Panel Data Methods
2: Short Panels - Extensions

© 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

- Go beyond basic fixed effects and random effects.
- Panel GMM estimation
  - IV estimator
  - Hausman Taylor estimator
  - Arellano-Bond estimator for dynamic models.
- Panel Random Slopes models
  - Mixed Linear Models
- Panel estimation commands can be used for clustered data
  - e.g. Village level data
- Presentation follows Cameron and Trivedi (2010) chapter 9.

Outline

1. Introduction
2. Panel GMM Theory
3. Linear Panel IV Estimation
4. Hausman-Taylor Estimator
5. Dynamic Models - Arellano Bond
6. Dynamic Models - Panel VAR
7. Mixed Linear Models
8. Clustered Data
9. Summary

2. Panel GMM Estimator

- Short panel: Independent over i, T fixed, $N \rightarrow \infty$.
- Linear panel model
$$y_{it} = \mathbf{x}_{it}'\beta + u_{it}.$$
- Stack all T observations for the i^{th} individual,
$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{u}_i,$$
where $\mathbf{y}_i$ and $\mathbf{u}_i$ are $T \times 1$ vectors and $\mathbf{X}_i$ is $T \times K$ with t^{th} row $\mathbf{x}_{it}'$, so
$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix}; \quad \mathbf{X}_i = \begin{bmatrix} \mathbf{x}_{i1}' \\ \vdots \\ \mathbf{x}_{iT}' \end{bmatrix}; \quad \mathbf{u}_i = \begin{bmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{bmatrix}.$$
- The model $\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{u}_i$ is a linear system of equations, so systems IV estimation.

Panel GMM Estimator (continued)

- Assume existence of a $T \times r$ matrix of instruments $\mathbf{Z}_i$ that satisfy

$$\mathrm{E}[\mathbf{Z}_i'\mathbf{u}_i] = \mathbf{0}.$$

  - $r \geq K$ is the number of instruments.
- GMM estimator minimizes the associated quadratic form

$$Q_N(\boldsymbol{\beta}) = \left[\sum_{i=1}^N \mathbf{Z}_i'\mathbf{u}_i\right]' \mathbf{W}_N \left[\sum_{i=1}^N \mathbf{Z}_i'\mathbf{u}_i\right],$$

  - where $\mathbf{W}_N$ denotes an $r \times r$ weighting matrix.
- Given $\mathbf{u}_i = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}$, some algebra yields the panel GMM estimator

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{\text{PGMM}} &= \left[\left(\sum_{i=1}^N \mathbf{X}_i'\mathbf{Z}_i\right)\mathbf{W}_N\left(\sum_{i=1}^N \mathbf{Z}_i'\mathbf{X}_i\right)\right]^{-1} \\ &\quad \times \left(\sum_{i=1}^N \mathbf{X}_i'\mathbf{Z}_i\right)\mathbf{W}_N\left(\sum_{i=1}^N \mathbf{Z}_i'\mathbf{y}_i\right).\end{aligned}$$

  - The essential condition for consistency of this estimator is $\mathrm{E}[\mathbf{Z}_i'\mathbf{u}_i] = \mathbf{0}$.

Instruments

- Usually need instrument $\mathbf{z}$ to be some variable other than $\mathbf{x}$.
- Here, however, if only current period $\mathbf{x}_{it}$ belong in the model for y_{it} then $\mathbf{x}_{it}'s$ from other periods may become available as instruments.
- Contemporaneous exogeneity assumption: it is only current $\mathbf{z}$ and ε that are uncorrelated

$$\mathrm{E}\left[\mathbf{z}_{it}u_{it}\right] = \mathbf{0}, \quad t = 1, ..., T,$$

  - current $\mathbf{z}_{it}$ only are available as instruments

- Weak exogeneity assumption(predetermined instruments) assumption: lagged values of $\mathbf{z}$ are uncorrelated with the current period error

$$\mathrm{E}\left[\mathbf{z}_{is}u_{it}\right] = \mathbf{0}, \quad s \leq t, \quad t = 1, ..., T.$$

  - $\mathbf{z}_{i1}, ..., \mathbf{z}_{it}$ can be instruments for u_{it}, though future values of $\mathbf{z}_{is}$ cannot be so used.

Instruments (continued)

- Strict exogeneity assumption future values of instruments are also uncorrelated with the current period error

$$\mathrm{E}\left[\mathbf{z}_{is} u_{it}\right] = \mathbf{0}, \quad s, t = 1, ..., T.$$

  - Current, past and future values of $\mathbf{z}_{is}$ are valid instruments for u_{it}
- Strict exogeneity is assumption made to date
  - $\mathrm{E}[u_{it}|\mathbf{x}_{i1}, ..., \mathbf{x}_{iT}] = 0$ implies $\mathrm{E}[u_{it}|\mathbf{x}_{is}] = 0$, $1 \leq s \leq T$, and hence $\mathrm{E}[\mathbf{x}_{is} u_{it}] = \mathbf{0}$.
  - appropriate for static models
  - for dynamic models at most weak exogeneity of instruments can be assumed.
- So many instruments may be available (over-identified)
  - though time-invariant regressors can only be used as an instrument once.

GMM Estimator Distribution

- Rewrite
$$\widehat{\beta}_{\text{PGMM}} = [\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{y},$$
  - where $\mathbf{X}' = [\mathbf{X}_1' \cdots \mathbf{X}_N']$, $\mathbf{Z}' = [\mathbf{Z}_1' \cdots \mathbf{Z}_N']$, $\mathbf{y}' = [\mathbf{y}_1' \cdots \mathbf{y}_N']$.
- Then $\widehat{\beta}_{\text{PGMM}}$ is asymptotically normal with
$$\widehat{\text{V}}[\widehat{\beta}_{\text{PGMM}}] = [\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_N(N\widehat{\mathbf{S}})\mathbf{W}_N'\mathbf{Z}'\mathbf{X}[\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1},$$
- Given independence over i , $\widehat{\mathbf{S}}$ is consistent estimate of the $r \times r$ matrix
$$\mathbf{S} = \text{plim}\,\frac{1}{N}\sum\nolimits_{i=1}^{N} \mathbf{Z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}.$$
- A White-type panel robust estimate of $\mathbf{S}$ is
$$\widehat{\mathbf{S}} = \frac{1}{N}\sum\nolimits_{i=1}^{N} \mathbf{Z}_i'\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{Z}_i,$$
  - where the $T \times 1$ estimated residual $\widehat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\widehat{\beta}$
  - controls for arbitrary heteroskedasticity and correlation
  - can panel jackknife or panel bootstrap to get this.

One-step and Two-step GMM

- One-step GMM or two-stage least squares estimator
  - $\mathbf{W}_N = [\sum_i \mathbf{Z}_i'\mathbf{Z}_i]^{-1} = [\mathbf{Z}'\mathbf{Z}]^{-1}$
  - $\widehat{\beta}_{2SLS} = [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$.
  - Optimal PGMM estimator based on $\mathrm{E}[\mathbf{Z}_i'\mathbf{u}_i] = \mathbf{0}$ if $\mathbf{u}_i|\mathbf{Z}_i$ is iid $[\mathbf{0}, \sigma^2\mathbf{I}_T]$.
- Two-step GMM estimator
  - $\mathbf{W}_N = \widehat{\mathbf{S}}^{-1}$ where $\widehat{\mathbf{S}}$ is consistent for $\mathbf{S}$
  - $\widehat{\beta}_{2SGMM} = [\mathbf{X}'\mathbf{Z}\widehat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}\widehat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{y}$
  - Optimal PGMM estimator based on $\mathrm{E}[\mathbf{Z}_i'\mathbf{u}_i] = \mathbf{0}$ given independence over i
  - $\widehat{\mathbf{S}}$ Comes from residuals after one-step 2SLS estimation.

Overidentifying Restrictions Test

- r instruments $>$ K parameters
  $\Rightarrow (r - K)$ overidentifying (or extra) restrictions
  - test whether $\sum_{i=1}^{N} \mathbf{Z}_i' \widehat{\mathbf{u}}_i$ is close to 0.
- After two-step GMM (not one-step) test of overidentifying restrictions (Sargan or Hansen test) is

$$\tau_N = \left[\sum_{i=1}^{N} \widehat{\mathbf{u}}_i' \mathbf{Z}_i\right] (N\widehat{\mathbf{S}})^{-1} \left[\sum_{i=1}^{N} \mathbf{Z}_i' \widehat{\mathbf{u}}_i\right]$$

  - where $\widehat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{Z}_i' \widehat{\beta}_{2SGMM}$
- $\tau_N \sim \chi^2(r - K)$ under H_0 : over-identifying restrictions are valid.
- If τ_N is large overidentifying moment conditions are rejected
  - conclude some of the instruments in $\mathbf{Z}_i$ are correlated with the error.
- If τ_N is small overidentifying moment conditions are not rejected
  - conclude that, conditional on K linear combinations of the instruments being valid, cannot reject that the remaining $(r - K)$ are not also valid.

3. Panel IV Estimation

- Consider model with possibly transformed variables:

$$y_{it}^* = \alpha + \mathbf{x}_{it}^{*\prime}\boldsymbol{\beta} + u_{it}^*,$$

- Transformations are

OLS	$y_{it}^* = y_{it}$
Between	$y_{it}^* = \bar{y}_i$
Fixed effects	$y_{it}^* = (y_{it} - \bar{y}_i)$
Random effects	$y_{it}^* = (y_{it} - \theta_i \bar{y}_i)$

- OLS is **inconsistent** if $\mathrm{E}[u_{it}^* | \mathbf{x}_{it}^*] \neq 0$.
- **IV estimation** with **instruments** $\mathbf{z}_{it}^*$ (also transformed) that satisfy $\mathrm{E}[u_{it}^* | \mathbf{z}_{it}^*] = 0$.
- Stata command `xtivreg` does this
  - natural extension of `ivregress` to panels
  - though only does 2SLS and not GMM.

IV Fixed Effects Example

- Regress lnwage on experience, experience^2, and weeks worked.
- Marital status instrument for weeks worked.
- Output on next slide:
  - Standard errors are more than 10 times larger than for regular FE
  - not unusual for IV
  - coefficients change sign though statistically insignificant.
- Standard errors assume iid errors
  - no `vce(robust)` option
  - instead use `vce(bootstrap)` or `vce(jackknife)`.

```
. * Panel IV example: FE with wks instrumented by external instrument ms
. use mus08psidextract.dta, clear
(PSID wage data 1976-82 from Baltagi and Khanti-Akom (1990))

. xtivreg lwage exp exp2 (wks = ms), fe

Fixed-effects (within) IV regression            Number of obs      =      4165
Group variable: id                              Number of groups   =       595

R-sq:  within  =      .                         Obs per group: min =         7
       between = 0.0172                                        avg =       7.0
       overall = 0.0284                                        max =         7

                                                Wald chi2(3)       =  700142.43
corr(u_i, Xb)  = -0.8499                        Prob > chi2        =     0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         wks |  -.1149742   .2316926    -0.50   0.620    -.5690832    .3391349
         exp |   .1408101   .0547014     2.57   0.010     .0335974    .2480228
        exp2 |  -.0011207   .0014052    -0.80   0.425    -.0038748    .0016334
       _cons |    9.83932   10.48955     0.94   0.348    -10.71983    30.39847
-------------+----------------------------------------------------------------
     sigma_u |  1.0980369
     sigma_e |  .51515503
         rho |  .81959748   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F  test that all u_i=0:     F(594,3567) =     4.62         Prob > F    = 0.0000
------------------------------------------------------------------------------
Instrumented:   wks
Instruments:    exp exp2 ms
```

4. Hausman-Taylor IV Estimator

- We want to estimate the coefficient of a time-invariant regressor.
- Problem: the fixed effect estimator is not identified for the-invariant regressor
- Solution: assume that some time-varying regressors are not correlated with either α_i or ε_{it}
  - then their values in periods other than the current period can be used as instruments
  - called the Hausman-Taylor estimator
- Stata command `xthtaylor` implements this
  - for panel-robust se's use `vce(bootstrap)` or `vce(jackknife)`.

Hausman-Taylor IV Estimator (continued)

- Hausman and Taylor (1981) model

$$y_{it} = \mathbf{x}'_{1it}\boldsymbol{\beta}_1 + \mathbf{x}'_{2it}\boldsymbol{\beta}_2 + \mathbf{w}'_{1i}\gamma_1 + \mathbf{w}'_{2i}\gamma_2 + \alpha_i + \varepsilon_{it},$$

  - $\mathbf{w}$ is introduced to denote time-invariant regressors
- Assume
  - $\mathbf{x}_{1it}$ and $\mathbf{w}_{1i}$ are uncorrelated with α_i
  - $\mathbf{x}_{2it}$ and $\mathbf{w}_{2i}$ are correlated with α_i
  - all regressors are assumed to be uncorrelated with ε_{it}.
- Then $\mathbf{x}_{1is}$ for $s \neq t$ available as instruments.
  - can test validity of overidentifying restrictions using OIR test.

Hausman-Taylor IV Estimator (continued)

- Actually assume a RE error model and do more efficient IV
  - first transform model as in FGLS

$$\widetilde{y}_{it} = \widetilde{\mathbf{x}}'_{1it}\boldsymbol{\beta}_1 + \widetilde{\mathbf{x}}'_{2it}\boldsymbol{\beta}_2 + \widetilde{\mathbf{w}}'_{1i}\gamma_1 + \widetilde{\mathbf{w}}'_{2i}\gamma_2 + v_{it}$$

  - where $\widetilde{\mathbf{x}}_{1it} = \widetilde{\mathbf{x}}_{1it} - \widehat{\lambda}\bar{\mathbf{x}}_{1i}$ and $\lambda = 1 - \sigma_\varepsilon / \sqrt{\sigma_\varepsilon^2 + T\sigma_\alpha^2}$.
- Then the instruments are $(\mathbf{x}_{1it} - \bar{\mathbf{x}}_{1i})$, $(\mathbf{x}_{2it} - \bar{\mathbf{x}}_{2i})$, $\mathbf{w}_{1i}$ and $\bar{\mathbf{x}}_{1i}$.
- Amemiya and MaCurdy (1986) propose additional instruments
  - option `amacurdy`
- Breusch et al (1989) propose even more.

Hausman-Taylor Example

- Full model: we used subset of regressors so far
  - want to instruments for time-invariant endogenous regressor ed
  - assume time-varying regressors occ, south, smsa and ind (industry) are exogenous

```
. * Hausman-Taylor example of Baltagi and Khanti-Akom (1990)
. use mus08psidextract.dta, clear
(PSID wage data 1976-82 from Baltagi and Khanti-Akom (1990))

. xthtaylor lwage occ south smsa ind exp exp2 wks ms union fem blk ed,  ///
>   endog(exp exp2 wks ms union ed)

Hausman-Taylor estimation                        Number of obs      =      4165
Group variable: id                               Number of groups   =       595

                                                 Obs per group: min =         7
                                                                avg =         7
                                                                max =         7

Random effects u_i ~ i.i.d.                      Wald chi2(12)      =   6891.87
                                                 Prob > chi2        =    0.0000
```

- Coefficient of `ed` now 0.138 (was 0.112 with `xtreg, re`)
  - se of `ed` increases to 0.084 from 0.0212.

lwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
TVexogenous						
occ	-.0207047	.0137809	-1.50	0.133	-.0477149	.0063055
south	.0074398	.031955	0.23	0.816	-.0551908	.0700705
smsa	-.0418334	.0189581	-2.21	0.027	-.0789906	-.0046761
ind	.0136039	.0152374	0.89	0.372	-.0162608	.0434686
TVendogenous						
exp	.1131328	.002471	45.79	0.000	.1082898	.1179758
exp2	-.0004189	.0000546	-7.67	0.000	-.0005259	-.0003119
wks	.0008374	.0005997	1.40	0.163	-.0003381	.0020129
ms	-.0298508	.01898	-1.57	0.116	-.0670508	.0073493
union	.0327714	.0149084	2.20	0.028	.0035514	.0619914
TIexogenous						
fem	-.1309236	.126659	-1.03	0.301	-.3791707	.1173234
blk	-.2857479	.1557019	-1.84	0.066	-.5909179	.0194221
TIendogenous						
ed	.137944	.0212485	6.49	0.000	.0962977	.1795902
_cons	2.912726	.2836522	10.27	0.000	2.356778	3.468674
sigma_u	.94180304					
sigma_e	.15180273					
rho	.97467788	(fraction of variance due to u_i)				

Note: TV refers to time varying; TI refers to time invariant.

5. Linear Dynamic Panel Models

- Simple dynamic model regresses y_{it} in **polynomial in time**.
  - e.g. Growth curve of child height or IQ as grow older
  - use previous models with $\mathbf{x}_{it}$ polynomial in time or age.
- Richer dynamic model regresses y_{it} on **lags** of y_{it}.

Linear Dynamic Panel Models with Individual Effects

- **Leading example**: AR(1) model with individual specific effects

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}.$$

- Four reasons for y_{it} being serially correlated over time:
  - **Unobserved heterogeneity:** via α_i (permanent)
  - **True state dependence:** via $y_{i,t-1}$ (dampens over time)
  - **Observed heterogeneity:** via $\mathbf{x}_{it}$ which may be serially correlated
  - **Error correlation:** via ε_{it}
- There is a literature on dynamic models with random effects.
- We focus on case where α_i is a **fixed effect**
  - and T small so cannot consistently estimate α_i
  - no problem if $T \rightarrow \infty$ and ε_{it} serially uncorrelated.

Fixed Effects Estimator is Inconsistent

- Mean difference yields

$$(y_{it} - \bar{y}_i) = \gamma(y_{i,t-1} - \bar{y}_{i,-1}) + (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i).$$

- Problem: regressor $(y_{i,t-1} - \bar{y}_{i,-1})$ correlated with $(\varepsilon_{it} - \bar{\varepsilon}_i)$
  - since y_{it} (part of $\bar{y}_{i,-1}$) is correlated with ε_{it}
  - and $y_{i,t-1}$ is correlated with ε_{it} (part of $\bar{\varepsilon}_i$).
- This inconsistency is called Nickell bias
  - inconsistency is $O(T^{-1})$.

Arellano-Bond Estimator

- **First-difference** to eliminate α_i (rather than mean-difference)

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

- **OLS inconsistent** as $(y_{i,t-1} - y_{i,t-2})$ correlated with $(\varepsilon_{it} - \varepsilon_{i,t-1})$ (even under assumption ε_{it} is serially uncorrelated).
- But $y_{i,t-2}$ is uncorrelated with $(\varepsilon_{it} - \varepsilon_{i,t-1})$, **if ε_{it} serially uncorrelated**
  so can use $y_{i,t-2}$ as an **instrument** for $(y_{i,t-1} - y_{i,t-2})$.
- Arellano-Bond is a variation that uses **unbalanced set** of instruments with further lags as instruments.
  For $t = 3$ can use y_{i1}, for $t = 4$ can use y_{i1} and y_{i2}, and so on.
- Stata commands
  - `xtabond` for Arellano-Bond
  - `xtdpdsys` for Blundell-Bond (more efficient than `xtabond`)
  - `xtdpd` for more complicated models than `xtabond` and `xtdpdsys`.

Example Stata

```
.  * Optimal or two-step GMM for a dynamic panel model
.  xtabond lwage occ south smsa ind, lags(2) maxldep(3) ///
    pre(wks,lag(1,2)) endogenous(ms,lag(0,2)) ///
.  endogenous(union,lag(0,2)) twostep vce(robust) ///
.  artests(3)
.  * Test whether error is serially correlated
.  estat abond
.  * Test of overidentifying restrictions
.  estat sargan
.  * Arellano/Bover or Blundell/Bond for dynamic panel
.  xtdpdsys lwage occ south smsa ind, lags(2) ///
    maxldep(3) pre(wks,lag(1,2)) endog(ms,lag(0,2)) ///
    endogenous(union,lag(0,2)) twostep vce(robust) ///
.  artests(3)
```

Data Example: Pure Time Series AR(2)

- Pure AR(2): $y_{it} = \gamma_1 y_{i,t-1} + \gamma_2 y_{i,t-2} + \Delta\varepsilon_{it}$, $t = 1, 2, ..., 7$.
- Sufficient data to obtain IV estimates in the differenced model

$$\Delta y_{it} = \gamma_1 \Delta y_{i,t-1} + \gamma_2 \Delta y_{i,t-2} + \Delta\varepsilon_{it}, \quad t = 4, 5, 6, 7.$$

- Instruments
  - At $t = 4$: 2 available instruments y_{i1} and y_{i2} as uncorrelated with $\Delta\varepsilon_{i4}$
  - At $t = 5$: 3 instruments, y_{i1}, y_{i2} and $y_{i3} \perp \Delta\varepsilon_{i5}$
  - At $t = 6$: 4 instruments $y_{i1}, ..., y_{i4}$
  - At $t = 7$: 5 instruments $y_{i1}, ..., y_{i5}$.
- In all $2 + 3 + 4 + 5 = 14$ instruments for $\Delta y_{i,t-1}$ and $\Delta y_{i,t-2}$
  - intercept is an instrument for itself.
- Estimation may be by 2SLS or by more efficient optimal GMM
  - instrument set is unbalanced: much easier to use command `xtabond` than it is to manually set up the instruments and use `ivregress`.

- One-step GMM

```
. xtabond lwage, lags(2) vce(robust)

Arellano-Bond dynamic panel-data estimation  Number of obs      =      2380
Group variable: id                           Number of groups   =       595
Time variable: t
                                             Obs per group:    min =         4
                                                               avg =         4
                                                               max =         4

Number of instruments =     15               Wald chi2(2)       =   1253.03
                                             Prob > chi2        =    0.0000
One-step results
                                     (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lwage |
         L1. |   .5707517   .0333941    17.09   0.000     .5053005    .6362029
         L2. |   .2675649   .0242641    11.03   0.000     .2200082    .3151216
             |
       _cons |   1.203588    .164496     7.32   0.000     .8811814    1.525994
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/.).lwage
Instruments for level equation
        Standard: _cons
```

Data Example: Pure Time Series (continued)

- There are $4 \times 595 = 2380$ observations as lost $t = 1, 2, 3$.
- Results are reported in terms of the original levels model
  - even though mechanically the FD model is fitted.
- There are 15 instruments
  - output `L(2/.)` means that $y_{i,t-2}$, $y_{i,t-3}$, ..., $y_{i,1}$ are the instruments used for period t.
- Wages depend greatly on past wages, with the lag weights summing to $.57 + .27 = .84$.
- Why is there a constant term?
  - The estimated model actually has no constant term.
  - Instead the estimated constant coefficient is recovered using an auxiliary moment condition (Blundell-Bond levels - considered later).
  - `xtabond` gives same $\widehat{\gamma}$ and $\widehat{\boldsymbol{\beta}}$ with or without `noconstant` option
  - advantage is want a constant if predict from levels equation.

- Two-step GMM - little change or efficiency gain

```
. * Optimal or two-step GMM for a pure time-series AR(2) panel model
. xtabond lwage, lags(2) twostep vce(robust)

Arellano-Bond dynamic panel-data estimation  Number of obs      =      2380
Group variable: id                           Number of groups   =       595
Time variable: t
                                             Obs per group:  min =         4
                                                             avg =         4
                                                             max =         4

Number of instruments =     15               Wald chi2(2)       =   1974.40
                                             Prob > chi2        =    0.0000
Two-step results
                                    (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |              WC-Robust
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lwage |
         L1. |   .6095931   .0330542    18.44   0.000      .544808    .6743782
         L2. |   .2708335   .0279226     9.70   0.000     .2161061    .3255608
             |
       _cons |   .9182262   .1339978     6.85   0.000     .6555952    1.180857
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/.).lwage
Instruments for level equation
        Standard: _cons
```

- Can reduce the number of lags used as instruments
  - too many instruments gives "weak" instrument problem where asymptotic theory works poorly in finite samples
  - here to only 1 lag leads to big efficiency loss (to 2 lags does not)

```
. * Reduce the number of instruments for a pure time-series AR(2) panel model
. xtabond lwage, lags(2) vce(robust) maxldep(1)

Arellano-Bond dynamic panel-data estimation  Number of obs      =      2380
Group variable: id                           Number of groups   =       595
Time variable: t
                                             Obs per group:    min =      4
                                                               avg =      4
                                                               max =      4

Number of instruments =      5               Wald chi2(2)       =   1372.33
                                             Prob > chi2        =    0.0000
One-step results
                                     (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lwage |
         L1. |   .4863642   .1919353     2.53   0.011      .110178    .8625505
         L2. |   .3647456   .1661008     2.20   0.028      .039194    .6902973
             |
       _cons |   1.127609   .2429357     4.64   0.000     .6514633    1.603754
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/2).lwage
Instruments for level equation
        Standard: _cons
```

Data Example: Regressors

- More generally

$$\begin{aligned} y_{it} &= \alpha_i + \sum_{j=1}^{J} \gamma y_{i,t-j} + \sum_{k=0}^{K} \mathbf{x}_{i,t-k}{}'\boldsymbol{\beta}_k + \varepsilon_{it} \\ \Delta y_{it} &= \sum_{j=1}^{J} \gamma_j \Delta y_{i,t-j} + \sum_{k=0}^{K} \Delta\mathbf{x}_{i,t-k}{}'\boldsymbol{\beta}_k + \Delta\varepsilon_{it}. \end{aligned}$$

- Regressors
  - Drop time-invariant regressors: they are eliminated after first-differencing
  - Lagged dependent: lag 2+ as instruments
    - `lwage` appears as first two lags, as before
  - Strictly exogenous $\mathrm{E}[x_{it}\varepsilon_{is}] = 0$ for all s, t : instruments for self
    - `occ`, `south`, `smsa`, and `ind`
  - Predetermined $\mathrm{E}[x_{it}\varepsilon_{is}] \neq 0$ for $s < t$: lag 1+ as instruments
    - `wks` appears current and one lag
  - Endogenous $\mathrm{E}[x_{it}\varepsilon_{it}] \neq 0$: lag 2+ as instruments
    - `ms` and `union`

Data Example: Regressors (continued)

- Stata coding:
  - strictly exogenous variables appear as regular regressors
  - option `maxldep(3)` means at most three lags of y are used as instruments
  - option `pre(wks,lag(1,2))` means `wks` and `L1.wks` are predetermined and only two additional lags are to be used as instruments.
  - option `endogenous(ms,lag(0,2))` means endogenous `ms` appears only as a contemporaneous regressor and that at most two additional lags are used as instruments
  - option `artests(3)` is needed for later `estat abond`

```
. * Optimal or two-step GMM for a dynamic panel model
. xtabond lwage occ south smsa ind, lags(2) maxldep(3)     ///
>   pre(wks,lag(1,2)) endogenous(ms,lag(0,2))              ///
>   endogenous(union,lag(0,2)) twostep vce(robust) artests(3)

Arellano-Bond dynamic panel-data estimation  Number of obs         =      2380
Group variable: id                           Number of groups      =       595
Time variable: t

                                             Obs per group:    min =         4
                                                               avg =         4
                                                               max =         4
```

- Regressors $\mathbf{x}_{it}$ less statistically significant compared to static model.

```
Number of instruments =      40               Wald chi2(10)        =   1287.77
                                              Prob > chi2          =    0.0000
Two-step results
                                      (Std. Err. adjusted for clustering on id)
```

lwage	Coef.	WC-Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
lwage						
L1.	.611753	.0373491	16.38	0.000	.5385501	.6849559
L2.	.2409058	.0319939	7.53	0.000	.1781989	.3036127
wks						
--.	-.0159751	.0082523	-1.94	0.053	-.0321493	.000199
L1.	.0039944	.0027425	1.46	0.145	-.0013807	.0093695
ms	.1859324	.144458	1.29	0.198	-.0972	.4690649
union	-.1531329	.1677842	-0.91	0.361	-.4819839	.1757181
occ	-.0357509	.0347705	-1.03	0.304	-.1038999	.032398
south	-.0250368	.2150806	-0.12	0.907	-.446587	.3965134
smsa	-.0848223	.0525243	-1.61	0.106	-.187768	.0181235
ind	.0227008	.0424207	0.54	0.593	-.0604422	.1058437
_cons	1.639999	.4981019	3.29	0.001	.6637377	2.616261

```
Instruments for differenced equation
        GMM-type: L(2/4).lwage L(1/2).L.wks L(2/3).ms L(2/3).union
        Standard: D.occ D.south D.smsa D.ind
Instruments for level equation
        Standard: _cons
```

Test Error Correlation

- Preceding requires ε_{it} serially uncorrelated
  - otherwise lagged $y's$ no longer valid instrument.
- Test ε_{it} serially uncorrelated $\Rightarrow$ $\Delta\varepsilon_{it}$ is only MA(1)
  - test based on $\Delta\widehat{\varepsilon}_{it}$ from fitted model

```
. * Test whether error is serially correlated
. estat abond

Arellano-Bond test for zero autocorrelation in first-differenced errors
```

Order	z	Prob > z
1	-4.5244	0.0000
2	-1.6041	0.1087
3	.35729	0.7209

```
H0: no autocorrelation
```

- Only reject no autocorrelation in $\Delta\varepsilon_{it}$ at order 1 (at level .05)
  - conclude we are okay as $\Delta\varepsilon_{it}$ is only MA(1)
- Note: if e.g. MA(2) then add more lagged y's to get to MA(1)
  - or use command `dpdsys` - allows e.g. ε_{it} is MA(1).

Test Overidentifying Restrictions

- There are 40 instruments and 11 parameters
  - so 29 over-identifying restrictions.
- Do OIR test: only after 2step GMM and without `vce(robust)`

```
. * Test of overidentifying restrictions (first estimate with no vce(robust))
. quietly xtabond lwage occ south smsa ind, lags(2) maxldep(3) ///
>   pre(wks,lag(1,2)) endogenous(ms,lag(0,2))              ///
>   endogenous(union,lag(0,2)) twostep artests(3)

. estat sargan
Sargan test of overidentifying restrictions
        H0: overidentifying restrictions are valid

        chi2(29)     =  39.87571
        Prob > chi2  =    0.0860
```

- Fail to reject H_0 at level .05
  - conclude that the overidentifying restrictions are not rejected.
- If $r - K$ is large there is bias for $p \rightarrow 1$.

Arellano-Bover / Blundell-Bond "Systems GMM"

- Arellano-Bond instruments the **differenced model** using lags levels as instruments
  - based on e.g. $\mathrm{E}[y_{is}\Delta(\alpha_i + \varepsilon_{it})] = 0$ for $s \leq t-2$
- Can also instrument **levels model** using lag changes as instruments
  - based on e.g. $\mathrm{E}[\Delta y_{i,t-1}(\alpha_i + \varepsilon_{it})] = 0$
    - ★ an additional assumption: $\Delta y_{i,t-1} \perp \alpha_i$ - deviations in y independent of fixed effect - see Blundell and Bond.
  - and also lagged once level of predetermined and endogenous
  - improves efficiency - intuition: model in levels has more info.

```
. * Arellano/Bover or Blundell/Bond for a dynamic panel model
. xtdpdsys lwage occ south smsa ind, lags(2) maxldep(3)     ///
>   pre(wks,lag(1,2)) endogenous(ms,lag(0,2))               ///
>   endogenous(union,lag(0,2)) twostep vce(robust) artests(3)

System dynamic panel-data estimation          Number of obs        =      2975
Group variable: id                            Number of groups     =       595
Time variable: t
                                              Obs per group:     min =         5
                                                                 avg =         5
                                                                 max =         5

Number of instruments =     60                Wald chi2(10)        =   2270.88
                                              Prob > chi2          =    0.0000
```

- Here standard errors are 10%-60% smaller
  - though most coefficients much smaller
    - ★ levels model and differences model give different answers - problem
  - and extra 20 instruments used.

```
Two-step results
------------------------------------------------------------------------------
             |              WC-Robust
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lwage |
         L1. |   .6017533   .0291502    20.64   0.000     .5446199    .6588866
         L2. |   .2880537   .0285319    10.10   0.000     .2321322    .3439752
             |
         wks |
         --. |  -.0014979   .0056143    -0.27   0.790    -.0125017     .009506
         L1. |   .0006786   .0015694     0.43   0.665    -.0023973    .0037545
             |
          ms |   .0395337   .0558543     0.71   0.479    -.0699386    .1490061
       union |  -.0422409   .0719919    -0.59   0.557    -.1833423    .0988606
         occ |  -.0508803   .0331149    -1.54   0.124    -.1157843    .0140237
       south |  -.1062817    .083753    -1.27   0.204    -.2704346    .0578713
        smsa |  -.0483567   .0479016    -1.01   0.313    -.1422422    .0455288
         ind |   .0144749    .031448     0.46   0.645    -.0471621    .0761118
       _cons |   .9584113   .3632287     2.64   0.008     .2464961    1.670327
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/4).lwage L(1/2).L.wks L(2/3).ms L(2/3).union
        Standard: D.occ D.south D.smsa D.ind
Instruments for level equation
        GMM-type: LD.lwage LD.wks LD.ms LD.union
        Standard: _cons
```

- Use Arellano-Bond when N large T small
  - need large N as asymptotics in N
  - keep number of instruments low
  - need small T as otherwise too many instruments, plus with larger T can just directly estimate fixed effects.
- Include time dummies, as require ε_{it} independent over i (not just t).
- User-written addon `xtabond2`
  - pre-dated Stata `xtabond`; still has some things that `xtabond does not`
  - forward orthogonal deviations (Helmert transformation) also eliminates α_i
    $y^{+}_{i,t=1} = \sqrt{T/(T+1)}(y_{it} - \frac{1}{T}\sum_{s=t+1}^{T} y_{is})$
    - ⋆ leads to fewer observations lost if unbalanced panel.
  - has a way to "collapse" instruments so less likely to have too many
  - David Roodman (2006): "How to Do xtabond2: An Introduction to "Difference" and "System" GMM in Stata" explains much about Arellano-Bond.

6. Short Panel Vector Autoregression

- Extend from Arellano-Bond for y_{it} to VAR for $\mathbf{y}_{it}$ (several y's)
  - Include individual specific effects $\boldsymbol{\alpha}$

$$\begin{aligned} \mathbf{y}_{it} &= \boldsymbol{\alpha}_i + \mathbf{A}_1 \mathbf{y}_{i,t-1} + \cdots + \mathbf{A}_p \mathbf{y}_{i,t-p} + \mathbf{u}_t \\ \mathbf{u}_t &\sim (\mathbf{0}, \Sigma), \end{aligned}$$

- Estimate by a systems version of Arellano and Bond (1992).
- This was actually considered in an earlier paper by Holtz-Eakin, D., W. Newey, and H.S. Rosen (1988), "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56, 1371-1395.
- Application: Love, I. and Ziccino, L. (2006), "Financial Development and Dynamic Investment Behaviour: Evidence from Panel VAR," *Quarterly Review of Economics and Finance*, 46, 190-210.

Short Panel VAR Example

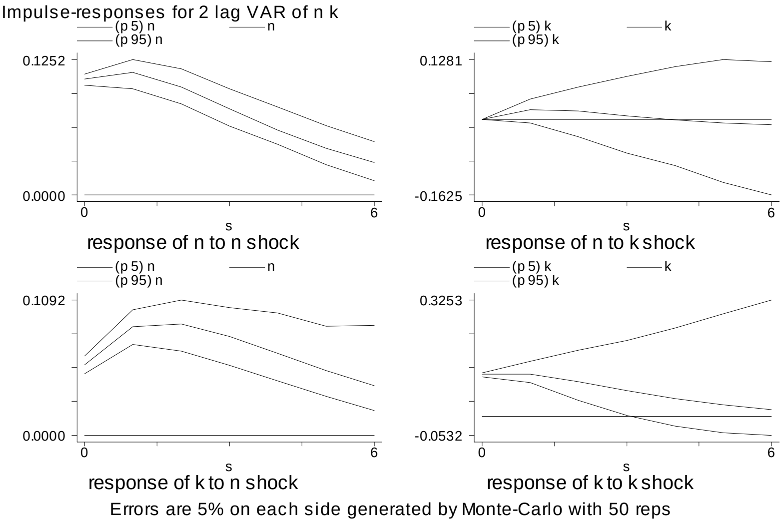
- Love has Stata user-written command: `pvar` (with `sgmm` and `helm`)
  - estimates model and graphs impulse-response functions
- As an example apply to Arellano-Bond data
  - 140 firms and 7 to 9 years of data
  - n is log employment
  - k is log capital
- First Helmert transform data
  - `use abdata.dta`
  - `helm n`
  - `helm k`
  - `pvar n k, lag(2) gmm impulse monte 50 decomp list_imp`

- Very high serial dependence

```
. pvar n k, lag(2) gmm impulse monte 50 decomp list_imp
GMM started : 15:33:18
accumulating matrices equation 1,2,calculating b2sls
calculating big zuuz matrix
finished accumulating zuuz
_______ Results of the Estimation by system GMM_________
number of observations used : 611
-----------------------------------------------------------------------------
EQ1: dep.var     : h_n

             b_GMM       se_GMM        t_GMM
 L.h_n   .96522745   .11722781   8.2337753
 L.h_k   .17832941   .15812582   1.1277691
L2.h_n  -.12987744   .10193137  -1.2741655
L2.h_k  -.20248717    .0534318  -3.7896382
-----------------------------------------------------------------------------
-----------------------------------------------------------------------------
EQ2: dep.var     : h_k

             b_GMM       se_GMM        t_GMM
 L.h_n   .28801842   .13862581    2.077668
 L.h_k   .99588169   .18395503   5.4137237
L2.h_n  -.16588097   .06990161  -2.3730636
L2.h_k  -.22096912   .05855934  -3.7734224
-----------------------------------------------------------------------------
just identified - Hansen statistic is not calculated
```


Impulse-responses for 2 lag VAR of n k
(p 5) n
n
(p 95) n
0.1252
0.0000
0
6
s
response of n to n shock
(p 5) k
k
(p 95) k
0.1281
-0.1625
0
6
s
response of n to k shock
(p 5) n
n
(p 95) n
0.1092
0.0000
0
6
s
response of k to n shock
(p 5) k
k
(p 95) k
0.3253
-0.0532
0
6
s
response of k to k shock
Errors are 5% on each side generated by Monte-Carlo with 50 reps

7. Mixed Linear Models: xtmixed and mixed

- Generalize random effects model to **random slopes**.
- Most applied statistics use mixed linear models
  - this is covered in detail below
  - `xtmixed` - Stata up to Stata 12
  - `mixed` - Stata 13 gives cluster-robust standard errors
- Econometricians also use **random coefficients model**

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

  - $(\alpha_i, \boldsymbol{\beta}_i)$ are iid with mean $(\alpha, \boldsymbol{\beta})$ and variance matrix Σ
  - ε_{it} is independent $(0, \sigma_i^2)$ so heteroskedastic over i and can do FGLS
  - command is `xtrc`. This is covered later.

Mixed or Multi-level or Hierarchical Model: xtmixed

- Not used in microeconometrics but used in many other disciplines.
- Stack all observations for individual i and specify

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i$$

  where $\mathbf{u}_i$ is iid $(\mathbf{0}, \mathbf{G})$ and $\mathbf{Z}_i$ is called a design matrix.
- Random effects: $\mathbf{Z}_i = \mathbf{e}$ (a vector of ones) and $\mathbf{u}_i = \alpha_i$
- Random coefficients: $\mathbf{Z}_i = \mathbf{X}_i$
  - Reason: $\boldsymbol{\beta}_i \sim (\boldsymbol{\beta}, \Sigma)$ so $\boldsymbol{\beta}_i \sim \boldsymbol{\beta} + \mathbf{u}_i$ where $\mathbf{u}_i \sim (\mathbf{0}, \Sigma)$
  - So $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i = \mathbf{X}_i(\boldsymbol{\beta} + \mathbf{u}_i) + \boldsymbol{\varepsilon}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{X}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i$.
  - Note: $\mathbf{y}_i|\mathbf{X}_i$ has mean $\mathbf{X}_i\boldsymbol{\beta}$ and variance $\mathbf{X}_i'\Sigma\mathbf{X}_i + \sigma_\epsilon^2\mathbf{I}$.
- Example where just `exp` and `wks` have random slopes:

```
. xtmixed lwage exp exp2 wks ed || id: exp wks, ///
. covar(unstructured) mle
```

Random Intercept

- The only random coefficient is that of the intercept
  - so RE model

```
. * Random intercept model estimated using xtmixed
. use mus08psidextract.dta, clear
(PSID wage data 1976-82 from Baltagi and Khanti-Akom (1990))

. xtmixed lwage exp exp2 wks ed || id:, mle

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood =  293.69563
Iteration 1:   log likelihood =  293.69563

Computing standard errors:

Mixed-effects ML regression                     Number of obs      =      4165
Group variable: id                              Number of groups   =       595

                                                Obs per group: min =         7
                                                               avg =       7.0
                                                               max =         7
```

- Results are identical to those from `xtreg, mle`

```
Log likelihood =  293.69563                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |   .1079955   .0024527    44.03   0.000     .1031883    .1128027
        exp2 |  -.0005202   .0000543    -9.59   0.000    -.0006266   -.0004139
         wks |   .0008365   .0006042     1.38   0.166    -.0003477    .0020208
          ed |   .1378559   .0125814    10.96   0.000     .1131968    .1625149
       _cons |   2.989858     .17118    17.47   0.000     2.654352    3.325365
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Identity                 |
                   sd(_cons) |   .8509015   .0278622       .798008    .9073008
-----------------------------+------------------------------------------------
                sd(Residual) |   .1536109   .0018574      .1500132    .1572949
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =  4576.13 Prob >= chibar2 = 0.0000
```

- Panel-robust standard errors
  - Stata 10-12: panel bootstrap takes time
  - Stata 13: `xtmixed lwage exp exp2 wks ed || id: , mle vce(robust)`.

Random Slopes

- Allow intercept and slopes of exp and wks to vary across individuals.

```
. * Random-slopes model estimated using xtmixed
. xtmixed lwage exp exp2 wks ed || id: exp wks, covar(unstructured) mle

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood =  397.61127  (not concave)
Iteration 1:   log likelihood =  427.28139  (not concave)
Iteration 2:   log likelihood =  481.60739  (not concave)
Iteration 3:   log likelihood =  493.48169
Iteration 4:   log likelihood =  495.95651
Iteration 5:   log likelihood =  508.23187
Iteration 6:   log likelihood =  509.00044
Iteration 7:   log likelihood =  509.00191
Iteration 8:   log likelihood =  509.00191

Computing standard errors:

Mixed-effects ML regression                     Number of obs      =      4165
Group variable: id                              Number of groups   =       595

                                                Obs per group: min =         7
                                                               avg =       7.0
                                                               max =         7
```

```
Log likelihood =  509.00191                     Prob > chi2       =    0.0000
```

lwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
exp	.0527159	.0032966	15.99	0.000	.0462546	.0591772
exp2	.0009476	.0000713	13.28	0.000	.0008078	.0010874
wks	.0006887	.0008267	0.83	0.405	-.0009316	.0023091
ed	.0868604	.0098652	8.80	0.000	.067525	.1061958
_cons	4.317674	.1420957	30.39	0.000	4.039171	4.596176

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
id: Unstructured				
sd(exp)	.043679	.0022801	.0394311	.0483846
sd(wks)	.0081818	.0008403	.00669	.0100061
sd(_cons)	.6042978	.0511419	.5119336	.7133265
corr(exp,wks)	-.2976598	.1000254	-.4792842	-.0915878
corr(exp,_cons)	.0036853	.0859701	-.1633389	.1705042
corr(wks,_cons)	-.4890482	.0835945	-.6352413	-.3090207
sd(Residual)	.1319489	.0017964	.1284745	.1355172

```
LR test vs. linear regression:      chi2(6) =  5006.75   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

Two-way Random Effects

- Have individual-specific intercepts and time-specific intercept that are both random
  - $y_{it} = \alpha_i + \delta_t + \mathbf{x}'_{it}\boldsymbol{\beta}_i + \varepsilon_{it}$
  - α_i iid $(\alpha, \sigma^2_\alpha)$ and δ_t iid $(0, \sigma^2_\delta)$.

```
* Two-way random-effects model estimated using xtmixed
xtmixed lwage exp exp2 wks ed || _all:  R.t || id:  , mle
```

8. Clustered Data

- **xt** commands are not just for panel data
  - can apply several of them to clustered cross-section data.
- Consider data on individual i in village j with **clustering on village**.
- A **cluster-specific model** (here village-specific) specifies

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

- Here clustering is on village (not individual) and the repeated measures are over individuals (not time).
- Assuming **equicorrelated errors** can be more reasonable here than with panel data (where correlation dampens over time).
  - So perhaps less need for `vce(robust)` after `xtreg`.

Estimators for Clustered Data

- First use `xtset village person` (versus `xtset id t` for panel).
- If α_i is **random** use:
  - `regress` with option `vce(cluster village)`
  - `xtreg,re`
  - `xtgee` with option `exchangeable`
  - `xtmixed` for Stata 10-12 for richer models of error structure.
  - `mixed` for Stata 13 for richer models of error structure.
- If α_i is **fixed** use:
  - `xtreg,fe`

9. Summary

- More linear panel estimators for short panels

Panel IV	`xtivreg`
Hausman-Taylor	`xthtaylor`
Dynamic FE models	`xtabond`, `xtdpdsys`, `xtdpd`
	`xtabond2`, `pvar` (user-written),
Random slopes	`xtmixed`; `quadchk`; `xtrc`
	`mixed` (from Stata 13 on)

- And many panel commands can be used for clustered data
  - e.g. cluster on village.

10. References

- Arellano, M., and S. Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," Review of Economic Studies, 58, 277-298.
- Holtz-Eakin, D., W. Newey, and H.S. Rosen (1988), "Estimating Vector Autoregressions with Panel Data," Econometrica, 56, 1371-1395.
- Love, I. and Ziccino, L. (2006), "Financial Development and Dynamic Investment Behaviour: Evidence from Panel VAR," Quarterly Review of Economics and Finance, 46, 190-210.
- Roodman, David (2006): "How to Do xtabond2: An Introduction to "Difference" and "System" GMM in Stata"