2022 Essex Summer School
3K: Dynamics and Heterogeneity

Robert W. Walker, Ph. D.
Associate Professor of Quantitative Methods
Atkinson Graduate School of Management
Willamette University
Salem, Oregon USA
rwalker@willamette.edu

August 15, 2022

Outline for Day 6

- With models, we need a model for comparison.
- Some Useful Notation
- Fixed and Random Effects
- Comparing FE and RE with Hausman
- The Multilevel generalization

The Dimensions of TSCS/CSTS and Summary

- Presence of a time dimensions gives us a natural ordering.
- Space is not irrelevant under the same circumstances as time – nominal indices are irrelevant on some level. Defining space is hard. Ex. targeting of Foreign Direct Investment and defining proximity.
- ANOVA is informative in this two-dimensional setting.
- A part of any good data analysis is summary and characterization. The same is true here; let's look at some examples of summary in panel data settings.

Basic xt commands

In Stata's language, xt is the way that one naturally refers to CSTS/TSCS data. Consider NT observations on some random variable y_{it} where $i \in N$ and $t \in T$. The TSCS/CSTS commands almost always have this prefix.

- xtset: Declaring xt data
- xtdes: Describing xt data structure
- xtsum: Summarizing xt data
- xttab: Summarizing categorical xt data.
- xttrans: Transition matrix for xt data.

- `xtline`: Line graphs for `xt` data.

A Primitive Question

Given two-dimensional data, how should we break it down? The most common method is unit-averages; we break each unit's time series on each element into deviations from their own mean. This is called the within transform. The between portion represents deviations between the unit's mean and the overall mean. Stationarity considerations are generically implicit.

Some Useful Variances and Notation

- W(ithin) for unit i[1]:

$$W_i = \sum_{t=1}^{T} (x_{it} - \overline{x}_i)^2$$

- B(etween):

$$B_T = \sum_{i=1}^{N} (\overline{x}_i - \overline{x})^2$$

- T(otal):

$$T = \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \overline{x})^2$$

[1] Thus the total within variance would be a summary over all $i \in N$

Some Useful Notation

- The Kronecker Product $\otimes$: This is a simple way of condensing the notation for sets of matrices. It is important to note that conformity is not required. So, for a general matrix A_{kl} and B_{mn}, we can write

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & a_{13}B \\ a_{21}B & a_{22}B & a_{23}B \\ a_{31}B & a_{32}B & a_{33}B \end{pmatrix}$$

  with a result C of dimension $(km)(ln)$.

- The inverse of a Kronecker product is well defined [under invertibility conditions]

$$[A \otimes B]^{-1} = [A^{-1} \otimes B^{-1}]$$

- As are products of Kronecker products

$$(A \otimes B)(C \otimes D) = AC \otimes BD$$

Why is the Notation Useful?

Let A be a variance/covariance matrix across panels and B be the same matrix for a given panel. This is a fairly general way to conceive of a panel data problem.

Heteroscedasticity?

Temporal Autocorrelation?

Spatial Autocorrelation?

Heteroscedasticity

The homoscedastic case is described by $\sigma^2 I$.

The [unit] heteroscedastic case is described, generally, by $\sigma_i^2 I_N \otimes I_t$

$$\begin{pmatrix} \sigma_1^2 I_T & 0 & 0 & 0 \\ 0 & \sigma_2^2 I_T & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_N^2 I_T \end{pmatrix}$$

The ultimate result will be of dimension $NT \times NT$. The first T entries will be σ_1^2, entries $T+1$ to $2T$ will be $\sigma_2^2 \ldots$ and the entries $(N-1)T+1$ to NT will be σ_N^2. If we believed that the heteroscedasticity arose from time points rather than units, replace N with T and vice versa; i becomes t.

The Managable Autocorrelation Structure

$$\Phi = \sigma^2\Psi = \sigma_e^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{T-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \dots & 1 \end{pmatrix}$$

given that $e_t = \rho e_{t-1} + \nu_t$. A Toeplitz form....

This allows us to calculate the variance of e using results from basic statistics, i.e. $Var(e_t) = \rho^2 Var(e_{t-1}) + Var(\nu)$. If the variance is stationary, we can rewrite,

$$\sigma_e^2 = \frac{\sigma_\nu^2}{1-\rho^2}$$

Autocorrelation

When discussing heteroscedasticity, we notice that the off-diagonal elements are all zeroes. This is the assumption of no correlation among [somehow] adjacent elements. The somehow takes two forms: (1) spatial and (2) temporal. Just as before where time-induced heteroscedasticity simply involved interchanging N and T and i and t; the same idea prevails here.

Aitken's Theorem?

In a now-classic paper, Aitken generalized the Gauss-Markov theorem to the class of Generalized Least Squares estimators. It is important to note that these are GLS and not FGLS estimators. What is the difference? The two GLS estimators considered by Stimson are not strictly speaking GLS.

Definition:

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

Properties: (1) GLS is unbiased. (2) Consistent. (3) Asymptotically normal. (4) MV(L)UE

What does the feasible do?

We need to estimate things to replace unknown covariance structures and coverage will depend on properties of the estimators of these covariances. Consistent estimators will work but there is euphemistically “considerable variation” in the class of consistent estimators. Contrasting the Beck and Katz/White approach with the GLS approach is a valid difference in philosophies.[2]

[2] We will return to this when we look at Hausman because this is the essential issue.

The Beck and Katz solution

Beck and Katz take a different tack to the general data types in common use (long T). The basic idea is to generate estimates using OLS because GLS can be quite bad. [What do we need to be able to do this?]

- Locate a specification to purge serial correlation (in t).

p. 638 Construct the panel corrected standard error. Construct Σ $(N \times N)$ using

$$\hat{\Sigma} = \frac{\sum_{t=1}^{T} e_{it} e_{jt}}{T}.$$

Estimate the cross-sectional correlation matrix. Kronecker product this in $\mathbf{I}_T$ remembering how we got $\mathbf{I}$.

- Inference with OLS and PCSE in the spirit of White, really Huber (1967) but the key is separable moments. Brief diversion here about separability; it turns out the result yesterday is what gives rise to the appropriate intuition.

Thinking about robust and cluster

Every Stata user is familiar with this, it seems. Though not developed by Stata (but Hardin, a student of Huber), the two are synonymous. What would these look like in an application?

- just robust is unstructured heteroscedastic
- cluster utilizes the multidimensional axes

`xtgls` **and** `xtpcse`

Two significant options of note

- `panels(iid,heteroscedastic,correlated)`
- `correlation(ar1,psar1,independent)`

panels

- `iid`

$$\epsilon\epsilon' = \sigma^2 \mathbf{I}_{N\times N}$$

gives us homscedasticity and no spatial correlation; σ^2 is scalar.

- `heteroscedastic`

$$\epsilon\epsilon' = \sigma_i^2 \mathbf{I}_{N\times N}$$

gives us heteroscedasticity and no spatial correlation; σ_i^2 is an N-vector.

- `correlated`

$$\epsilon\epsilon' = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1N} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \dots & \sigma_{2N} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \dots & \sigma_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \sigma_{N3} & \dots & \sigma_N^2 \end{pmatrix}$$

gives us heteroscedastic and (contemporaneously) spatially correlated errors

correlation

- `independent`

$$\epsilon\epsilon' = \mathbf{I}_{T\times T}$$

gives us no autoregression.

- `ar1`

$$\epsilon\epsilon' = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & 1 \end{pmatrix}$$

gives us a global autoregressive parameter for the errors. In simple terms, all cross-sections share the same "level" of serial correlation.

- `psar1`

$$\epsilon\epsilon' = \begin{pmatrix} 1 & \rho_i & \rho_i^2 & \dots & \rho_i^{T-1} \\ \rho_i & 1 & \rho_i & \dots & \rho_i^{T-2} \\ \rho_i^2 & \rho_i & 1 & \dots & \rho_i^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_i^{T-1} & \rho_i^{T-2} & \rho_i^{T-3} & \dots & 1 \end{pmatrix}$$

gives us an autoregressive parameter for the errors that is unique to each cross-section. Each cross-section has a distinct “level” of serial correlation.

Unit Heterogeneity

Most discussions of panel data estimators draw on a fixed versus random effects distinction. The subtle distinction is important but perhaps overstated.

Definitions

Let's construct a general model:

$$y_{it} = \alpha_{it} + \mathbf{X}_{it}\beta_{it} + \epsilon_{it}$$

1. Pooled Model: $y_{it} = \alpha + \boldsymbol{X_{it}\beta} + \epsilon_{it}$

2. Year Dummies Model: $y_{it} = \boldsymbol{\alpha}_t + \boldsymbol{X_{it}\beta} + \epsilon_{it}$

3. (Two-way) LSDV: $y_{it} = \boldsymbol{\alpha_i} + \boldsymbol{\alpha_t} + \boldsymbol{X_{it}\beta} + \epsilon_{it}$

4. Unit Dummies Model: $y_{it} = \boldsymbol{\alpha_i} + \boldsymbol{X_{it}\beta} + \epsilon_{it}$

   (a) Fixed effects: $y_{it} - \bar{y}_i = \Delta_i \boldsymbol{X}_{it}\boldsymbol{\beta} + \Delta_i \epsilon_{it}$
   (b) Random effects $\boldsymbol{\alpha_i} \perp \boldsymbol{X_{it}}$: $\boldsymbol{\alpha_i} \sim [\alpha, \sigma^2_\alpha]\ \ \epsilon_{it} \sim [0, \sigma^2_\epsilon]$

Why does heterogeneity matter?

- If $\alpha \neq \alpha_i \forall i$, then serial correlation is induced in the errors. At a minimum, this implies incorrect standard errors for inference and inefficiency.

- If $\mathbb{E}[X_{it}\alpha_i] \neq 0$, then $(\alpha_i - \alpha)$ is an omitted variable with a consequent bias induced. We can draw a picture of this.

A brief simulation.

Some ANCOVA

- Pooled Slope and Intercepts
- Pooled Intercepts
- Pooled Slopes

Constructing Estimators

- Pooled Estimator

$$\hat{\beta}_T = T_{xx}^{-1} T_{xy} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

- Within Estimator

$$\hat{\beta}_W = W_{xx}^{-1} W_{xy}$$

- Between Estimator

$$\hat{\beta}_B = B_{\overline{xx}}^{-1} B_{\overline{xy}}$$

A Variation Identity

- $T = W_{xx} + W_{\overline{xx}}$
- In different notation, $T = W + B$ or $S^t = S^w + S^b$.

$$\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\overline{x})^2 = \sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}^2 - NT\overline{x}^2 \quad (2)$$

$$= \sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}^2 - T_i\overline{x}_i + T_i\overline{x}_i - NT\overline{x}^2 \quad (3)$$

$$= \underbrace{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\overline{x}_i)^2}_{W_i} + \underbrace{\sum_{i=1}^{N}T_i(\overline{x}_i-\overline{x})^2}_{B_T} \quad (4)$$

Back to ANCOVA

1. RSS from W_i with DF $= NT - NK - N$

2. RSS from W with DF $= NT - N - K$

3. RSS from T with DF $= NT - K - 1$

For total pooling, we can F-test $\frac{3-1}{1}$. This is the least and most restricted sets of models. If we can reject this, pooling is (perhaps) justified? Now let's construct some others. Suppose we reject total pooling. Is it intercepts, slopes, or both? Imposing a slope restriction gives us 2, the F we want is $\frac{2-1}{1}$. What do we get from $\frac{3-2}{2}$? NB: It's conditional. We can also do this with time. This is a good starting point, but it is not as clean as we might like.

OLS as Weighted Average

Walk through proof here.

$$\hat{\beta}_{OLS} = [S^t_{xx}]^{-1} S^t_{xy} \quad (5)$$

$$= [S^w_{xx} + S^b_{xx}]^{-1} (S^w_{xy} + S^b_{xy}) \quad (6)$$

$$= [S^w_{xx} + S^b_{xx}]^{-1} S^w_{xy} + [S^w_{xx} + S^b_{xx}]^{-1} S^b_{xy} \quad (7)$$

Let $F^w = [S^w_{xx} + S^b_{xx}]^{-1} S^w_{xx} \rightarrow F^b = I - F^w = [S^w_{xx} + S^b_{xx}]^{-1} S^b_{xx}$. My claim is that $\hat{\beta}_{OLS} = F^w \beta^w + F^b \beta^b$.

$$= [S^w_{xx} + S^b_{xx}]^{-1} \underbrace{S^w_{xx}[S^w_{xx}]^{-1}}_{I} S^w_{xy} + [S^w_{xx} + S^b_{xx}]^{-1} \underbrace{S^b_{xx}[S^b_{xx}]^{-1}}_{I} S^b_{xy}$$

$$= [S^w_{xx} + S^b_{xx}]^{-1} S^w_{xy} + [S^w_{xx} + S^b_{xx}]^{-1} S^b_{xy}$$

A Random Effects Estimator

- Assume that the unit means have some distribution rather than being some fixed constant.
- This allows (under normality) us to partition the global error into components.
- The method is the same, the difference is the weighting by a covariance matrix with a known structure.
- As we noted, there is a simple problem with the application of the OLS estimator if the error is correlated with the regressors.
- How might we think about remedying this?

Comparing Fixed and Random Effects

- The Hausman test: smart and broadly applicable idea. Wish it worked better... See V. E. Troeger.
- Mundlak's argument merits consideration.

Hausman's Idea

The basic idea is that the fixed effects estimator is consistent but potentially inefficient. The random effects estimator is only consistent under the null. We can leverage this to form a test in the Hausman family using the result proved in the paper. This is implemented in Stata using model storage capabilities.

- Estimate a consistent model
- Store the result as XXX.
- Estimate an efficient model
- Store the result as YYY.
- `hausman` XXX YYY

Mundlak

The basic idea behind Mundlak's paper is that the fixed versus random effects debate is ill conceived. Moreover, there is a "right model". Why and how?

- Conditional versus unconditional inference.
- FE problem is inefficiency.
- RE problem can be bias.
- Maybe we want an MSE criterion?
- As usual, N and T matter in size. Plug-in estimators in general.

Bell, Fairbrother, and Jones

Estimate a variant of the Mundlak model that accommodates all the concerns.

$$y_{it} = \beta_0 + \beta_{1W}(x_{it} - \overline{x}_i) + \beta_{2B}\overline{x}_i + \beta_3 z_i + (\nu_i + \epsilon_{it})$$

First-Differences

Define $\mathbf{\Delta}$ to be a difference operator so that we can define

$$\begin{aligned} \mathbf{\Delta X} &= \mathbf{X}_{it} - \mathbf{X}_{i,t-1} && (8) \\ \mathbf{\Delta y} &= \mathbf{y}_{it} - \mathbf{y}_{i,t-1} && (9) \end{aligned}$$

Observation: N(T-1) observations if $T_i \geq 2 \quad \forall i$. Equality case is interesting. The first-difference estimator is then:

$$\mathbf{\Delta y} = \boldsymbol{\beta}(\mathbf{\Delta X}) + \epsilon_{it}$$

And an OLS estimator would simply look like:

$$\hat{\beta} = (\mathbf{\Delta X}'\mathbf{\Delta X})^{-1}\mathbf{\Delta X}'\mathbf{\Delta y} \tag{10}$$

NB: For $T = 2$ show that FE is FD.

First Differences/Fixed Effects

Either transformation removes heterogeneity. The difference is that the two estimators operate at different orders of integration. The difference is not purely convenience; there is substance to this and theory can help. At the same time, the statistics matter.

Specification Testing and Interpretation in the Fixed Effects Model

- F-test of the dummy variables. What does this mean?
- Above can be done in one- and two- way frameworks.
- The substance depends on the first-order question. Under what conditions are first-order effects unbiased (we know this)? The RE/GLS approach works when the orthogonality is maintained.
- Example from Arellano, p. 40

Conditional versus unconditional prediction?

Stata Implementation

- `xtreg`: contains five estimators. For now, we will skip (`pa`).

  `be`: the between effects estimator.

$$\overline{y}_i = \overline{x}_i + \epsilon_i$$

  `fe`: the fixed effects or within estimator.

$$y^{C_i} = \mathbf{X}^{C_i} + \epsilon_{it}$$

  `re`: the standard GLS random effects estimator.
  `mle`: the maximum likelihood random effects estimator.

Random Effects in Estimation

- The between estimator ignores all within variation $(\psi = 0)$.
- OLS is a weighted average of between and within $(\psi = 1)$.
- GLS is an optimally determined compromise given the orthogonality assumption $(0 \geq \psi \geq 1)$.

That weight is not in any sense optimally determined, it is a function of the relative ratio of the two quantities (all variance counts the same). As Hsiao (p. 37) points out that the random effects estimator is often known as a quasi-demeaning estimator. This is because it is a partial within transformation.

Details on Random Effects GLS (FGLS)

We will start with the model we defined as random effects before. We defined random effects $\boldsymbol{\alpha_i} \perp \boldsymbol{X_{it}}$: $\boldsymbol{\alpha_i} \sim [\alpha, \sigma^2_\alpha]$ $\epsilon_{it} \sim [0, \sigma^2_\epsilon]$. Consider $\nu_{it} = \alpha_i + \epsilon_{it}$. For a single cross-section (remembering the Kronecker product will help us here)

$$\mathbb{E}(\nu_{it}\nu'_{it}) = \sigma^2_\epsilon \mathbf{I_T} + \sigma^2_\alpha \mathbf{1}_T = \Omega$$

The inverse is given by

$$\Omega^{-1} = \frac{1}{\sigma^2_\epsilon}\left[\mathbf{I_T} - \frac{\sigma^2_\alpha}{\sigma^2_\epsilon + T\sigma^2_\alpha}\mathbf{1}_T\right]$$

We can also estimate this by using ordinary least squares applied to transformed data. The quasi-demeaning can be done in a first-stage with OLS estimates on the quasi-demeaned data. Recall the pooled regression uses no transformation. The within estimator uses complete demeaning. The random effects estimator is somewhere in between.

Random Effects Variance

Breusch and Pagan (modified by Baltagi and Li) have developed a Lagrange multiplier test of whether or not the random effects have a variance. The test statistic is defined as:

$$LM = \frac{NT}{2(T-1)}\left[\frac{\sum_N \left(\sum_T \epsilon_{it}\right)^2}{\sum_N \sum_T \epsilon_{it}^2} - 1\right] \sim \chi_1^2$$

```
. xtreg growth lagg opengdp openex openimp leftc central inter, re
Random-effects GLS regression                   Number of obs      =       240
Group variable (i): country                     Number of groups   =        16
------------------------------------------------------------------------------
      growth |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lagg1 |    .151848   .0865508     1.75   0.079    -.0177884    .3214843
     opengdp |   .0082889   .0010012     8.28   0.000     .0063267    .0102511
      openex |   .0019834   .0005903     3.36   0.001     .0008263    .0031404
     openimp |  -.0047988   .0010474    -4.58   0.000    -.0068518   -.0027459
       leftc |  -.0268801   .0108211    -2.48   0.013     -.048089   -.0056711
     central |  -.7428119   .2547157    -2.92   0.004    -1.242045   -.2435784
```

```
       inter |   .0138935   .0041671     3.33   0.001     .0057261    .0220609
       _cons |   3.607517    .571187     6.32   0.000     2.488011    4.727023
-------------+----------------------------------------------------------------
     sigma_u |  .36517121
     sigma_e |  2.0094449
         rho |  .03196908   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

R-squareds

```
. xtreg growth lagg1 opengdp, fe

Fixed-effects (within) regression               Number of obs      =       240
Group variable (i): country                     Number of groups   =        16

R-sq:  within  = 0.2562                         Obs per group: min =        15
       between = 0.0031                                        avg =      15.0
       overall = 0.1563                                        max =        15

                                                F(2,222)           =     38.23
corr(u_i, Xb)  = -0.3888                        Prob > F           =    0.0000

------------------------------------------------------------------------------
      growth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lagg1 |   .2647972   .0851979     3.11   0.002     .0968971    .4326972
     opengdp |   .0094949   .0011229     8.46   0.000      .007282    .0117078
       _cons |   .5289261   .3719065     1.42   0.156    -.2039929    1.261845
-------------+----------------------------------------------------------------
     sigma_u |   1.142546
     sigma_e |  2.0889953
         rho |  .23025918   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(15, 222) =     3.55             Prob > F = 0.0000
```

```
. reg Cgrowth Clagg1 Copengdp

      Source |       SS       df       MS              Number of obs =     240
-------------+------------------------------           F(  2,   237) =   40.81
       Model |  333.650655     2  166.825327           Prob > F      =  0.0000
    Residual |  968.786108   237   4.0877051           R-squared     =  0.2562
-------------+------------------------------           Adj R-squared =  0.2499
       Total |  1302.43676   239   5.4495262           Root MSE      =  2.0218

------------------------------------------------------------------------------
     Cgrowth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Clagg1 |   .2647972   .0824577     3.21   0.002     .1023536    .4272408
    Copengdp |   .0094949   .0010868     8.74   0.000     .0073539    .0116359
       _cons |   1.30e-08   .1305071     0.00   1.000    -.2571021    .2571021
------------------------------------------------------------------------------
```

Betweens

```
. by country: egen gmean = mean(growth)
. by country: egen glmean = mean(lagg1)
. by country: egen opengdpmean = mean(opengdp)
. gen yhatb = _b[_cons] + _b[lagg1]*glmean + _b[opengdp]*opengdpmean
. reg gmean yhatb

      Source |       SS       df       MS              Number of obs =     240
-------------+------------------------------           F(  1,   238) =    0.75
       Model |  .445360906     1  .445360906           Prob > F      =  0.3868
    Residual |  140.975583   238  .592334381           R-squared     =  0.0031
-------------+------------------------------           Adj R-squared = -0.0010
       Total |  141.420943   239  .591719429           Root MSE      =  .76963

------------------------------------------------------------------------------
       gmean |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       yhatb |  -.0570801   .0658282    -0.87   0.387    -.1867605    .0726003
       _cons |   3.185291   .2044862    15.58   0.000     2.782457    3.588125
------------------------------------------------------------------------------
```

Total

```
gen yhatT = _b[_cons] + _b[lagg1]*lagg1 + _b[opengdp]*opengdp

. fit growth yhatT

      Source |       SS       df       MS              Number of obs =     240
-------------+------------------------------           F(  1,   238) =   44.11
       Model |  225.744206     1  225.744206           Prob > F      =  0.0000
    Residual |  1218.11349   238  5.11812392           R-squared     =  0.1563
-------------+------------------------------           Adj R-squared =  0.1528
       Total |   1443.8577   239   6.0412456           Root MSE      =  2.2623

------------------------------------------------------------------------------
      growth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       yhatT |   .6927893   .1043154     6.64   0.000       .48729    .8982887
       _cons |   .9257153   .3465985     2.67   0.008     .2429227    1.608508
------------------------------------------------------------------------------
```

Extending this basic logic will hold for all `xtreg` estimators. Basically, think about them as projecting any given model result to the centered data, to group means, and to all data.

Random Coefficients

We saw fixed and random effects. The basic idea generalizes to regression coefficients on variables that are not unit-specific factors/indicators.

- Random Coefficients Specifications (Swamy 1970)

$$\boldsymbol{y_{it}} = \alpha + (\overline{\boldsymbol{\beta}} + \boldsymbol{\mu_i})\boldsymbol{X_{it}} + \epsilon_{it} \quad (11)$$

$$\mathbb{E}[\alpha_i] = 0; \mathbb{E}[\alpha_i \boldsymbol{X}_{it}] = 0 \quad (12)$$

$$\mathbb{E}[\alpha_i \alpha_j] = \begin{cases} \Delta & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (13)$$

Hsiao and Pesaran (2004, IZA DP 136) show that the GLS estimator is a matrix weighted average of the OLS estimator applied to each unit separately with weights inversely proportional to the covariance matrix for the unit.

`xtrc`: Implementing Random Coefficients

`xtrc` estimates the Swamy random coefficients model and provides us with a test statistic of parameter constancy. If the statistic is significantly different from zero, parameter constancy is rejected. Option `betas` gives us the unit-specifics. We have `vce` options here also.

Note, as with many `xt` commands, the jackknife is unit-based.

xtmixed

Stata has a mixed effects module that we can use for some things we have already seen and for extensions. I should say in passing that this also works for dimensions with nesting properties, though we are looking at two-dimensional data structures.

```
. sum

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        year |       240        1977    4.329523       1970       1984
     country |       240         8.5    4.619406          1         16
      growth |       240    3.013292    2.457895       -3.6        9.8
       lagg1 |       240    3.119855    1.652682   -2.40641   6.683519
     opengdp |       240    174.6452    146.2456      -32.1     736.02
-------------+--------------------------------------------------------
      openex |       240    489.7662    420.4374      30.94     2879.2
     openimp |       240    482.8254    267.6722      64.96     1415.2
       leftc |       240    34.79583    39.56008          0        100
     central |       240     2.02421    .9593759   .4054115   3.618419
       inter |       240    91.33376    117.5622          0   361.8419
```

```
.  xtreg growth lagg1 opengdp openimp openex leftc, re

Random-effects GLS regression                   Number of obs      =       240
Group variable (i): country                     Number of groups   =        16

R-sq:  within  = 0.2960                         Obs per group: min =        15
       between = 0.2038                                        avg =      15.0
       overall = 0.2811                                        max =        15

Random effects u_i ~ Gaussian                   Wald chi2(5)       =     92.41
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      growth |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lagg1 |   .2194248   .0875581     2.51   0.012     .0478142    .3910355
     opengdp |   .0077965   .0009824     7.94   0.000      .005871    .0097219
     openimp |  -.0053695   .0009868    -5.44   0.000    -.0073035   -.0034355
      openex |   .0019647   .0006047     3.25   0.001     .0007796    .0031498
       leftc |   .0030365   .0036142     0.84   0.401    -.0040472    .0101202
       _cons |   2.491734   .4633904     5.38   0.000     1.583505    3.399962
-------------+----------------------------------------------------------------
     sigma_u |  .21759529
     sigma_e |  2.0364407
         rho |  .01128821   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

An MLE

```
.  xtreg growth lagg1 opengdp openimp openex leftc, mle
Random-effects ML regression                    Number of obs      =       240
Group variable (i): country                     Number of groups   =        16

Random effects u_i ~ Gaussian                   Obs per group: min =        15

                                                LR chi2(5)         =     81.33
Log likelihood  =  -514.4714                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      growth |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lagg1 |   .1873509   .0881362     2.13   0.034       .014607    .3600947
     opengdp |   .0077706   .0009913     7.84   0.000      .0058276    .0097136
     openimp |  -.0055243   .0010506    -5.26   0.000     -.0075835   -.0034651
      openex |   .0020447   .0005936     3.44   0.001      .0008812    .0032082
       leftc |   .0044378   .0039745     1.12   0.264     -.0033521    .0122277
       _cons |   2.583146   .5204807     4.96   0.000      1.563022    3.603269
-------------+----------------------------------------------------------------
    /sigma_u |   .5100119   .1962033                      .2399497    1.084028
    /sigma_e |   2.018389   .0957214                      1.839233    2.214995
         rho |   .0600166   .0445522                      .0110832    .2056057
------------------------------------------------------------------------------
Likelihood-ratio test of sigma_u=0: chibar2(01)=    3.56 Prob>=chibar2 = 0.030
```

```
. xtmixed growth lagg1 opengdp openimp openex leftc || R.country, mle
Mixed-effects ML regression                     Number of obs      =       240
Group variable: _all                            Number of groups   =         1
                                                Wald chi2(5)       =     97.44
Log likelihood =  -514.4714                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      growth |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lagg1 |   .1873501   .0859494     2.18   0.029     .0188925    .3558078
     opengdp |   .0077706   .0009911     7.84   0.000     .0058281     .009713
     openimp |  -.0055243   .0010452    -5.29   0.000    -.0075729   -.0034757
      openex |   .0020447   .0005915     3.46   0.001     .0008854    .0032039
       leftc |   .0044378   .0038479     1.15   0.249    -.003104     .0119796
       _cons |   2.583148   .5173579     4.99   0.000     1.569145    3.597151
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
_all: Identity               |
               sd(R.country) |   .5100191   .1962046      .2399545    1.084037
-----------------------------+------------------------------------------------
                sd(Residual) |   2.018388   .0957229       1.83923    2.214997
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =     3.56 Prob >= chibar2 = 0.0296
```

General Stata things, `, vce()`

For virtually all Stata commands, we can acquire multiple variance/covariance matrices of the parameters.

`, robust` sometimes

`, cluster()` sometimes

`, vce(boot)`

`, vce(jack)`

`xtmixed`

Will allow us to do tons of things. In particular, we can play with the residual correlation matrix using the option `residuals`. One can recreate virtually everything that we have seen so far this way. The remaining task for you in the lab is to figure out what all you can make it do.

- exchangeable
- ar
- ma
- unstructured
- banded

- toeplitz
- exponential

Mixed Effects Models in Stata with `xtmixed`

Mixed effects models will allow us to estimate many interesting models for `xt` data.

- Simple random effects
- Crossed random effects
- Random Coefficients
- Determined random coefficients

Examples

For the simple random effects estimator, there are two ways to do it via ML.

```
xtreg depvar indvars, mle

xtmixed depvar indvars || _all:  R.UnitID, mle
```

```
. xtreg growth lagg1 opengdp openimp openex leftc, mle
                                                LR chi2(5)          =     81.33
Log likelihood  =  -514.4714                    Prob > chi2         =    0.0000

------------------------------------------------------------------------------
      growth |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lagg1 |   .1873509   .0881362     2.13   0.034      .014607    .3600947
     opengdp |   .0077706   .0009913     7.84   0.000     .0058276    .0097136
     openimp |  -.0055243   .0010506    -5.26   0.000    -.0075835   -.0034651
      openex |   .0020447   .0005936     3.44   0.001     .0008812    .0032082
       leftc |   .0044378   .0039745     1.12   0.264    -.0033521    .0122277
       _cons |   2.583146   .5204807     4.96   0.000     1.563022    3.603269
-------------+----------------------------------------------------------------
    /sigma_u |   .5100119   .1962033                      .2399497    1.084028
    /sigma_e |   2.018389   .0957214                      1.839233    2.214995
         rho |   .0600166   .0445522                      .0110832    .2056057
------------------------------------------------------------------------------
Likelihood-ratio test of sigma_u=0: chibar2(01)=    3.56 Prob>=chibar2 = 0.030

. xtmixed growth lagg1 opengdp openimp openex leftc || _all: R.country, mle
                                                Wald chi2(5)        =     97.44
Log likelihood =  -514.4714                     Prob > chi2         =    0.0000

------------------------------------------------------------------------------
      growth |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
```

```
-------------+----------------------------------------------------------------
       lagg1 |   .1873501   .0859494     2.18   0.029     .0188925    .3558078
     opengdp |   .0077706   .0009911     7.84   0.000     .0058281     .009713
     openimp |  -.0055243   .0010452    -5.29   0.000    -.0075729   -.0034757
      openex |   .0020447   .0005915     3.46   0.001     .0008854    .0032039
       leftc |   .0044378   .0038479     1.15   0.249     -.003104    .0119796
       _cons |   2.583148   .5173579     4.99   0.000     1.569145    3.597151
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
_all: Identity               |
               sd(R.country) |   .5100191   .1962046      .2399545    1.084037
-----------------------------+------------------------------------------------
                sd(Residual) |   2.018388   .0957229       1.83923    2.214997
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =     3.56 Prob >= chibar2 = 0.0296
```

Crossed Random Effects

```
Mixed-effects ML regression                     Number of obs      =       240
Group variable: _all                            Number of groups   =         1

                                                Obs per group: min =       240
                                                               avg =     240.0
                                                               max =       240


                                                Wald chi2(5)       =      7.18
Log likelihood = -503.45468                     Prob > chi2        =    0.2076

------------------------------------------------------------------------------
      growth |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lagg1 |   .0059048   .1296512     0.05   0.964    -.2482069    .2600164
     opengdp |   .0001904   .0016087     0.12   0.906    -.0029626    .0033433
     openimp |  -.0030722   .0015617    -1.97   0.049     -.006133   -.0000114
      openex |    .002307   .0010185     2.27   0.024     .0003108    .0043032
       leftc |   .0048234   .0036133     1.33   0.182    -.0022585    .0119053
       _cons |   3.147245   .7630121     4.12   0.000     1.651768    4.642721
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
```

```
-----------------------------+------------------------------------------------
_all: Identity               |
               sd(R.country) |   .6667379   .1900389      .3813634    1.165658
-----------------------------+------------------------------------------------
_all: Identity               |
                  sd(R.year) |   1.554459   .4033566      .9347738     2.58495
-----------------------------+------------------------------------------------
                sd(Residual) |   1.752177   .0885389      1.586961    1.934595
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(2) =    25.59   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference

. estimates store MLEtwowayRE

. lrtest MLEtwowayRE MLEunitRE

Likelihood-ratio test                                 LR chibar2(01)   =     22.03
(Assumption: MLEunitRE nested in MLEtwowayRE)     Prob > chibar2   =    0.0000

. qui xtmixed growth lagg1 opengdp openimp openex leftc || _all: R.year, mle

. lrtest MLEtwowayRE .

Likelihood-ratio test                                 LR chibar2(01)   =     10.04
(Assumption: . nested in MLEtwowayRE)             Prob > chibar2   =    0.0008
```

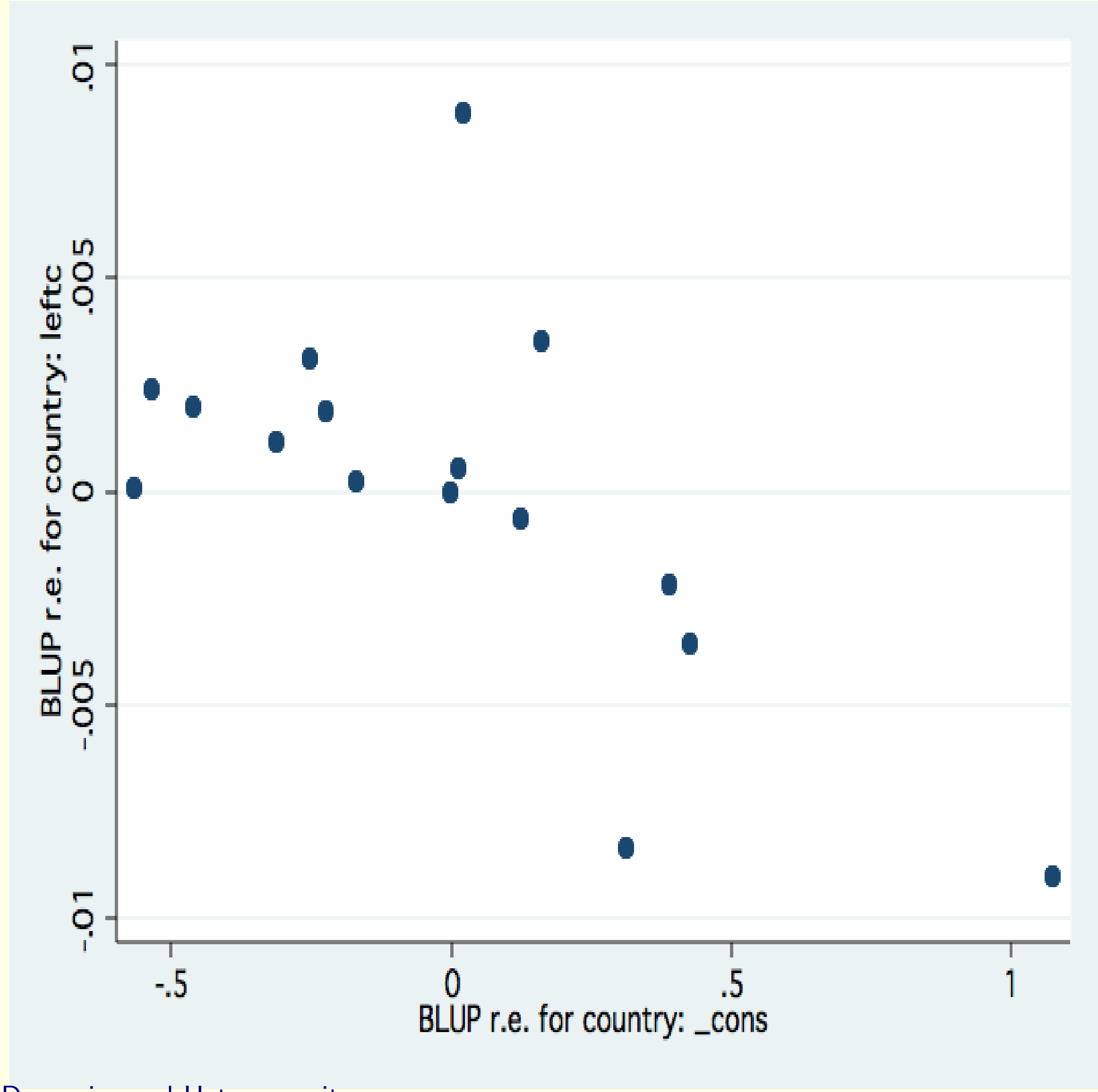
```
. xtmixed growth lagg1 opengdp openimp openex leftc || country: leftc, covariance(unstructured)

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log restricted-likelihood = -540.17955
Iteration 1:   log restricted-likelihood = -540.15493
Iteration 2:   log restricted-likelihood = -540.15472
Iteration 3:   log restricted-likelihood = -540.15472

Computing standard errors:

Mixed-effects REML regression                   Number of obs      =       240
Group variable: country                         Number of groups   =        16

                                                Obs per group: min =        15
                                                               avg =      15.0
                                                               max =        15


                                                Wald chi2(5)       =     95.70
Log restricted-likelihood = -540.15472          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      growth |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
```

```
-------------+----------------------------------------------------------------
       lagg1 |    .170562   .0869219     1.96   0.050     .0001982    .3409259
     opengdp |   .0078608   .0010053     7.82   0.000     .0058905    .0098312
     openimp |  -.0055371   .0010763    -5.14   0.000    -.0076465   -.0034277
      openex |   .0020745   .0005967     3.48   0.001     .0009051    .0032439
       leftc |   .0039332   .0046265     0.85   0.395    -.0051346     .013001
       _cons |   2.570449   .5444497     4.72   0.000     1.503347    3.637551
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
country: Unstructured        |
                  sd(leftc) |   .0089451   .0078813      .0015908    .0502989
                  sd(_cons) |   .6566839   .2658791      .2969756    1.452085
          corr(leftc,_cons) |  -.6168731   .5300418     -.9835763    .7429732
-----------------------------+------------------------------------------------
               sd(Residual) |   2.022226    .098202       1.83863    2.224156
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =     5.40   Prob > chi2 = 0.1445

Note: LR test is conservative and provided only for reference

. * The coefficient is insignificant as is the randomness
. estat recovariance
```

```
Random-effects covariance matrix for level country

             |     leftc      _cons
-------------+----------------------
       leftc |    .00008
       _cons | -.0036236   .4312338

. capture drop u1 u2

. predict u*, reffects
```

```
. by country, sort: sum u*
------------------------------------------------------------------------------------------
-> country = AUL
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          u1 |        15   -.0006591           0  -.0006591  -.0006591
          u2 |        15    .1237475           0   .1237475   .1237475
-> country = AUS
          u1 |        15    .0005591           0   .0005591   .0005591
          u2 |        15    .0125652           0   .0125652   .0125652
-> country = BEL
          u1 |        15   -.0000316           0  -.0000316  -.0000316
          u2 |        15   -.0002924           0  -.0002924  -.0002924
-> country = CAN
          u1 |        15   -.0035756           0  -.0035756  -.0035756
          u2 |        15    .4255248           0   .4255248   .4255248
-> country = DEN
          u1 |        15    .0019625           0   .0019625   .0019625
          u2 |        15    -.462575           0   -.462575   -.462575
-> country = FIN
          u1 |        15     .003543           0    .003543    .003543
          u2 |        15    .1606634           0   .1606634   .1606634
-> country = FRA
          u1 |        15   -.0083416           0  -.0083416  -.0083416
          u2 |        15    .3128709           0   .3128709   .3128709
-> country = GER
```

```
         u1 |        15    .0011514           0   .0011514   .0011514
         u2 |        15   -.3119804           0  -.3119804  -.3119804
-> country = IRE
         u1 |        15   -.0021854           0  -.0021854  -.0021854
         u2 |        15    .3908045           0   .3908045   .3908045
 -> country = ITA
         u1 |        15    .0002358           0   .0002358   .0002358
         u2 |        15   -.1705837           0  -.1705837  -.1705837
-> country = JAP
         u1 |        15   -.0090248           0  -.0090248  -.0090248
         u2 |        15    1.074025           0   1.074025   1.074025
-> country = NET
         u1 |        15    .0031352           0   .0031352   .0031352
         u2 |        15   -.2520462           0  -.2520462  -.2520462
-> country = NOR
         u1 |        15    .0088704           0   .0088704   .0088704
         u2 |        15    .0223926           0   .0223926   .0223926
-> country = SWE
         u1 |        15     .002398           0    .002398    .002398
         u2 |        15   -.5351107           0  -.5351107  -.5351107
-> country = UK
         u1 |        15     .000085           0    .000085    .000085
         u2 |        15   -.5665398           0  -.5665398  -.5665398
-> country = USA
         u1 |        15    .0018777           0   .0018777   .0018777
         u2 |        15   -.2234658           0  -.2234658  -.2234658
```


BLUP r.e. for country: leftc
.01
.005
0
-.005
-.01
-.5
0
.5
1
BLUP r.e. for country: _cons

Wilson and Butler

- Survey of papers using TSCS data and methods(?)

- Vast majority do nothing about space or time.

- Does it matter?

  Table 3
  Table 4

- What do we do? Raise the bar for positive findings and look at multiple models trying to tease out the role of particular assumptions as necessary and/or sufficient for results.

More on xtpcse

Holding on to data

- `preserve`
- `restore`

Testing the Null Hypothesis of No Random Effects

```
. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects:

        growth[country,t] = Xb + u[country] + e[country,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                  growth |   6.041246       2.457895
                       e |   4.147091       2.036441
                       u |   .0473477       .2175953

        Test:   Var(u) = 0
                            chi2(1) =     4.39
```

```
Prob > chi2 =     0.0361
```

xttest

```
. xttest1

Tests for the error component model:

        growth[country,t] = Xb + u[country] + v[country,t]
           v[country,t] = rho v[country,(t-1)] + e[country,t]

        Estimated results:
                                 Var     sd = sqrt(Var)
                ---------+-----------------------------
                  growth |   6.041246       2.457895
                       e |   4.037869      2.0094449
                       u |     .13335      .36517121

        Tests:
           Random Effects, Two Sided:
           LM(Var(u)=0)       =     1.00   Pr>chi2(1) =  0.3174
           ALM(Var(u)=0)      =     0.54   Pr>chi2(1) =  0.4610

           Random Effects, One Sided:
           LM(Var(u)=0)       =     1.00   Pr>N(0,1)  =  0.1587
           ALM(Var(u)=0)      =     0.74   Pr>N(0,1)  =  0.2305

           Serial Correlation:
           LM(rho=0)          =     0.74   Pr>chi2(1) =  0.3906
```

```
          ALM(rho=0)          =    0.28   Pr>chi2(1) =  0.5961

          Joint Test:
          LM(Var(u)=0,rho=0) =    1.28   Pr>chi2(2) =  0.5271

* We cannot reject the null hypothesis of no variation in the random effects.
Also no evidence of serial correlation.
Remember, with the lagged endogenous variable on the right hand side,
the random effects are included if they are there.
```

`xttest1`

1. LM test for random effects, assuming no serial correlation
2. Adjusted LM test for random effects, which works even under serial correlation
3. One-sided version of the LM test for random effects
4. One-sided version of the adjusted LM test for random effects
5. LM joint test for random effects and serial correlation
6. LM test for first-order serial correlation, assuming no random effects
7. Adjusted test for first-order serial correlation, which works even under random effects

`xtgls`

- `corr`: t structure ([ar] or [ps]ar) is ρ common or not.
- `panels`: i structure (iid, [h]eteroscedastic, [c]orrelated (and [h]))
- `rhotype`: regress (regression using lags), dw - Durbin-Watson, freg (forward regression uses leads), nagar, theil, tscorr
- `igls` (iterate or two-step)
- `force` for unbalanced.

`xttest2` and `xttest3`

After `fe` or `xtgls`, we have two tests pre-programmed.

1. We have a test of independence (within) in `xttest2`

2. We have a test of homoscedasticity (within) in `xttest3`

`xtserial`

Wooldridge presents a test for serial correlation.

`xtcsd`

How do we test for cross-sectional dependence?

- Generally used for small T and large N settings.
- Three methods: `xtcsd, pesaran friedman frees`
- This is the panel correction in PCSE

`xtscc`

Driscoll and Kraay (1998) describe a robust covariance matrix estimator for pooled and fixed effects regression models that contain a large time dimension. The approach is robust to heteroscedasticity, autocorrelation, and spatial correlation.

We're Here for Fancy Estimators, Why is Everything OLS?

There are limitation imposed by what people have programmed in terms of regression diagnostics. However, if we can fit the same model by OLS, we can use standard regression diagnostics post-estimation to avoid calculating the diagnostics by hand. Many diagnostics are pre-programmed.

OLS Diagnostics

- We could also use other standard diagnostics in the OLS framework. If you are going to intensively use Stata, books like Statistics with Stata are quite useful.

  `estat ovtest, [rhs]` will give us Ramsey's RESET test. The option gives us RHS variables, otherwise we just use fitted values. The default is a Wald test applied to the regression

$$y_{it} = X_{it}\beta + \hat{y}^2\gamma_1 + \hat{y}^3\gamma_2 + \hat{y}^4\gamma_3 + \epsilon_{it}$$

  and with option `rhs` the powers are applied to the right-hand side variables.

  `predict ... , dfits` and `dfbeta`: We also have the various `dffits` and `dfbeta` statistics for use in diagnosing leverage. The dfit is the studentized residual multiplied by the square root of h_j over $(1 - h_j)$;

basically a scaled measure of the difference between in-sample and out-of-sample predictions. The `dfit` is obtained as a post-regression prediction using predict. Define `dfbeta` as:

$$DFBETA_j = \frac{r_j v_j}{\sqrt{v^2(1-h_j)}}$$

where h is the j^{th} item in $\mathbf{P}$, r_j is the studentized residual, v_j are the residuals from a regression not containing the regressor in question, and v^2 is their sum of squares. Suggested cutoffs are $2\sqrt{\frac{k}{N}}$ for dfit and $\frac{2}{\sqrt{N}}$ for dfbeta. There is also the Cook's distance (`cooksd`) and Welsch distance (`welsch`).

`estat hettest [varlist] [, rhs [normal | iid | fstat] mtest[(spec)]]` gives us a variety of tests for heteroscedasticity. The `rhs` option gives structure from covariates. `mtest` is important because we are doing multiple testing (often).

`estat vif` gives us some collinearity diagnostics. The statistic is essentially $\frac{1}{1-R^2_{(-k)}}$.

`estat imtest [, preserve white]` where the default is Cameron-Trivedi, we can request White's version, and preserve maintains the original data (saves time often). As a general misspecification test, the Information Matrix test is shown by Hall (1987) to decompose into heteroscedasticity, skewness, and kurtosis of residuals and has some suboptimal properties.

Plots

- avplot: added-variable plot
- avplots: all added-variable plots in one image
- cprplot: component-plus-residual plot
- lvr2plot: leverage-versus-squared-residual plot
- rvfplot: residual-versus-fitted plot
- rvpplot: residual-versus-predictor plot

Panel Unit Root Testing in Stata

As of Stata 11, a battery of panel unit-root tests have emerged. There are many and they operate under differing sets of assumptions.

- Levin-Lin-Chu (`xtunitroot llc`): trend nocons (unit specific) demean (within transform) lags. Under (crucial) cross-sectional independence, the test is an advancement on the generic Dickey-Fuller theory that allows the lag lengths to vary by cross-sections. The test relies on specifying a kernel (beyond our purposes) and a lag length (upper bound). The test statistic has a standard normal basis with asymptotics in $\frac{\sqrt{N_T}}{T}$ (T grows faster than N). The test is of either all series containing unit roots (H_0) or all stationary; this is a limitation. It is recommended for moderate to large T and N.

1. Perform separate ADF regressions:

$$\Delta y_{it} = \rho_i \Delta y_{i,t-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{i,t=L} + \alpha_{mi} d_{mt} + \epsilon_{it}$$

with d_{mt} as the vector of deterministic variables (none, drift, drift and trend). Select a max L and use t on $\hat{\theta}_{iL}$ to attempt to simplify. Then use $\Delta y_{it} = \Delta y_{i,t-L}$ and d_{mt} for residuals

- Harris-Tzavalis (`xtunitroot ht`): trend nocons (unit specific) demean (within transform) altt (small sample adjust) Similar to the previous, they show that $T \to \infty$ faster than N (rather than T fixed) leads to size distortions.

- Breitung (`xtunitroot breitung`): trend nocons (unit specific) demean (within transform) robust (CSD) lags.
  Similar to LLC with a common statistic across all i.

- Im, Pesaran, Shin (`xtunitroot ips`): trend demean (within transform) lags. They free ρ to be ρ_i and average individual unit root statistics. The null is that all contain unit roots while the alternative specifies at least some to be stationary. The test relies on sequential asymptotics (first T, then N). Better in small samples than LLC, but note the differences in the alternatives.

- Fisher type tests (`xtunitroot fisher`): dfuller pperron demean lags.

- Hadri (LM) (`xtunitroot hadri`): trend demean robust

All but the last are null hypothesis unit-root tests. Most assume balance but the fisher and IPS versions can work for unbalanced panels.

ADL/Canonical models

We can consider some very basic time series models.

- Koyck/Geometric decay:
  short run and long-run effects are parametrically identified (given $\mathcal{M}$).

- Almon (more arbitrary decay):

$$y_{it} = \sum_{t_A=0}^{T_F} \rho_{t_A} x_{t-t_A} + \epsilon_t$$

  with coefficients that are ordinates of some general polynomial of degree $T_F >> q$. The $\rho_{t_A} = \sum_{k=0}^{T_F} \gamma_k t^k$.

- Prais-Winston, etc. are basically FGLS implementations of AR(1).

Prais-Winsten/Cochrane-Orcutt

$$y_{it} = X_{it}\beta + \epsilon_{it}$$

where

$$\epsilon_{it} = \rho\epsilon_{i,t-1} + \nu_{it}$$

and $\nu_{it} \sim N(0, \sigma_\nu^2)$ with stationarity forcing $|\rho| < 1$. We will use iterated FGLS. First, estimate the regression recalling our unbiasedness condition. Then regress $\hat{\epsilon}_{it}$ on $\hat{\epsilon}_{i,t-1}$. Rinse and repeat until ρ doesn't change. The transformation applied to the first observation is distinct, you can look this up.... In general, the transformed regression is:

$$y_{it} - \rho y_{i,t-1} = \alpha(1-\rho) + \beta(X_{it} - \rho X_{i,t-1}) + \nu_{it}$$

with ν white noise.

Beck

- Static model: Instantaneous impact.

$$y_{i,t} = X_{i,t}\beta + \nu_{i,t}$$

- Finite distributed lag: lags of x finite horizon impact (defined by lags).

$$y_{i,t} = X_{i,t}\beta + \sum_{k=1}^{K} X_{i,t-k}\beta_k + \nu_{i,t}$$

- AR(1): Errors decay geometrically, X instantaneous. (Suppose unmeasured x and think this through).

$$y_{i,t} = X_{i,t}\beta + \nu_{i,t} + \theta\epsilon_{i,t-1}$$

- Lagged dependent variable: lags of y [common geometric decay]

$$y_{i,t} = X_{i,t}\beta + \phi y_{i,t-1} + \nu_{i,t}$$

- ADL: current and lagged x and lagged y.

$$y_{i,t} = X_{i,t}\beta + X_{i,t-1}\gamma + \phi y_{i,t-1} + \epsilon_{i,t}$$

- Panel versions of transfer function models from Box and Jenkins time series. (each x has an impact and decay function)

Brief Comment on Hurwicz/Nickell Bias

- Bias is of stochastic order $\frac{1}{T}$.
- Less bad as more T

Interpretation of dynamic models

- Do it.
- Whitten and Williams `dynsim` uses `Clarify`[3] to do this.
- Their paper is "But Wait, There's More! Maximizing Substantive Inferences from TSCS Models". Easy to find on the web and on the website.

[3] If you do not know what Clarify is, please ask: estimate, set, simulate.

Details

$$y_{it} = \alpha + \gamma y_{i,t-1} + X_{it}\beta + \epsilon_{it} \quad (14)$$

$$y_{it} = \alpha + \gamma[\alpha + \gamma y_{i,t-2} + X_{i,t-1}\beta + \epsilon_{i,t-1}] + X_{it}\beta + \epsilon_{it} \quad (15)$$

$$y_{it} = \alpha + \gamma[\alpha + \gamma(\alpha + \gamma y_{i,t-3} + X_{i,t-2}\beta + \epsilon_{it}) + X_{i,t-1}\beta + \epsilon_{i,t-1}] + X_{it}\beta + \epsilon_{it} \quad (16)$$

We can continue substituting through to conclude that we have a geometrically decaying impact so that the long-run effect of a one-unit change in X is

$$\frac{\beta}{1-\gamma}$$

But γ has uncertainty, it is an estmate. To show the realistic long-run impact, we need to incorporate that uncertainty.