ECON6300/7320/8300
Advanced Microeconometrics
Panel Data II

Christiern Rose

[1]University of Queensland

Lecture 6

Fixed versus Random Effects Models

- Fundamental distinction is between models with and without fixed effects
- The econometrics literature emphasizes fixed effects.
- Some authors use the notation

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

  in to make it very clear that the individual effect is a random variable in both fixed and random effects models.
- Both models assume ε_{it} is uncorrelated with c_i and $\mathbf{x}_{it}$, so

$$\mathrm{E}[y_{it}|c_i, \mathbf{x}_{it}] = c_i + \mathbf{x}_{it}'\beta.$$

Identification in FE & RE Models

- Individual-specific effect c_i is unknown and in short panels cannot be consistently estimated, so we cannot estimate $\mathrm{E}[y_{it}|c_i, \mathbf{x}_{it}]$. Instead we can eliminate c_i be taking the expectations conditional on $\mathbf{x}_{it}$, leading to

$$\mathrm{E}[y_{it}|\mathbf{x}_{it}] = \mathrm{E}[c_i|\mathbf{x}_{it}] + \mathbf{x}_{it}'\beta.$$

- For the RE model it is assumed that $\mathrm{E}[c_i|\mathbf{x}_{it}] = \alpha$ so $\mathrm{E}[y_{it}|\mathbf{x}_{it}] = \alpha + \mathbf{x}_{it}'\beta$ and hence it is possible to identify $\mathrm{E}[y_{it}|\mathbf{x}_{it}]$.

Identification in FE & RE Models

- In the FE model, however, $\mathrm{E}[c_i|\mathbf{x}_{it}]$ varies with $\mathbf{x}_i$, and it is not known how it varies, so we cannot identify $\mathrm{E}[y_{it}|\mathbf{x}_{it}]$.
- It is nonetheless possible to consistently estimate β in the FE model with short panels.

- Thus it is possible in the FE model to identify the marginal effect

$$\beta = \partial \mathrm{E}[y_{it}|c_i, \mathbf{x}_{it}]/\partial \mathbf{x}_{it},$$

even though the conditional mean is not identified.

Two-step RE FGLS Estimator (1)

- Begin with the individual-specific effect model, but assume a random effects model where α_i and ε_{it} are iid.
- Pooled OLS is consistent but pooled GLS will be more efficient.
- General form of GLS using transformations based on the decomposition $\Omega^{-1} = \mathbf{T}'\mathbf{T}$:

$$
\begin{aligned}
\widehat{\beta}_{GLS} &= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{y} \\
&= (\mathbf{X}'\mathbf{T}'\mathbf{T}\mathbf{X})^{-1}\mathbf{X}'\mathbf{T}'\mathbf{T}\mathbf{y} \text{ as } \Omega^{-1} = \mathbf{T}'\mathbf{T} \\
&= (\mathbf{X}^{*\prime}\mathbf{X}^{*})^{-1}\mathbf{X}^{*\prime}\mathbf{y}^{*} \text{ where } \mathbf{X}^{*} = \mathbf{T}(\lambda)\mathbf{X}
\end{aligned}
$$

Two-step RE FGLS Estimator (2)

- Which transformation?

$$y^*_{it} = y_{it} - \widehat{\lambda}\bar{y}_i;\ \mathbf{x}^*_{it} = (\mathbf{x}_{it} - \widehat{\lambda}\bar{x}_i)'$$

where $\lambda = 1 - \sigma_\varepsilon / \sqrt{\sigma^2_\varepsilon + T\sigma^2_\alpha}$

- Feasible GLS estimator of the RE model, called the **random effects estimator**, can be calculated from OLS estimation of the transformed model

$$y_{it} - \widehat{\lambda}\bar{y}_i = (\mathbf{x}_{it} - \widehat{\lambda}\bar{x}_i)'\beta + v_{it}, \qquad (1)$$

where $v_{it} = (1 - \widehat{\lambda})\alpha_i + (\varepsilon_{it} - \widehat{\lambda}\bar{\varepsilon}_i)$ is asymptotically iid, and $\widehat{\lambda}$ is consistent for λ

- How to estimate σ^2_α and σ^2_ε and hence λ? Note that $\widehat{\lambda} = 0$ corresponds to pooled OLS, $\widehat{\lambda} = 1$ is within estimation and $\lambda \to 1$ as $T \to \infty$.

Two-step RE FGLS Estimator (3)

- To implement GLS for the RE model we need consistent estimates of σ^2_ε and σ^2_α.
- From the within or FE regression of $(y_{it} - \bar{y}_i)$ on $(\mathbf{x}_{it} - \bar{x}_i)$, which has errors $(\varepsilon_{it} - \bar{\varepsilon}_i)$,

$$\widehat{\sigma}^2_\varepsilon = \frac{1}{N(T-1)-K} \sum_i \sum_t ((y_{it} - \bar{y}_i) - (\mathbf{x}_{it} - \bar{x}_i)'\widehat{\beta}_{\mathrm{W}}))^2. \quad (2)$$

- From the between regression of $\bar{y}_i$ on an intercept and $\bar{\mathbf{x}}_i$, an equation which has has error term $(\alpha_i - \alpha + \bar{\varepsilon}_i)$ with variance $\sigma^2_\alpha + \sigma^2_\varepsilon / T$

$$\widehat{\sigma}^2_\alpha = \frac{1}{N-(K+1)} \sum_i (\bar{y}_i - \bar{x}_i'\widehat{\beta}_{\mathrm{B}})^2 - \frac{1}{T}\widehat{\sigma}^2_\varepsilon. \quad (3)$$

Two-step RE FGLS Estimator (4)

- More efficient estimators of the variance components σ_ε^2 and σ_α^2 are possible. See for example Amemiya (1985), but these will not necessarily increase the efficiency of $\widehat{\beta}_{\text{RE}}$.

- Variance estimator (3) <u>can be</u> negative, in which case programs often set $\widehat{\sigma}_\alpha^2 = 0$ so $\widehat{\lambda} = 0$ and estimation is then by pooled OLS.

Panel robust variance (1)

- Consider the following regression:

$$\widetilde{y}_{it} = \widetilde{\mathbf{x}}_{it}'\beta + \widetilde{u}_{it}, \tag{4}$$

where different panel estimators correspond to different transformations $\widetilde{y}_{it}$, $\widetilde{\mathbf{x}}_{it}$ and $\widetilde{u}_{it}$ and we incorporate the constant into $\boldsymbol{x}_{it}$.

- The key is that $\widetilde{y}_{it}$ is a known function of only $y_{i1}, ..., y_{iT}$, and similarly for $\widetilde{x}_{it}$ and $\widetilde{u}_{it}$. (i.e. it depends only on individual i).

- In the simplest case of pooled OLS, no transformation is necessary.

- For the within estimator $\widetilde{y}_{it} = y_{it} - \bar{y}_i$, $\widetilde{\mathbf{x}}_{it} = (\mathbf{x}_{it} - \bar{x}_i)$ where only time-varying regressors appear.
- For first differences estimation $\widetilde{y}_{it} = y_{it} - y_{i,t-1}$, $\widetilde{\mathbf{x}}_{it} = (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})$ where only time-varying regressors appear.
- For random effects $\widetilde{y}_{it} = y_{it} - \widehat{\lambda}\bar{y}_i$ and $\widetilde{\mathbf{x}}_{it}' = (\mathbf{x}_{it} - \widehat{\lambda}\bar{x}_i)$.

Panel robust variance (2)

- Stack observations over time periods for a given individual, leading to
$$\widetilde{\mathbf{y}}_i = \widetilde{\mathbf{X}}_i\beta + \widetilde{\mathbf{u}}_i,$$
where $\widetilde{\mathbf{y}}_i$ is a $T \times 1$ vector in the preceding examples, except for the first difference model where it is $(T-1) \times 1$, and $\widetilde{\mathbf{X}}_i$ is a $T \times q$ or, for the first difference model, a $(T-1) \times q$ matrix.
- Further stacking over the N individuals yields
$$\widetilde{\mathbf{y}} = \widetilde{\mathbf{X}}\beta + \widetilde{\mathbf{u}}.$$

Panel robust variance (3)

- Three representations of the OLS estimator are:

$$
\begin{aligned}
\widehat{\beta}_{\text{OLS}} &= [\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}]^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{y}} \\
&= \left[\sum\nolimits_{i=1}^{N} \widetilde{\mathbf{X}}_i{}'\widetilde{\mathbf{X}}_i\right]^{-1} \sum \widetilde{\mathbf{X}}_i{}'\widetilde{\mathbf{y}}_i \\
&= \left[\sum\nolimits_{i=1}^{N} \sum\nolimits_{t=1}^{T} \widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}_{it}'\right]^{-1} \sum\nolimits_{i=1}^{N} \sum\nolimits_{t=1}^{T} \widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{y}}_{it},
\end{aligned}
$$

where in third expression the sum is from $t = 2$ to T in the case of the first difference estimator.

Panel Robust Variance estimation (4)

- To consider consistency, note that if the model is correctly specified then the usual algebra yields

$$\widehat{\beta}_{\text{OLS}} = \beta + \left[\sum_{i=1}^{N} \widetilde{\mathbf{X}}_i{}'\widetilde{\mathbf{X}}_i\right]^{-1} \sum_{i=1}^{N} \widetilde{\mathbf{X}}_i{}'\widetilde{\mathbf{u}}_i.$$

Given independence over i the essential condition for consistency is $\mathrm{E}[\widetilde{\mathbf{X}}_i{}'\widetilde{\mathbf{u}}_i] = 0$. A sufficient assumption is that of strict exogeneity.

- Asymptotic variance of $\widehat{\beta}_{\text{OLS}}$ is

$$\mathrm{V}[\widehat{\beta}_{\text{OLS}}|\mathbf{X}] = \left[\sum_{i=1}^{N} \widetilde{\mathbf{X}}_i{}'\widetilde{\mathbf{X}}_i\right]^{-1} \sum_{i=1}^{N} \widetilde{\mathbf{X}}_i{}'\mathrm{E}[\widetilde{\mathbf{u}}_i\widetilde{\mathbf{u}}_i'|\widetilde{\mathbf{X}}_i]\widetilde{\mathbf{X}}_i \left[\sum_{i=1}^{N} \widetilde{\mathbf{X}}_i{}'\widetilde{\mathbf{X}}_i\right]^{-1}$$

given independence of errors over i.

Variance estimation (1)

- Consistent estimation of $\mathrm{V}[\widehat{\beta}_{\mathrm{OLS}}]$ is analogous to the cross-section problem of obtaining a consistent estimate of $\mathrm{V}[\widehat{\beta}_{\mathrm{OLS}}]$ that is robust to heteroskedasticity of unknown form.

- The only complication is the appearance of a vector $\mathbf{u}_i$ rather than a scalar u_i, which poses no problem if the panel is short.

- **Panel robust estimate** of the asymptotic variance matrix of OLS estimator, controlling for both SC and H,

$$\widehat{\mathrm{V}}[\widehat{\beta}_{\mathrm{OLS}}] = \left[\sum\nolimits_{i=1}^{N} \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\right]^{-1} \sum\nolimits_{i=1}^{N} \widetilde{\mathbf{X}}_i'\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\widetilde{\mathbf{X}}_i \left[\sum\nolimits_{i=1}^{N} \widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\right]^{-1},$$

assuming independence over i and $N \to \infty$. $\mathrm{V}[u_{it}]$ and $\mathrm{Cov}[u_{it}, u_{is}]$ may vary with i, t and s.

Variance estimation (2)

- Above formula uses matrix notation which suppresses the t dimension of the data.
- Equivalently (to bring out the i and the t dimensions)

$$\widehat{V}[\widehat{\beta}_{\text{OLS}}] = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}_{it}'\right]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}_{is}'\widehat{u}_{it}\widehat{u}_{is}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\mathbf{x}}_{it}\widetilde{\mathbf{x}}_{it}'\right]^{-1}$$

  This estimator was proposed by Arellano (1987) for the fixed effects estimator.
- Can be computed by use of a regular OLS command, if the command has a cluster robust standard error option. Then one selects the identifier for individual i as the cluster variable.

Hausman Test (1)

- For FE $\widehat{\beta}_{\mathrm{W}}$ is consistent while the random effects estimator of coefficients of just the time-varying regressors $\widetilde{\beta}_{\mathrm{RE}}$ is inconsistent
- Test for presence of fixed effects are present by a Hausman test of whether there is a statistically significant difference between these estimators.
- A large Hausman test statistic $\Rightarrow$ rejection of the null hypothesis that the individual specific effects are uncorrelated with regressors.
- Then use FE or add additional regressors.

Hausman Test Computation

- Assume RE with α_i iid $[0, \sigma_\alpha^2]$ uncorrelated with regressors and error ε_{it} iid $[0, \sigma_\varepsilon^2]$.
- Let β_1 be the parameters of the time-varying regressors (i.e. those which are estimable under the FE model)
- Then the estimator $\widetilde{\beta}_{\mathrm{RE}}$ is fully efficient, so

$$\mathrm{H} = \left(\widetilde{\beta}_{1,\mathrm{RE}} - \widehat{\beta}_{\mathrm{W}}\right)' \left[\widehat{\mathrm{V}}[\widehat{\beta}_{1,\mathrm{W}}] - \widehat{\mathrm{V}}[\widetilde{\beta}_{1,\mathrm{RE}}]\right]^{-1} \left(\widetilde{\beta}_{1,\mathrm{RE}} - \widehat{\beta}_{1,\mathrm{W}}\right),$$

is asymptotically $\chi^2(\dim(\beta_1))$ distributed under the null hypothesis.

Computation when RE is not fully efficient

- H-test invalid if α_i or ε_{it} are not iid, but (say) heteroskedastic. Then the RE estimator is not fully efficient under the null hypothesis so the expression $\widehat{\mathrm{V}}[\widehat{\beta}_{\mathrm{W}}] - \widehat{\mathrm{V}}[\widetilde{\beta}_{\mathrm{RE}}]$ in the formula for H needs to be replaced by the more general $\widehat{\mathrm{V}}[\widetilde{\beta}_{\mathrm{RE}} - \widehat{\beta}_{\mathrm{W}}]$.

- For short panels this can be consistently estimated by bootstrap resampling over i.

- See last week's computer class for another method by Wooldridge (2002)

Cluster-robust inference

- Many methods assume ε_{it} and α_i (if present) are iid.
- Yields **wrong standard errors** if heteroskedasticity or if errors not equicorrelated over time for a given individual.
- For short panel can relax and use **cluster-robust inference.**
  - Allows heteroskedasticity and general correlation over time for given i.
  - Independence over i is still assumed.
- For `xtreg` the option `vce(robust)` does cluster-robust
- For some other `xt` commands use option `vce(cluster)`
- And for some other `xt` commands there is no option but may be able to do a cluster bootstrap.

Fixed effects versus random effects

- Prefer RE as can estimate all parameters and more efficient.
- But RE is inconsistent if fixed effects present.
- Use **Hausman test** to discriminate between FE and RE.
  - This tests difference between FE and RE estimates is statistically significantly different from zero.
  - If we do not reject the null, then favour RE. Otherwise, favour FE.
  - Many (almost all) econometricians always use FE.
- Problem: `hausman` command gives incorrect statistic as it assumes RE estimator is fully efficient, usually not the case.
- Solution: do a panel bootstrap of the Hausman test or use the Wooldridge (2002) robust version of Hausman test.

Panel IV: xtivreg command

- Command `xtivreg` is natural extension of `ivregress` to panels.
- Consider model with possibly transformed variables:

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

where transformations are

OLS	$y_{it}^{*} = y_{it}$
Between	$y_{it}^{*} = \bar{y}_i$
Within	$y_{it}^{*} = (y_{it} - \bar{y}_i)$
FD	$y_{it}^{*} = (y_{it} - y_{it-1})$
Random effects	$y_{it}^{*} = (y_{it} - \lambda\bar{y}_i)$

- OLS is **consistent** if $\mathrm{E}[u_{it}^{*}|\mathbf{x}_{it}^{*}] = 0$.
- **IV estimation** with **instruments** $\mathbf{z}_{it}^{*}$ satisfying $\mathrm{E}[u_{it}^{*}|\mathbf{z}_{it}^{*}] = 0$.
- Example: `xtivreg lwage exp exp2 (wks = ms), fe`

IV for FE and RE Models

- Model is a **random effects** model if instruments $\mathbf{Z}_i$ exist that satisfy $\mathrm{E}[\mathbf{Z}_i'(\alpha_i + \varepsilon_{it})] = \mathbf{0}$. Then the above methods will permit consistent estimation of all regression parameters.

- If there are instruments such that $\mathrm{E}[\mathbf{Z}_i'\varepsilon_i] = \mathbf{0}$, but $\mathrm{E}[\mathbf{Z}_i'\alpha_i] \neq \mathbf{0}$, model is a **fixed effects** model.
  - Then α_i must be eliminated, in which case only the coefficients of time-varying regressors will be identified.

Example: Dynamic panel models

- Consider the dynamic panel data model:

$$y_{it} = \beta y_{it-1} + \alpha_i + \epsilon_{it}, \quad T \geq 2$$

$$E[\epsilon_{it}] = 0, \quad E[\epsilon_{it}\epsilon_{is}] = 0, \quad E[\alpha_i \epsilon_{it}] = 0 \quad \forall i, t, s \neq t$$

- Which is appropriate: RE or FE?
- If FE, which transformation to eliminate α_i? (Within? First differences?)
- Which instruments?! (You only have data on y_{it})

IV for Fixed Effects Models (I)

- The various differencing operations generate a transformed model of the form

$$\widetilde{y}_{it} = \widetilde{\mathbf{x}}_{it}'\beta + \widetilde{\varepsilon}_{it},$$

  where the tilda denotes a differencing transformation that eliminates α_i. Upon stacking

$$\widetilde{\mathbf{y}}_i = \widetilde{\mathbf{X}}_i\beta + \widetilde{\varepsilon}_i. \tag{5}$$

  If $\mathrm{E}[\mathbf{x}_{it}\varepsilon_{it}] \neq 0$ then $\mathrm{E}[\widetilde{\mathbf{x}}_{it}\widetilde{\varepsilon}_{it}] \neq 0$ and LS estimation of (5) $\Rightarrow$ inconsistent estimates.
- If instruments $\mathbf{Z}_i$ satisfy $\mathrm{E}[\mathbf{Z}_i'\widetilde{\varepsilon}_i] = \mathbf{0}$, panel GMM estimation (IV, 2SLS or 2SGMM) of (5) with instruments $\mathbf{Z}_i$ yields consistent estimates of the coefficients of time-varying regressors.

- Panel robust standard errors can be computed as discussed earlier.
- Valid IVs are strictly exogenous regressors in periods other than t (e.g. w_{it-2}).

IV for Fixed Effects Models (II)

- Primitive assumptions for instrument availability are those on correlation between $\mathbf{z}_{is}$ and ε_{it}. But here it is correlation between $\mathbf{z}_{is}$ and the differenced error $\widetilde{\varepsilon}_{it}$ that matters.
- In general differencing, necessary to eliminate the fixed effect, reduces the number of available instruments.

- Some differencing operations lead to greater loss than others and can even lead to inconsistent IV estimation.
- We consider three differencing operations with focus on weakly exogenous instruments.

IV for the First Differences Model

- The **first difference IV estimator** is the IV or 2SLS estimator of the first differenced model

$$y_{it} - y_{i,t-1} = (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})'\beta + (\varepsilon_{it} - \varepsilon_{i,t-1}) \quad (6)$$

  Weak exogeneity assumption:

$$E[\mathbf{z}_{is}\varepsilon_{it}] = \mathbf{0} \text{ for } s \leq t \implies E[\mathbf{z}_{is}(\varepsilon_{it} - \varepsilon_{i,t-1})] = \mathbf{0}$$

  for $s \leq t - 1$. First differencing therefore reduces the available instrument set by one period

- Using lagged regressors as instruments was first proposed by Anderson and Hsiao (1981) in the context of dynamic models and expanded upon by Holtz-Eakin, Newey and Rosen (1988) and Arellano and Bond (1991).

IV for Within (Mean Differenced) Model

- The **within IV estimator** is the IV or 2SLS estimator of the within model

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{x}_i)'\beta + (\varepsilon_{it} - \bar{\varepsilon}_i). \quad (7)$$

  Then $\mathrm{E}[\mathbf{z}_{is}\varepsilon_{it}] = \mathbf{0}$ for $s < t$ no longer implies $\mathrm{E}[\mathbf{z}_{is}(\varepsilon_{it} - \bar{\varepsilon}_i)] = \mathbf{0}$.

- Thus IV estimation of the within model leads to inconsistent estimation of β if the instruments are weakly exogenous, or satisfy the even weaker assumptions of contemporaneous exogeneity.

- The within transformation can only be used if the instruments are strictly exogenous.

IV for Random Effects Models (1)

- Stacked model where $\mathbf{1}_T$ is a $T \times 1$ vector of ones:

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{1}_T\alpha_i + \varepsilon_i$$

- Consistent but inefficient estimates can be obtained by directly applying the panel GMM estimators given instruments $\mathbf{Z}_i$, satisfying $\mathrm{E}[\mathbf{Z}_i'(\mathbf{1}_T\alpha_i + \varepsilon_i)] = \mathbf{0}$.

- Consider more efficient estimation. Assume that the instruments $\mathbf{Z}_i$ satisfy $\mathrm{E}[\mathbf{u}_i|\mathbf{Z}_i] = \mathbf{0}$ and $\mathrm{V}[\mathbf{u}_i|\mathbf{Z}_i] = \Omega_i$.
- Given $\mathrm{E}[\mathbf{u}_i|\mathbf{Z}_i] = \mathbf{0}$, optimal unconditional moment condition is

$$\mathrm{E}[\mathbf{Z}_i'\Omega_i^{-1}\mathbf{u}_i] = \mathrm{E}[(\Omega_i^{-1/2}\mathbf{Z}_i)'(\Omega_i^{-1/2}\mathbf{u}_i)] = \mathbf{0}.$$

IV for Random Effects Models (2)

- GMM estimation in the transformed system $\mathbf{y}_i^* = \mathbf{X}_i^*\beta + \mathbf{u}_i^*$ with transformed instruments $\mathbf{Z}_i^*$, where the asterisk denotes premultiplication by the $T \times T$ matrix $\Omega_i^{-1/2}$ or a consistent estimate $\widehat{\Omega}_i^{-1/2}$.

- Premultiplication by $\widehat{\Omega}_i^{-1/2}$ leads to the model

$$y_{it} - \widehat{\lambda}\bar{y}_i = (\mathbf{x}_{it} - \widehat{\lambda}\bar{\mathbf{x}}_i)'\beta + \{(1-\widehat{\lambda})\alpha_i + (\varepsilon_{it} - \widehat{\lambda}\bar{\varepsilon}_i)\}, \quad (8)$$

where $\widehat{\lambda}$ is a consistent estimate of $\lambda = 1 - \sigma_\varepsilon / \sqrt{\sigma_\varepsilon^2 + T\sigma_\alpha^2}$.

- **RE IV estimator** is the IV or 2SLS estimator of this model with transformed instruments $\widetilde{\mathbf{z}}_{it} = (\mathbf{z}_{it} - \widehat{\lambda}\bar{z}_i)$, or equivalently with instruments $\mathbf{z}_{it} - \bar{z}_i$ and $\bar{z}_i$.

- This method requires a consistent estimate $\widehat{\lambda}$ of λ. Results are dependent on specification of a particular functional form for Ω_i. This method can only be used if the original instruments are strictly exogenous.

Panel GMM (PGMM) (1)

- Model:

$$y_{it} = \mathbf{x}_{it}'\beta + u_{it}, \tag{9}$$

  where the regressors $\mathbf{x}_{it}$ may have both time-varying and time-invariant components and include an intercept.
- No individual specific effect α_i, and $\mathbf{x}_{it}$ is assumed to include only current period variables.
- Assume: observations independent over i and a short panel with T fixed and $N \to \infty$.

Stack all T observations for the i^{th} individual,

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{u}_i, \tag{10}$$

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

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix}; \quad \mathbf{X}_i = \begin{bmatrix} \mathbf{x}_{i1}' \\ \vdots \\ \mathbf{x}_{iT}' \end{bmatrix}; \quad \mathbf{u}_i = \begin{bmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{bmatrix}.$$

Panel GMM (2)

- Assume the existence of a $T \times r$ matrix of instruments $\mathbf{Z}_i$, where $r \geq K$ is the number of instruments, that satisfy

$$\mathrm{E}[\mathbf{Z}_i'\mathbf{u}_i] = \mathrm{E}[\mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\beta)] = \mathbf{0}. \tag{11}$$

- GMM estimator based on these moment conditions minimizes the associated quadratic form

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

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

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

Panel GMM (3)

- In many applications $\mathbf{Z}_i$ is composed of current and lagged values of exogenous regressors.
- If, for example, all regressors are contemporaneously exogenous, then $\mathrm{E}[\mathbf{x}_{it} u_{it}] = \mathbf{0}$ implies (11) with $\mathbf{Z}_i = [\mathbf{x}'_{i1} \ldots \mathbf{x}'_{iT}]$.
- In this case the model is just identified and, since $\mathbf{Z}_i = \mathbf{X}_i$, $\widehat{\beta}_{\mathrm{PGMM}}$ simplifies to the pooled OLS estimator.
- If $\mathrm{E}[\mathbf{x}_{it-1} u_{it}] = \mathbf{0}$, then $\mathbf{x}_{it-1}$ is available as **additional instruments** for the it^{th} observation, the model is over-identified and more efficient estimation is possible using the PGMM estimator.

Panel Variance estimator (1)

- Using more compact notation. Rewrite

$$\widehat{\beta}_{\text{PGMM}} = [\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{y}, \tag{12}$$

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 estimated asymptotic variance matrix

$$\widehat{\mathrm{V}}[\widehat{\beta}_{\text{PGMM}}] = [\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_N(N\widehat{\mathbf{S}})\mathbf{W}_N'\mathbf{Z}'\mathbf{X}[\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}, \tag{13}$$

where $\widehat{\mathbf{S}}$ is a consistent estimate of the $r \times r$ matrix

$$\mathbf{S} = \text{plim}\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i, \tag{14}$$

and independence over i has been assumed.

Panel Variance estimator (2)

- Essential assumption for this result is that $N^{-1/2}\mathbf{Z}'\mathbf{u} = N^{-1/2}\sum_i \mathbf{Z}_i\mathbf{u}_i \xrightarrow{d} \mathcal{N}[\mathbf{0}, \mathbf{S}]$.
- White-type robust estimate of $\mathbf{S}$ is

$$\widehat{\mathbf{S}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{Z}_i'\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\mathbf{Z}_i \qquad (15)$$

where the $T \times 1$ estimated residual $\widehat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\widehat{\beta}$.
- The estimate (15) yields panel robust standard errors allowing for both heteroskedasticity and correlation over time. Alternatively the panel bootstrap could be used.

One-Step GMM

- The **one-step GMM** or **two-stage least squares estimator** uses weighting matrix $\mathbf{W}_N = [\sum_i \mathbf{Z}_i'\mathbf{Z}_i]^{-1} = [\mathbf{Z}'\mathbf{Z}]^{-1}$, leading to

$$\widehat{\beta}_{2SLS} = [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}. \qquad (16)$$

  The motivation for this estimate is that it can be shown to be the optimal PGMM estimator based on (11) if $\mathbf{u}_i|\mathbf{Z}_i$ is iid $[\mathbf{0}, \sigma^2\mathbf{I}_T]$.
- Called one-step GMM as given the data it can be directly calculated using the formula (16).
- Called 2SLS as it can instead be obtained in two stages by (1) OLS of $\mathbf{X}_i$ on $\mathbf{Z}_i$ yielding prediction $\widehat{\mathbf{X}}_i$; and (2) OLS of $\mathbf{y}_i$ on $\widehat{\mathbf{X}}_i$.
- Panel and heteroskedasticity robust estimate of the variance matrix of $\widehat{\beta}_{2SLS}$ is that given in (13) with $\mathbf{W}_N = [\mathbf{Z}'\mathbf{Z}]^{-1}$.

Two-Step GMM

- The most efficient GMM estimator based on the unconditional moment condition (11) uses weighting matrix $\mathbf{W}_N = \widehat{\mathbf{S}}^{-1}$ where $\widehat{\mathbf{S}}$ is consistent for $\mathbf{S}$ defined in (14)
- Using $\widehat{\mathbf{S}}$ in (15) yields the **two-step GMM estimator**

$$\widehat{\beta}_{2\mathrm{SGMM}} = [\mathbf{X}'\mathbf{Z}\widehat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}\widehat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{y}. \tag{17}$$

- Then (13) simplifies and $\widehat{\mathrm{V}}[\widehat{\beta}_{2\mathrm{SGMM}}] = [\mathbf{X}'\mathbf{Z}(N\widehat{\mathbf{S}})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}$.
- Called two-step GMM since a first step consistent estimator of β such as $\widehat{\beta}_{2\mathrm{SLS}}$ is needed to form the residuals $\widehat{\mathbf{u}}_i$ used to compute $\widehat{\mathbf{S}}$.

Efficiency Gains

- We consider situations where $\mathbf{Z}$ cannot contain all of $\mathbf{X}$, due to endogeneity of some components of $\mathbf{X}$. Then panel GMM provides consistent estimates when OLS does not.
- 2SGMM provides the most efficient estimator based on the moment condition $\mathrm{E}[\mathbf{Z}_i'\mathbf{u}_i] = \mathbf{0}$.
- Even under strict exogeneity assumption, 2SGMM is more efficient than pooled OLS.
  - Explanation: suppose $\mathbf{X}$ is strictly exogenous. Set $\mathbf{Z} = \mathbf{X}$, then 2SGMM estimator simplifies to $[\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{y}$ and there is no benefit to panel GMM.
  - But if instead $\mathbf{Z} = \mathbf{X}$ **plus** some additional variables, such as powers of the regressors or regressor values in periods other than the current period, then the 2SGMM is at least as efficient as OLS, with equality if the error u_{it} is i.i.d.

Test of Over-identifying Restrictions

- If there are r instruments and only K parameters to estimate, then panel GMM estimations leaves $(r - K)$ overidentifying restrictions.

- Leads to a **test of overidentifying restrictions**

$$\text{OIR} = \left[\sum_{i=1}^{N} \widehat{\mathbf{u}}_i' \mathbf{Z}_i\right] (N\widehat{\mathbf{S}})^{-1} \left[\sum_{i=1}^{N} \mathbf{Z}_i' \widehat{\mathbf{u}}_i\right], \tag{18}$$

where $\widehat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i' \widehat{\beta}_{2SGMM}$, $\widehat{\mathbf{S}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{Z}_i' \widehat{\mathbf{u}}_i \widehat{\mathbf{u}}_i' \mathbf{Z}_i$ and independence over i is assumed but heteroskedasticity and correlation over t for given i is permitted.

- This test statistic is distributed as $\chi^2(r - K)$ under the H_0 that the overidentifying restrictions are valid.

Selection of Instruments

- Discussion so far assumes the existence of a $T \times r$ matrix of instruments $\mathbf{Z}_i$ that satisfies (11).
- With cross-section data, endogenous variables are instrumented by variables that do not appear as regressors in the equation of interest. Such variables can also be used as instruments in the panel case.
- *With panel data, however, the additional periods of data provide additional moment conditions and additional instruments that can easily lead to identification or over-identification of* β
- Number of moment conditions and instruments available **expands as progressively stronger assumptions** are made about the correlation between u_{it} and $\mathbf{z}_{is}$, $s, t = 1, ..., T$.