Panel Data Methods
3: Long Panels - Basics

© A. Colin Cameron
Univ. of Calif. Davis

.

Based on A. Colin Cameron and Pravin K. Trivedi (2009, 2010), Microeconometrics using Stata (MUS), Stata Press. and A. Colin Cameron and Pravin K. Trivedi (2005), Microeconometrics: Methods and Applications (MMA), C.U.P.

November 2013

1. Introduction

- These slides focus on static regressions with predetermined regressors.
  - then y_{it} on $\mathbf{x}_{it}$ is consistent even if error u_{it} is serially correlated
  - whereas e.g. y_{it} on y_{it-1} is inconsistent if error u_{it} is serially correlated and stationary
  - dynamics, especially unit roots and cointegration are deferred to separate slides
- For **long panels** asymptotics are for $T \rightarrow \infty$
  - A dynamic model for the errors may be specified, such as AR(1) error
  - Errors may be correlated over individuals
  - Individual-specific effects can be modelled as individual dummies
  - Presentation follows Cameron and Trivedi (2010) chapter 8.10.
- We also introduce heterogeneous panels
  - coefficients can vary over i.

- **Heterogeneous** panels can be more flexible as $T \longrightarrow \infty$
  - To date focused on just intercept α_i varying in short panels
  - Now can allow slopes to differ across individuals
  - Additionally may control for correlation over i due to common shocks or spatial correlation.
  - Again defer complication of unit roots and cointegration to separate slides
  - See Breitung and Pesaran (2008) for extensive survey.

Outline

2. Panel Data Example: Cigarettes Sales

- U.S. state-year panel $N = 10$ states and $T = 30$ (1963-92)
- Source: Baltagi, Griffin and Xiong (2000).
- Goal: estimate response of per capita cigarette consumption (lnc) to real cigarette prices (lnp).
  - Price varies across states, due in large part to different levels of taxation, as well as over time.

Data Description

- We will regress `lnc` on `lnp`, `lnpmin` and `lny` (own price, other price, income)

```
. * Description of cigarette dataset
. use mus08cigar.dta, clear

. describe

Contains data from mus08cigar.dta
  obs:           300
 vars:             6                          26 Nov 2008 17:14
 size:         9,600 (99.9% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
state           float  %9.0g                  U.S. state
year            float  %9.0g                  Year 1963 to 1992
lnp             float  %9.0g                  Log state real price of pack of cigarettes
lnpmin          float  %9.0g                  Log of min real price in adjoining states
lnc             float  %9.0g                  Log state cigarette sales in packs per capita
lny             float  %9.0g                  Log state per capita disposable income
```

Summary Statistics

```
. * Summary of cigarette dataset
. summarize, separator(6)

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       state |       300         5.5     2.87708          1         10
        year |       300        77.5    8.669903         63         92
         lnp |       300    4.518424    .1406979   4.176332    4.96916
      lnpmin |       300      4.4308    .1379243     4.0428   4.831303
         lnc |       300    4.792591    .2071792   4.212128   5.690022
         lny |       300    8.731014    .6942426   7.300023    10.0385

.
. * Pooled GLS with error correlated across states and state-specific AR(1)
. xtset state year
       panel variable:  state (strongly balanced)
        time variable:  year, 63 to 92
                delta:  1 unit
```

Within versus Between Variation

- Within variation is generally greater.

```
. xtsum lnc lnp lny lnpmin
```

Variable		Mean	Std. Dev.	Min	Max	Observations
lnc	overall	4.792591	.2071792	4.212128	5.690022	N = 300
	between		.1328465	4.643053	5.037004	n = 10
	within		.1642764	4.164742	5.445609	T = 30
lnp	overall	4.518424	.1406979	4.176332	4.96916	N = 300
	between		.0572717	4.451445	4.623794	n = 10
	within		.1297459	4.23715	4.972691	T = 30
lny	overall	8.731014	.6942426	7.300023	10.0385	N = 300
	between		.1924964	8.477204	9.019063	n = 10
	within		.6697103	7.553833	9.817204	T = 30
lnpmin	overall	4.4308	.1379243	4.0428	4.831303	N = 300
	between		.059411	4.313441	4.521246	n = 10
	within		.1258403	4.149695	4.840707	T = 30

Data Autocorrelations

- Looks like AR(1) will model `lnc` and `lnp` well.

```
. pwcorr lnc L.lnc L2.lnc L3.lnc L4.lnc L5.lnc L6.lnc
```

	lnc	L.lnc	L2.lnc	L3.lnc	L4.lnc	L5.lnc	L6.lnc
lnc	1.0000						
L.lnc	0.9740	1.0000					
L2.lnc	0.9373	0.9721	1.0000				
L3.lnc	0.8888	0.9351	0.9717	1.0000			
L4.lnc	0.8346	0.8882	0.9374	0.9737	1.0000		
L5.lnc	0.7749	0.8367	0.8934	0.9404	0.9738	1.0000	
L6.lnc	0.7126	0.7836	0.8499	0.9008	0.9442	0.9747	1.0000

```
. pwcorr lnp L.lnp L2.lnp L3.lnp L4.lnp L5.lnp L6.lnp
```

	lnp	L.lnp	L2.lnp	L3.lnp	L4.lnp	L5.lnp	L6.lnp
lnp	1.0000						
L.lnp	0.8893	1.0000					
L2.lnp	0.7570	0.8810	1.0000				
L3.lnp	0.5886	0.7196	0.8688	1.0000			
L4.lnp	0.4481	0.5443	0.6993	0.8619	1.0000		
L5.lnp	0.3280	0.3972	0.5120	0.6802	0.8537	1.0000	
L6.lnp	0.2080	0.2743	0.3613	0.4852	0.6640	0.8480	1.0000

3. Pooled OLS and FGLS

- FGLS is more efficient and gives correct standard errors if have correct model for the error

$$\widehat{\beta} = (\mathbf{X}'\widehat{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\Omega}^{-1}\mathbf{y}.$$

- Problem: Wide range of models for how u_{it} varies over i and t.
- For short panels $N \rightarrow \infty$ analysis was made easy by assuming
  - independence over i
  - finite T so correlation over t becomes a minor nuisance
  - use `xtgee` or `xtreg, pa` with option `vce(robust)`
- For long panels $(T \rightarrow \infty)$ need to be more explicit in modelling correlation over t
  - also some methods allow for possibility of correlation over i
  - `xtgls`, `xtpcse`, `xtscc`

Example: AR(1) error

- Model
  - $y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}$; with N small; $T \rightarrow \infty$
  - regressors $\mathbf{x}_{it}$ can include individual dummies but not time dummies
  - AR(1) error $u_{it} = \rho_i u_{i,t-1} + \varepsilon_{it}$
    - ε_{it} is uncorrelated over t but potentially correlated over i
- Estimator that controls for error AR(1) but not for correlation over i
  - Cochrane-Orcutt transformation (or variation Prais-Winsten)
  - $y_{it} - \rho_i y_{i,t-1} = (\mathbf{x}_{it} - \rho_i x_{i,t-1})'\boldsymbol{\beta} + \varepsilon_{it}$ (since $u_{it} - \rho_i u_{i,t-1} = \varepsilon_{it}$)
  - OLS of y_{it} on $\mathbf{x}_{it}$ gives consistent (inefficient) $\widehat{\boldsymbol{\beta}}_{OLS}$
  - Then $\widehat{\rho}_i$ from OLS of $\widehat{u}_{it}$ on $\widehat{u}_{i,t-1}$ where $\widehat{u}_{it} = y_{it} - \mathbf{x}_{it}'\widehat{\boldsymbol{\beta}}_{OLS}$
  - Finally $\widehat{\boldsymbol{\beta}}_{FGLS}$ from OLS of $y_{it} - \widehat{\rho}_i y_{i,t-1}$ on $\mathbf{x}_{it} - \widehat{\rho}_i x_{i,t-1}$
- Robust variance matrix that allows for correlation over i
  - $\mathrm{V}[\widehat{\boldsymbol{\beta}}_{FGLS}] = (\mathbf{X}^{*\prime}\mathbf{X}^{*})^{-1} \times \sum_t \mathbf{X}_t^{*\prime}\widehat{\boldsymbol{\varepsilon}}_t^{*}\widehat{\boldsymbol{\varepsilon}}_t^{*\prime}\mathbf{X}_t^{*} \times (\mathbf{X}'\mathbf{X})^{-1}$
  - where $\mathbf{x}_{it}^{*} = \mathbf{x}_{it} - \widehat{\rho}_i x_{i,t-1}$; $\widehat{\varepsilon}_{it}^{*} = \mathbf{y}_{it}^{*} - \mathbf{x}_{it}^{*\prime}\widehat{\boldsymbol{\beta}}_{FGLS}$

xtgls Command

- `xtgls`
  - FGLS with very simple model over t
    - u_{it} is independent over t or AR(1) with parameter ρ_i or parameter ρ
  - $+$ FGLS allowing for correlation over i with limited heteroskedasticity
    - $\text{Var}[u_{it}] = \sigma_{ii}$ and $\text{Cov}[u_{it}, u_{jt}] = \sigma_{ij}$ do not vary with t but do over i.
  - standard errors are not robust - assumes correct Ω
  - original (1960's) approach - Kmenta
  - limitations
    - need $T > N$
    - underestimates standard errors with finite N, T even if correct Ω

- FGLS assuming error AR(1) and correlated over i
  - standard errors assume correct Ω

```
. xtgls lnc lnp lny lnpmin year, panels(correlated) corr(psar1)

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic with cross-sectional correlation
Correlation:   panel-specific AR(1)

Estimated covariances      =        55          Number of obs      =       300
Estimated autocorrelations =        10          Number of groups   =        10
Estimated coefficients     =         5          Time periods       =        30
                                                Wald chi2(4)       =    342.15
                                                Prob > chi2        =    0.0000
```

lnc	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnp	-.3260683	.0218214	-14.94	0.000	-.3688375	-.2832991
lny	.4646236	.0645149	7.20	0.000	.3381768	.5910704
lnpmin	.0174759	.0274963	0.64	0.525	-.0364159	.0713677
year	-.0397666	.0052431	-7.58	0.000	-.0500429	-.0294902
_cons	5.157994	.2753002	18.74	0.000	4.618416	5.697573

xtpcse Command

- `xtpcse` (pcse means "panel corrected standard errors")
- Estimator can control for error correlation over t **but not** i
  - does OLS
  - or does FGLS with AR(1) error with parameter ρ or ρ_i
- But variance matrix does control for error correlation over i
- Example: command `xtpcse y x, corr(psar1)`
- The estimator is not fully efficient as it does not control for across individual error correlation
  - but its finite sample (finite N, T) performance is better than `xtfgls` method

- FGLS assuming error AR(1) error (common ρ) & independent over i
  - but standard errors then control for error contemporaneous correlated over i

```
. xtpcse lnc lnp lny lnpmin year, corr(psar1)
(note: estimates of rho outside [-1,1] bounded to be in the range [-1,1])

Prais-Winsten regression, correlated panels corrected standard errors (PCSEs)

Group variable:   state                       Number of obs      =       300
Time variable:    year                        Number of groups   =        10
Panels:           correlated (balanced)       Obs per group: min =        30
Autocorrelation:  panel-specific AR(1)                       avg =        30
                                                             max =        30
Estimated covariances      =        55        R-squared          =    0.9807
Estimated autocorrelations =        10        Wald chi2(4)       =     65.29
Estimated coefficients     =         5        Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |           Panel-corrected
         lnc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnp |  -.2674266   .0485877    -5.50   0.000    -.3626567   -.1721964
         lny |   .4820779   .1324463     3.64   0.000     .2224879    .7416679
      lnpmin |   .0796523   .0638603     1.25   0.212    -.0455116    .2048162
        year |  -.0449654    .010324    -4.36   0.000    -.0652002   -.0247307
       _cons |   4.879688   .5575454     8.75   0.000     3.786919    5.972456
-------------+----------------------------------------------------------------
        rhos =  .9474947   .8697296          1  .9991112  .9571484 ...   .991599
------------------------------------------------------------------------------
```

Cluster Robust se's for OLS

- Recall for OLS with independence over i and $N \rightarrow \infty$ and T fixed

$$\mathrm{V}[\widehat{\beta}_{\mathrm{OLS}}] = \left[\sum_i \mathbf{X}_i{}'\mathbf{X}_i\right]^{-1} \left[\sum_i \mathrm{E}[\mathbf{X}_i{}'\mathbf{u}_i\mathbf{u}_i'\mathbf{X}_i']\right] \left[\sum_i \mathbf{X}_i{}'\mathbf{X}_i\right]^{-1}.$$

$$\widehat{\mathrm{V}}[\widehat{\beta}_{\mathrm{OLS}}] = \left[\sum_i \mathbf{X}_i{}'\mathbf{X}_i\right]^{-1} \sum_i \mathbf{X}_i{}'\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{X}_i' \left[\sum_i \mathbf{X}_i{}'\mathbf{X}_i\right]^{-1}.$$

- What if $T \rightarrow \infty$?
- Hansen (2007) considers several cases of N, T with independence over i
  - 1. $N \rightarrow \infty$, T fixed. Original short panel case. $\widehat{\mathrm{V}}[\widehat{\beta}_{\mathrm{OLS}}]$ okay.
  - 2. $N \rightarrow \infty$, $T \rightarrow \infty$. $\widehat{\mathrm{V}}[\widehat{\beta}_{\mathrm{OLS}}]$ okay for either mixing or no mixing in time.
  - 3. N fixed, $T \rightarrow \infty$ problem. Need to ensure that mixing, so correlations over time go to zero. Use HAC.
- Bottom line
  - can use the usual panel-robust if $T \rightarrow \infty$ in addition to $N \rightarrow \infty$.

Spatial and Cluster Robust se's

- For $T \rightarrow \infty$ the HAC extends to spatially correlated (correlated over i)
  - Driscoll and Kraay (1998) and add-on command `xtscc`
- Begin with pure time series background:
  - $y_t = \mathbf{x}_t' \beta + u_t$
  - $\text{Cor}[u_t, u_{t-k}] \neq 0$ for $k \leq m$ and $\text{Cor}[u_t, u_{t-k}] = 0$ for $k > m$
  - Example is an MA(m) error.

- Then given errors correlated only up to m periods apart

$$
\begin{aligned}
\mathrm{V}[\widehat{\beta}_{OLS}] &= (\mathbf{X}'\mathbf{X})^{-1}\mathrm{V}\left[\sum\nolimits_t \mathbf{x}_t u_t\right](\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathrm{E}\left[\sum\nolimits_t \sum\nolimits_s \mathbf{x}_t u_t \mathbf{x}_s' u_s\right](\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\left[\sum\nolimits_t \sum\nolimits_{s;|t-s|<m} \mathrm{E}[\mathbf{x}_t u_t \mathbf{x}_s' u_s]\right](\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\left[\sum\nolimits_{k=-m}^{m} \sum\nolimits_t \mathrm{E}[\mathbf{x}_t u_t \mathbf{x}_{t-k}' u_{t-k}]\right](\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\left[\Omega_0 + \sum\nolimits_{k=1}^{m}(\Omega_k + \Omega_k')\right](\mathbf{X}'\mathbf{X})^{-1};\ \Omega_k = \sum\nolimits_t \mathrm{E}
\end{aligned}
$$

- Newey-West HAC (heteroskedasticity- and autocorrelation-consistent) variance matrix estimate

$$
\begin{aligned}
\widehat{\mathrm{V}}[\widehat{\beta}_{OLS}] &= (\mathbf{X}'\mathbf{X})^{-1}\left[\Omega_0 + \sum\nolimits_{k=1}^{m}(\widehat{\Omega}_k + \widehat{\Omega}_k')\right](\mathbf{X}'\mathbf{X})^{-1} \\
\widehat{\Omega}_k &= (1 - \frac{k}{m+1})\sum\nolimits_{t=k+1}^{T} \widehat{u}_t \widehat{u}_{t-k} \mathbf{x}_t \mathbf{x}_{t-k}'.
\end{aligned}
$$

Spatial and Cluster Robust se's

- For panel data the HAC extends to spatially correlated (correlated over i)
  - Driscoll and Kraay (1998) and add-on command xtscc

$$\begin{aligned}
\mathrm{V}[\widehat{\beta}_{OLS}] &= (\mathbf{X}'\mathbf{X})^{-1}\mathrm{V}\left[\sum_i \mathbf{X}_i{}'\mathbf{u}_i\right](\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathrm{V}\left[\sum_i \sum_t \mathbf{x}_{it}u_{it}\right](\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathrm{E}\left[\sum_t \sum_s \sum_i \sum_j \mathbf{x}_{it}u_{it}\mathbf{x}'_{js}u_{js}\right](\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\left[\sum_t \sum_{s;|t-s|<m} \sum_i \sum_j \mathrm{E}[\mathbf{x}_{it}u_{it}\mathbf{x}'_{js}u_{js}]\right](\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\left[\sum_{k=-m}^{m} \sum_t \sum_i \sum_j \mathrm{E}[\mathbf{x}_{it}u_{it}\mathbf{x}'_{j,t-k}u_{j,t-k}]\right](\mathbf{X}'\mathbf{X}
\end{aligned}$$

- where assume errors correlated only m periods apart (MA(m)) and $T \rightarrow \infty$.

- So

$$
\begin{aligned}
V[\widehat{\beta}_{OLS}] &= (\mathbf{X}'\mathbf{X})^{-1}\mathrm{E}\left[\Omega_0 + \textstyle\sum_{k=1}^{m}(\Omega_k + \Omega_k')\right](\mathbf{X}'\mathbf{X})^{-1} \\
\Omega_k &= \textstyle\sum_t \sum_i \sum_j \mathrm{E}[u_{it}u_{j,t-k}\mathbf{x}_{it}\mathbf{x}_{j,t-k}']
\end{aligned}
$$

- Then panel HAC variance estimate

$$
\begin{aligned}
\widehat{V}[\widehat{\beta}_{OLS}] &= (\mathbf{X}'\mathbf{X})^{-1}\mathrm{E}\left[\widehat{\Omega}_0 + \textstyle\sum_{k=1}^{m}(\widehat{\Omega}_k + \widehat{\Omega}_k')\right](\mathbf{X}'\mathbf{X})^{-1} \\
\widehat{\Omega}_k &= (1 - \frac{j}{m+1})\textstyle\sum_t \sum_i \sum_j \widehat{u}_{it}\widehat{u}_{j,t-k}\mathbf{x}_{it}\mathbf{x}_{j,t-k}'.
\end{aligned}
$$

- Requires $T \rightarrow \infty$
  - either N fixed or $N \rightarrow \infty$ is okay.
- Extends to FE estimator and to nonlinear m-estimators e.g. logit.

Comparisons

```
. * Comparison of various pooled OLS and GLS estimators
. quietly xtpcse lnc lnp lny lnpmin year, corr(ind) independent nmk

. estimates store OLS_iid

. quietly xtpcse lnc lnp lny lnpmin year, corr(ind)

. estimates store OLS_cor

. quietly xtscc lnc lnp lny lnpmin year, lag(4)

. estimates store OLS_DK

. quietly xtpcse lnc lnp lny lnpmin year, corr(ar1)

. estimates store AR1_cor

. quietly xtgls lnc lnp lny lnpmin year, corr(ar1) panels(iid)

. estimates store FGLSAR1

. quietly xtgls lnc lnp lny lnpmin year, corr(ar1) panels(correlated)

. estimates store FGLSCAR
```

- OLS_iid: OLS with default se's (assuming u_{it} is iid)
- OLS_iid: OLS with spatial but no HAC ($\simeq$ xtscc, lags(0))
- OLS_DK: OLS with spatial and HAC - has the largest se's
- AR1_corr: FGLS AR(1) error no spatial but spatial corrected se's
- FGLSAR1: FGLS AR(1) error no spatial and FGLS default se's
- FGLSCAR: FGLS AR(1) error and spatial and FGLS default se's
  - efficiency gain assuming correct model for errors.

```
. estimates table OLS_iid OLS_cor OLS_DK AR1_cor FGLSAR1 FGLSCAR, b(%7.3f) se
```

Variable	OLS_iid	OLS_cor	OLS_DK	AR1_cor	FGLSAR1	FGLSCAR
lnp	-0.583	-0.583	-0.583	-0.266	-0.264	-0.330
	0.129	0.169	0.273	0.049	0.049	0.026
lny	0.365	0.365	0.365	0.398	0.397	0.407
	0.049	0.080	0.163	0.125	0.094	0.080
lnpmin	-0.027	-0.027	-0.027	0.069	0.070	0.036
	0.128	0.166	0.252	0.064	0.059	0.034
year	-0.033	-0.033	-0.033	-0.038	-0.038	-0.037
	0.004	0.006	0.012	0.010	0.007	0.006
_cons	6.930	6.930	6.930	5.115	5.100	5.393
	0.353	0.330	0.515	0.544	0.414	0.361

legend: b/se

4. Individual Fixed Effects

- Individual Fixed Effects straightforward here
  - no incidental parameters problem as $T \rightarrow \infty$.
- But rather than start with OLS, **more efficient** to start with AR(1) error model
  - first transform to control for serial correlation:
    - $(y_{it} - \widehat{\rho} y_{i,t-1}) = (1 - \widehat{\rho})\alpha_i + (\mathbf{x}_{it} - \widehat{\rho}\mathbf{x}_{i,t-1})'\boldsymbol{\beta}$
  - then mean-difference to eliminate $(1 - \widehat{\rho})\alpha_i$
  - or assume random effect
- Command `xtregar` with options `fe` and `re`.

Interactive Effects

- Bai (2009) considers the model

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \lambda'_i F_t + \varepsilon_{it}$$

  - e.g. $F_t = \begin{bmatrix} 1 \\ \delta_t \end{bmatrix}$ and $\lambda_i = \begin{bmatrix} 1 \\ \delta_i \end{bmatrix}$ so $\lambda'_i F_t = \alpha_i + \delta_t$
  - λ_i and F_t are possibly correlated with $\mathbf{x}_{it}$
  - ε_{it} may be serially and cross-sectionally dependent
  - $N \rightarrow \infty$ and $T \rightarrow \infty$ with $T/N \rightarrow 0$ or $N/T \rightarrow 0$ or $T/N \rightarrow \rho > 0$
- Stack $\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{F}\boldsymbol{\lambda}_i + \boldsymbol{\varepsilon}_i$ and $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{F}\boldsymbol{\Lambda} + \boldsymbol{\varepsilon}$
  - normalization (necessary) $\mathbf{F}'\mathbf{F} = \mathbf{I}$ and $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}$ diagonal
  - $\widehat{\beta}$, $\widehat{\mathbf{F}}$, $\widehat{\boldsymbol{\Lambda}}$ minimize $\sum_i (\mathbf{Y}_i - \mathbf{X}_i \beta - \mathbf{F}\boldsymbol{\lambda}_i)'(\mathbf{Y}_i - \mathbf{X}_i \beta - \mathbf{F}\boldsymbol{\lambda}_i)$

5. Heterogeneous Panels: Introduction

- With $T \rightarrow \infty$ can allow coefficients to vary over cross-sections

$$y_{it} = \alpha_i + \mathbf{x}_{it}{}'\boldsymbol{\beta}_i + u_{it}.$$

- Also concerned about common shocks (f_t) over time
  - can't just put in time dummies (since $T \rightarrow \infty$)
  - time trends not enough.
- Here introduce estimators for static models
  - they were actually proposed for dynamic models
  - will return to them later consider dynamic models.
- Estimators considered
  - mean grouped estimator - Pesaran and Smith (1995)
  - common correlated effects estimator (for common shocks) - Pesaran (2006)
  - augmented mean group estimator (for common shocks) - Eberhardt and Teal (2009).
  - all can be estimated with Stata addon `xtmg`

- With N fixed (and "small") and $T \rightarrow \infty$ can run N separate regressions

$$y_{it} = \alpha_i + \mathbf{x}_{it}{}'\boldsymbol{\beta}_i + u_{it}.$$

  - assume no fixed effects problem due to either lack of correlation or $T \rightarrow \infty$.
  - and can test for poolability: $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all i.
- Can estimate using the `statsby` prefix command.

```
. * Run separate regressions for each state
. statsby _b, by(state) clear: regress lnc lnp lny lnpmin year
(running regress on estimation sample)

      command:  regress lnc lnp lny lnpmin year
           by:  state

Statsby groups
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..........
```

- Some variation but mostly sales rise as price falls and income rises.

```
. * Report regression coefficients for each state
. format _b* %9.2f

. list, clean

       state    _b_lnp    _b_lny   _b_lnp~n   _b_year   _b_cons
  1.       1     -0.36      1.10       0.24     -0.08      2.10
  2.       2      0.12      0.60      -0.45     -0.05      5.14
  3.       3     -0.20      0.76       0.12     -0.05      2.72
  4.       4     -0.52     -0.14      -0.21     -0.00      9.56
  5.       5     -0.55      0.71       0.30     -0.07      4.76
  6.       6     -0.11      0.21      -0.14     -0.02      6.20
  7.       7     -0.43     -0.07       0.18     -0.03      9.14
  8.       8     -0.26      0.89       0.08     -0.07      3.67
  9.       9     -0.03      0.55      -0.36     -0.04      4.69
 10.      10     -1.41      1.12       1.14     -0.08      2.70
```

Mean-Group Estimator

- Let $y_{it} = \alpha_i + \mathbf{x}_{it}{}'\boldsymbol{\beta}_i + u_{it} = \mathbf{w}_{it}'\boldsymbol{\delta}_i + u_{it}$
- The mean group estimator is $\widehat{\boldsymbol{\delta}} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\delta}}_i$
  - viewed as an estimator of $\boldsymbol{\delta} = \mathrm{E}[\boldsymbol{\delta}_i]$.
  - Pesaran and Smith (1995)
  - say use $\widehat{\mathrm{V}}[\widehat{\boldsymbol{\delta}}] = \frac{1}{N(N-1)}\sum_{i=1}^{N}(\widehat{\boldsymbol{\delta}}_i - \boldsymbol{\delta})(\widehat{\boldsymbol{\delta}}_i - \boldsymbol{\delta})'$.
- Default for user-written addon `xtmg` does this.

```
. * Mean-group estimator
. xtmg lnc lnp lny lnpmin year


Pesaran & Smith (1995) Mean Group estimator

All coefficients represent averages across groups (group variable: state)
Coefficient averages computed as unweighted means
```

```
Mean Group type estimation                       Number of obs      =       300
Group variable: state                            Number of groups   =        10

                                                 Obs per group: min =        30
                                                                avg =      30.0
                                                                max =        30

                                                 Wald chi2(4)       =    272.08
                                                 Prob > chi2        =    0.0000
```

lnc	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnp	-.3752339	.1336287	-2.81	0.005	-.6371414	-.1133264
lny	.5741365	.1409129	4.07	0.000	.2979524	.8503206
lnpmin	.0902451	.1419863	0.64	0.525	-.188043	.3685332
year	-.0509173	.0083476	-6.10	0.000	-.0672784	-.0345562
_cons	5.069284	.8177948	6.20	0.000	3.466436	6.672133

```
Root Mean Squared Error (sigma): 0.0436
```

- Can get estimates for each state (option `full`)
  - same as earlier `statsby`

Group-specific coefficients

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Group 1						
lnp	-.3572344	.2149646	-1.66	0.097	-.7785572	.0640884
lny	1.102418	.3575748	3.08	0.002	.4015847	1.803252
lnpmin	.2403018	.192765	1.25	0.213	-.1375107	.6181142
year	-.0806138	.0286329	-2.82	0.005	-.1367333	-.0244942
_cons	2.098109	1.463726	1.43	0.152	-.7707406	4.966959
Group 2						
lnp	.1158018	.1958822	0.59	0.554	-.2681204	.4997239
lny	.60493	.2204498	2.74	0.006	.1728563	1.037004
lnpmin	-.4469771	.1725522	-2.59	0.010	-.7851732	-.108781
year	-.054202	.0175606	-3.09	0.002	-.0886201	-.0197838
_cons	5.13989	1.069976	4.80	0.000	3.042775	7.237005

Random Coefficients Estimator

- A weighted average of the estimators that predated mean-group estimator.
  - stack so $\mathbf{y}_i = \mathbf{W}_i \delta_i + \mathbf{u}_i$
  - let $\delta_i = \delta + \mathbf{v}_i$ where $\mathrm{E}[\mathbf{v}_i | \mathbf{W}_i] = 0$
  - then $\mathbf{y}_i = \mathbf{W}_i \delta + (\mathbf{W}_i \mathbf{v}_i + \mathbf{u}_i)$
  - can do OLS of $\mathbf{y}_i$ on $\mathbf{W}_i$
  - Swamy (1970) proposed FGLS using a consistent estimate of the variances and covariances (over i) of the error $(\mathbf{W}_i \mathbf{v}_i + \mathbf{u}_i)$.
  - do this assume homoskedastic errors in each state but allowing correlation across states, so $\mathrm{Cov}[u_{it}, u_{jt}] = \sigma_{ij}$.
- Compared to MG, RC in theory will do better when there are outlying values of δ_i as it shrinks them towards zero.

```
. * Randon coefficients estimator
. xtrc lnc lnp lny lnpmin year

Random-coefficients regression                  Number of obs      =       300
Group variable: state                           Number of groups   =        10

                                                Obs per group: min =        30
                                                               avg =      30.0
                                                               max =        30

                                                Wald chi2(4)       =    170.13
                                                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         lnc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnp |  -.3727744   .1427454    -2.61   0.009    -.6525503   -.0929985
         lny |   .5351386   .1593635     3.36   0.001     .2227919    .8474852
      lnpmin |   .0696587    .151666     0.46   0.646    -.2276013    .3669187
        year |   -.047699   .0101616    -4.69   0.000    -.0676153   -.0277827
       _cons |   5.233397   .8821246     5.93   0.000     3.504465     6.96233
------------------------------------------------------------------------------
Test of parameter constancy:    chi2(45) =  3680.53      Prob > chi2 = 0.0000
```

Common Shocks

- The MG and RC estimator assume independence over individuals (here states)
  - putting in a time trend will help sop up some correlation but may not be enough.
- Correlation over i may be due to spatial considerations
  - neighboring states are similar
  - need to deal with by specifying a spatial model for errors
  - for $y_{it} = \mathbf{w}_{it}'\delta_i + u_{it}$ stack u_{it} into $\mathbf{u}_t$ and $\mathbf{u}_t = \mathbf{S}\varepsilon_t$, $\mathbf{S}$ is specified and ε_t is cross-sectionally independent
  - not covered here.
- Or it may be due to common shocks
  - a common way to deal with this is a factor model.

Common Factor Model

- Before $y_{it} = \mathbf{w}_{it}'\delta_i + u_{it}$
- Now suppose $y_{it} = \mathbf{w}_{it}'\delta_i + \mathbf{f}_t'\gamma_i + \varepsilon_{it}$
  - where $\varepsilon_{it} \sim [0, \sigma_i^2]$ are independent over i and $\mathbf{f}_t$ are "factors"
  - simplest case one factor $y_{it} = \mathbf{w}_{it}'\delta_i + f_t\gamma_i + \varepsilon_{it}$.
- The factors $\mathbf{f}_t$ are unobservable
  - they are data determined to fit well the correlations over i of $\mathbf{f}_t'\gamma_i + \varepsilon_{it}$
    - formally principal components are often used in economics
  - $\mathbf{f}_t$ possibly correlated with $\mathbf{w}_{it}$
  - so ignoring them could lead to inconsistent $\widehat{\delta}_i$
  - they are common to each individual
  - but are given different weights for different individuals via γ_i.

Aside: Principal Components

- Principal components reduces dimensionality.
- Consider the $N \times k$ matrix $\mathbf{X}$ and form $\mathbf{X}'\mathbf{X}$
  - let λ_j be the ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$
  - let $\mathbf{v}_j$ be the corresponding eigenvector
- The first principal component of $\mathbf{X}$ is $\frac{1}{\lambda_1}\mathbf{X}\mathbf{v}_1$
  - it explains $\lambda_1 / \sum_{j=1}^{k} \lambda_j^2$ of the total variation of $(\sum_{j=1}^{k} \sum_{i=1}^{N} \mathbf{X}_{ji}) = tr(\mathbf{X}'\mathbf{X})$
  - the second principal component is $\frac{1}{\lambda_2}\mathbf{X}\mathbf{v}_2$
- Results simplify if we first demean and use the correlation or covariance matrix
- And can interpret principal components as factor analysis with homoskedastic residuals: $\mathbf{X}_{ij} = \mathbf{f}_i'\gamma_j + \varepsilon_{ij}$.

Common Correlated Effects Estimator

- Pesaran (2006) shows that for large N, T can simply
  - add the time-averages of y_{it} and $\mathbf{x}_{it}$ as regressors in each of the N regressions
- So do OLS in $y_{it} = \mathbf{w}_{it}'\delta_i + \pi_{0i}\bar{y}_t + \overline{\mathbf{w}}_t'\boldsymbol{\pi}_i + u_{it}$
  - note that the coefficients of the time-averages vary with i
- Called the common correlated effects estimator.
- This is option cce of `xtmg`

```
. xtmg lnc lnp lny lnpmin year, cce


Pesaran (2006) Common Correlated Effects Mean Group estimator

All coefficients represent averages across groups (group variable: state)
Coefficient averages computed as unweighted means
```

- All coefficients below are averages over i

```
Mean Group type estimation                    Number of obs      =        300
Group variable: state                         Number of groups   =         10

                                              Obs per group: min =         30
                                                             avg =       30.0
                                                             max =         30

                                              Wald chi2(4)       =      30.76
                                              Prob > chi2        =     0.0000
```

lnc	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnp	-.4669704	.1063	-4.39	0.000	-.6753145	-.2586263
lny	.7238857	.2364108	3.06	0.002	.2605291	1.187242
lnpmin	.1101311	.0909168	1.21	0.226	-.0680625	.2883248
year	-.0047461	.0087429	-0.54	0.587	-.0218819	.0123897
lnc_avg	1.077181	.1482423	7.27	0.000	.7866314	1.367731
lnp_avg	.2925321	.2757283	1.06	0.289	-.2478854	.8329496
lny_avg	-.6450694	.2439767	-2.64	0.008	-1.123255	-.1668839
lnpmin_avg	.1229866	.1975405	0.62	0.534	-.2641856	.5101588
year_avg	(omitted)					
_cons	-.9610782	.7264015	-1.32	0.186	-2.384799	.4626426

```
Root Mean Squared Error (sigma): 0.0264
Cross-section averaged regressors are marked by the suffix avg.
```

Augmented Mean Group Estimator

- Bond and Eberhardt (2009) and Eberhardt and Teal (2010) propose an augmented mean group estimator
  - estimate the common factors and include them as explicit regressors
  - whereas CCE just controlled for their presence
  - option `augment` of `xtmg`

```
. xtmg lnc lnp lny lnpmin year, augment


Augmented Mean Group estimator (Bond & Eberhardt, 2009; Eberhardt & Teal, 2010)

Common dynamic process included as additional regressor
All coefficients represent averages across groups (group variable: state)
Coefficient averages computed as unweighted means
```

```
Mean Group type estimation                      Number of obs      =       300
Group variable: state                           Number of groups   =        10

                                                Obs per group: min =        30
                                                               avg =      30.0
                                                               max =        30

                                                Wald chi2(4)       =    300.99
                                                Prob > chi2        =    0.0000
```

lnc	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnp	-.5328879	.0828091	-6.44	0.000	-.6951906	-.3705851
lny	.7429772	.0768996	9.66	0.000	.5922568	.8936976
lnpmin	.0709764	.0926518	0.77	0.444	-.1106178	.2525706
year	-.0593132	.0070495	-8.41	0.000	-.0731301	-.0454964
c_d_p	1.138518	.1819006	6.26	0.000	.7819994	1.495037
_cons	4.99731	.4492237	11.12	0.000	4.116848	5.877772

```
Root Mean Squared Error (sigma): 0.0293
Variable c_d_p refers to the common dynamic process.
```

Comparison of Estimators

```
. * Compare estimators
. quietly xtscc lnc lnp lny lnpmin year, lag(4)

. estimates store OLS_DK

. quietly xtmg lnc lnp lny lnpmin year

. estimates store MG

. quietly xtrc lnc lnp lny lnpmin year

. estimates store RC

. quietly xtmg lnc lnp lny lnpmin year, cce

. estimates store CCE

. estimates table OLS_DK MG RC CCE, b(%7.3f) se(%7.2f)
```

- Biggest variation in coefficient of `lny`
  - note that Driscoll-Kraay standard errors are larger

Variable	OLS_DK	MG	RC	CCE
lnp	-0.583	-0.375	-0.373	-0.467
	0.27	0.13	0.14	0.11
lny	0.365	0.574	0.535	0.724
	0.16	0.14	0.16	0.24
lnpmin	-0.027	0.090	0.070	0.110
	0.25	0.14	0.15	0.09
year	-0.033	-0.051	-0.048	-0.005
	0.01	0.01	0.01	0.01
lnc_avg				1.077
				0.15
lnp_avg				0.293
				0.28
lny_avg				-0.645
				0.24
lnpmin_avg				0.123
				0.20
year_avg				0.000
				0.00
_cons	6.930	5.069	5.233	-0.961
	0.51	0.82	0.88	0.73

legend: b/se

6. Summary

- Basics for $T \rightarrow \infty$ and stationary

xtgls	Pooled OLS and FGLS
xtpcse	Pooled OLS and FGLS
xtscc	Spatial and cluster robust (Stata addon)
xtregar	FE and RE with AR(1) error

- With heterogeneous (and stationary) panels

xtmg	Mean-group and CCE (Stata addon)
xtrc	Random Coefficients

7. References

- Bai, J. (2009), "Panel data models with interactive fixed effects," Econometrica, 1229–1279.
- Eberhardt, M. (2012), "Estimating Panel Time Series Models with Heterogeneous Slopes," Stata Journal, 12, 61-71.
- Eberhardt, M., C. Helmers, and H. Strauss (2013), "Do Spillovers Matter When Estimating Private Returns to R&D?," Review of Economics and Statistics, 95, 436-448..
- Pesaran, M.H. (2006), "Estimation and inference in large heterogeneous panels with a multifactor error structure," Econometrica, 74, 967–1012.