_{i,4}$

Arellano and Bond Estimator

Even more instrumental variables - Strictly exogenous X

$y_{i,3} - y_{i,2} = (\mathbf{x}_{i,3} - \mathbf{x}_{i,2})'\boldsymbol{\beta} + \delta(y_{i,2} - y_{i,1}) + (\varepsilon_{i,3} - \varepsilon_{i,2})$

Can use y_{i1} and $\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, ..., \mathbf{x}_{i,T}$ (all periods)

$y_{i,4} - y_{i,3} = (\mathbf{x}_{i,4} - \mathbf{x}_{i,3})'\boldsymbol{\beta} + \delta(y_{i,3} - y_{i,2}) + (\varepsilon_{i,4} - \varepsilon_{i,3})$

Can use $y_{i,1}, y_{i2}, \mathbf{x}_{i,1}, \mathbf{x}_{i,2}, ..., \mathbf{x}_{i,T}$

$y_{i,5} - y_{i,4} = (\mathbf{x}_{i,5} - \mathbf{x}_{i,4})'\boldsymbol{\beta} + \delta(y_{i,4} - y_{i,3}) + (\varepsilon_{i,5} - \varepsilon_{i,4})$

Can use $y_{i,1}, y_{i2}, y_{i,3}, \mathbf{x}_{i,1}, \mathbf{x}_{i,2}, ..., \mathbf{x}_{i,T}$

The number of potential instruments is huge.

These define the rows of $\mathbf{Z}_i$. These can be used for simple instrumental variable estimation.

Application: Maquiladora

The U.S. and Mexico: Are We Still Connected?

Federal Reserve Bank of Dallas, El Paso Branch

Network of Border Economics (Red de la Economía Fronteriza)

Centro de Investigación y Docencia Económicas A.C.

Houston, Texas. November 18, 2005

Maquila: volatility and Mexico-US economic integration

Gustavo Félix Verduzco
Centro de Investigaciones Socioeconómicas
Universidad Autónoma de Coahuila
gfelix@cise.uadec.mx

http://www.dallasfed.org/news/research/2005/05us-mexico_felix.pdf

Maquiladora

Model: Labor Demand in Maquila Industry

Dynamic Panel Data:

$$Ltrab_{it} = \alpha_0 + \alpha_1 Ltrab_{i(t-1)} + \alpha_2 Ltrab_{i(t-2)} + \beta_1 Lrppd_{it} + \beta_2 Lpibusa_{it} + v_i + u_{it}$$

t= 1990.1 – 2005.3 quarterly

i = The Following 13 States where maquila mainly operates: Baja California, Sonora, Chihuahua, Coahuila, Nuevo León, Tamaulipas, Durango, Aguascalientes, Jalisco, Guanajuato, Mexico-DF, Puebla y Yucatán.

Variables:

Ltrab= log of maquila employment

Lrppd = wage per worker in dollars

Lpibusa = log of: USA GDP (2000 prices) over distance

Estimates

Model: Labor Demand in Maquila Industry

```
Arellano-Bond dynamic panel-data estimation     Number of obs      =       695
Group variable (i): estado                      Number of groups   =        13
                                                Wald chi2(4)       =  18500.45
Time variable (t): trim                         Obs per group: min =        35
                                                               avg =  53.46154
                                                               max =        59

One-step results
------------------------------------------------------------------------------
D.ltrab      |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ltrab        |
          LD |   1.220175   .0362107    33.70   0.000     1.149204    1.291147
         L2D |   -.262198   .0355168    -7.38   0.000    -.3318095   -.1925864
lrppd        |
          D1 |  -.0804483   .0115187    -6.98   0.000    -.1030246   -.0578721
lpibusa      |
          D1 |   .4801248   .1643802     2.92   0.003     .1579454    .8023041
_cons        |  -.0023032   .0012531    -1.84   0.066    -.0047592    .0001528
------------------------------------------------------------------------------
Sargan test of over-identifying restrictions:
         chi2(1827) =   695.25    Prob > chi2 = 1.0000
Arellano-Bond test that average autocovariance in residuals of order 1 is 0:
         H0: no autocorrelation   z = -13.42   Pr > z = 0.0000
Arellano-Bond test that average autocovariance in residuals of order 2 is 0:
     H0: no autocorrelation  z = -1.30  Pr > z = 0.1927
```