Lecture 6: Linear Panel Data Model, II

February 6, 2025

First-Differencing Estimation

- Like FE, FD removes c_i. But it does it by differencing adjacent observations. FE and FD are the same when $T = 2$, but differ otherwise. Again, start with the original equation:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, t = 1, \ldots, T.$$

For FD, we explicitly lose the first time period:

$$\Delta y_{it} = \boldsymbol{\Delta}\mathbf{x}_{it}\boldsymbol{\beta} + \Delta u_{it}, t = 2, \ldots, T.$$

The FD estimator is pooled OLS on the first differences.

- In practice, might not difference period dummies, unless interested in the year intercepts in the original levels.
- FD also requires a kind of strict exogeneity. The weakest assumption is

$$E\left(\boldsymbol{\Delta}\mathbf{x}_{it}'\Delta u_{it}\right) = 0, t = 2, \ldots, T.$$

- Failure of strict exogeneity will cause different inconsistencies in FE and FD when $T > 2$.

First-Differencing Estimation

- A sufficient condition is

  Assumption FD.1: Same as FE. 1, $E\left(u_{it} \mid \mathbf{x}_i, c_i\right) = 0, t = 1, \ldots, T$.

  Assumption FD.2: Let $\Delta\mathbf{X}_i$ be the $(T-1) \times K$ matrix with rows $\Delta\mathbf{x}_{it}$. Then,

  $$\operatorname{rank} E\left(\Delta\mathbf{X}_i'\Delta\mathbf{X}_i\right) = K$$

- Should make inference robust to serial correlation and heteroskedasticity in the differenced errors, $e_{it} \equiv u_{it} - u_{i,t-1}$. For example, if $\{u_{it}\}$ is uncorrelated, $\operatorname{Corr}\left(e_{it}, e_{i,t+1}\right) = -.5$.

- After POLS on the first differences, let

  $$\hat{e}_{it} = \Delta y_{it} - \Delta\mathbf{x}_{it}\hat{\beta}_{FD}, t = 2, \ldots, T; i = 1, \ldots, N$$

  and let $\hat{\mathbf{e}}_i = \left(\hat{e}_{12}, \ldots, \hat{e}_{iT}\right)'$ be the $(T-1) \times 1$ residuals. Then

  $$\widehat{\text{Avar}}\left(\hat{\boldsymbol{\beta}}_{FD}\right) = \left(\sum_{i=1}^{N} \Delta\mathbf{X}_i'\Delta\mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \Delta\mathbf{X}_i'\hat{e}_i\hat{\mathbf{e}}_i'\Delta\mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \Delta\mathbf{X}_i'\Delta\mathbf{X}_i\right)^{-}$$

  is the fully robust variance matrix estimator.

- Use pooled OLS, on the first differences and then use a "cluster" option.

First-Differencing Estimation

Assumption FD.3:

$$E\left(\mathbf{e}_i\mathbf{e}_i' \mid \Delta\mathbf{X}_i\right) = \sigma_e^2\mathbf{I}_T$$

where $\sigma_e^2 = E\left(e_{it}^2\right)$ for all t.

- Under Assumption FE.3, the usual POLS statistics in the FD regression are asymptotically valid.
- If we believe FD.3, then $u_{it} = u_{i,t-1} + e_{it}$ is a random walk. In a pure time series setting, this means the regression would be "spurious."
- Here we can allow random walk behavior in $\{u_{it}\}$ with a short T because we have cross section variation driving the large-sample analysis.

Serial Correlation

- Testing for serial correlation in $\{e_{it} = \Delta u_{it}\}$ is easy. If we start with $T \geq 3$, then use a t test or heteroskedasticity-robust version for $\hat{\delta}$, where $\hat{\delta}$ is the coefficient on $\hat{e}_{i,t-1}$ in the pooled dynamic OLS regression

$$\hat{e}_{it} \text{ on } \hat{e}_{i,t-1}, t = 3, \ldots, T; i = 1, \ldots, N.$$

- We can also use this regression to test whether $\operatorname{Cor}(e_{it}, e_{i,t-1}) = -.5$, as implied by FE.3. But then the standard error of $\hat{\delta}$ should be made robust to serial correlation. The t statistic in this case is

$$\frac{(\hat{\delta} + .5)}{\mathrm{se}(\hat{\delta})}.$$

Serial Correlation

- Can use the FD residuals to recover an estimate of ρ if we think $\{u_{it} : t = 1, 2, \ldots T\}$ follows a stationary AR(1) process. Then $\operatorname{Cov}(u_{it}, u_{i,t-h}) = \rho^h \sigma_u^2, h = 0, 1, \ldots$. Therefore

$$\begin{aligned}
\operatorname{Cov}(e_{it}, e_{i,t-1}) &= \operatorname{Cov}(u_{it} - u_{i,t-1}, u_{i,t-1} - u_{i,t-2}) \\
&= \rho\sigma_{it}^2 - \rho^2\sigma_{it}^2 - \sigma_u^2 + \rho\sigma_u^2 \\
&= -\sigma_u^2\left(1 - 2\rho + \rho^2\right) \\
&= -\sigma_u^2(1-\rho)^2
\end{aligned}$$

- Further,

$$\begin{aligned}
\operatorname{Var}(e_{it}) &= \sigma_u^2 - 2\operatorname{Cov}(u_{it}, u_{i,t-1}) + \sigma_u^2 \\
&= 2\sigma_u^2(1-\rho)
\end{aligned}$$

Serial Correlation

- It follows that

$$\mathrm{Corr}\,(e_{it}, e_{i,t-1}) = \frac{-\sigma_u^2(1-\rho)^2}{2\sigma_u^2(1-\rho)} = \frac{(\rho-1)}{2}$$

  Letting $\delta \equiv \mathrm{Corr}\,(e_{it}, e_{i,t-1})$, we can write

$$\rho = 1 + 2\delta$$

- Notice we get the right answer when $\delta = 0$: namely, $\rho = 1$ (so that $\{u_{it}\}$ follows a random walk). So we can use

$$\hat{\rho} = 1 + 2\hat{\delta}$$

  as a consistent estimator of ρ for $\delta \leq 0$.

Example

Airline Market Competition

Comparison of Estimators

FE versus FD

- Estimates and inference are identical when $T = 2$. Generally, can see differences as T increases.

- Usually think a significant difference signals violation of $\mathrm{Cov}\left(\mathbf{x}_{is}, u_{it}\right) = 0$, all s, t. FE has some robustness if $\mathrm{Cov}\left(\mathbf{x}_{it}, u_{it}\right) = 0$ but $\mathrm{Cov}\left(\mathbf{x}_{it}, u_{is}\right) = 0$, some $s \neq t$: The "bias" is of order 1/T. FD does not average out the bias over T.

- To see this, maintain contemporaneous exogeneity:

$$E\left(\mathbf{x}_{it}' u_{it}\right) = 0.$$

- Generally, under Assumption FE.2, we can write

$$\underset{N \to \infty}{\mathrm{plim}} \left(\hat{\beta}_{FE}\right) = \beta + \left[T^{-1} \sum_{i=1}^{T} E\left(\ddot{\mathbf{x}}_{it}' \ddot{\mathbf{x}}_{it}\right)\right]^{-1} \left[T^{-1} \sum_{i=1}^{T} E\left(\ddot{\mathbf{x}}_{it}' u_{it}\right)\right].$$

Comparison of Estimators

- Under contemporaneous exogeneity,

$$E\left(\ddot{\mathbf{x}}_{it}' u_{it}\right) = -E\left(\overline{\mathbf{x}}_i' u_{it}\right)$$

and so

$$T^{-1} \sum_{t=1}^{T} E\left(\ddot{\mathbf{x}}_{it}' u_{it}\right) = -T^{-1} \sum_{t=1}^{T} E\left(\overline{\mathbf{x}}_i' u_{it}\right) = -E\left(\overline{\mathbf{x}}_i' \bar{u}_i\right).$$

- Under stationarity and weak dependence, $E\left(\overline{\mathbf{x}}_i' \bar{u}_i\right) = O\left(T^{-1}\right)$ because, by the Cauchy-Schwartz inequality, for each j,

$$\left|\text{Cov}\left(\bar{x}_{ij}, \bar{u}_i\right)\right| \leq sd\left(\bar{x}_{ij}\right) sd\left(\bar{u}_i\right)$$

and $sd\left(\bar{x}_{ij}\right), sd\left(\bar{u}_i\right)$ are $O\left(T^{-1/2}\right)$ where each series is weakly dependent.

- Further, $T^{-1} \sum_{t-1}^{T} E\left(\ddot{\mathbf{x}}_{it}' \ddot{\mathbf{x}}_{it}\right)$ is bounded as a function of T. It follows that

$$\operatorname*{plim}_{N\to\infty} \left(\hat{\beta}_{FE}\right) = \beta + O(1) \cdot O\left(T^{-1}\right) = \beta + O\left(T^{-1}\right).$$

Comparison of Estimators

- For the first difference estimator, the general probability limit is

$$\underset{N\to\infty}{\text{plim}}\left(\hat{\beta}_{FD}\right) = \beta + \left[(T-1)^{-1}\sum_{t=2}^{T} E\left(\Delta\mathbf{x}_{it}'\Delta\mathbf{x}_{it}\right)\right]^{-1} \cdot \left[(T-1)^{-1}\sum_{i=2}^{T} E\left(\Delta\mathbf{x}_{it}'\Delta u_{it}\right)\right]$$

- If $\{\mathbf{x}_{it} : t = 1, 2, \ldots\}$ is weakly dependent, so is $\Delta\mathbf{x}_{it}$, and so the first average is generally bounded. (In fact, under stationarity this average does not depend on T.

- As for the second average,

$$E\left(\Delta\mathbf{x}_{it}'\Delta u_{it}\right) = -\left[E\left(\mathbf{x}_{it}'u_{i,t-1}\right) + E\left(\mathbf{x}_{i,t-1}'u_{it}\right)\right]$$

which is constant under stationarity (and generally nonzero). So

$$\underset{N\to\infty}{\text{plim}}\left(\hat{\beta}_{FD}\right) = \beta + O(1)$$

even if $E\left(\mathbf{x}_{i,t-1}'u_{it}\right) = \mathbf{0}$ (so the dynamics given the elements of $\mathbf{x}_{it}$ are correct).

Comparison of Estimators

FE versus RE

- Time-consistent variables drop out of FE estimation. On the time-varying covariates, are FE and RE so different after all? Define the parameter

$$\lambda = 1 - \left[\frac{1}{1 + T\left(\sigma_c^2/\sigma_u^2\right)} \right]^{1/2},$$

  which is consistently estimated (for fixed T) by $\hat{\lambda}$. (Some authors use θ as the symbol.) The, the RE estimate can be obtained from the pooled OLS regression

$$y_{it} - \hat{\lambda}\bar{y}_i \text{ on } \mathbf{x}_{it} - \hat{\lambda}\overline{\mathbf{x}}_i, t = 1, \ldots, T; i = 1, \ldots, N.$$

- Call $y_{it} - \hat{\lambda}\bar{y}_i$ a "quasi-time-demeaned" variable: only a fraction of the mean is removed.

$$\hat{\lambda} \approx 0 \Rightarrow \hat{\beta}_{RE} \approx \hat{\beta}_{POLS}$$

$$\hat{\lambda} \approx 1 \Rightarrow \hat{\beta}_{RE} \approx \hat{\beta}_{FE}$$

  λ increases to unity as (i) σ_c^2/σ_u^2 increases or (ii) T increases. With large T, FE and RE are often similar.

- If $\mathbf{x}_{it}$ includes time-constant variables $\mathbf{z}_i$, then $(1 - \hat{\lambda})\mathbf{z}_i$ appears as a regressor.

Comparison of Estimators

Efficiency of RE

- Can show that RE is asymptotically more efficient than FE under RE.1, RE.2, FE.2, and RE.3. Assume, for simplicity, $\mathbf{x}_{it}$ has all time-varying elements. (See text Section 10.7.2 for more general case.)
- Then

$$Avar\left(\hat{\beta}_{FE}\right) = \sigma_u^2 \left[E\left(\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i\right)\right]^{-1}/N$$

- Let $\breve{\mathbf{x}}_{it} = \mathbf{x}_{it} - \lambda\overline{\mathbf{x}}_i$ be the quasi-time demeaned time-varying covariates. Then

$$Avar\left(\hat{\beta}_{RE}\right) = \sigma_u^2 \left[E\left(\breve{\mathbf{X}}_i'\breve{\mathbf{X}}_i\right)\right]^{-1}/N$$

- Using $\sum_{t=1}^{T} \ddot{\mathbf{x}}_{it} = \mathbf{0}$ we have

$$\begin{aligned}
\breve{\mathbf{X}}_i'\breve{\mathbf{X}}_i &= \sum_{t=1}^{T} \breve{\mathbf{x}}_{it}'\breve{\mathbf{x}}_{it} = \sum_{t=1}^{T} \left[\ddot{\mathbf{x}}_{it} + (1-\lambda)\overline{\mathbf{x}}_i\right]' \left[\ddot{\mathbf{x}}_{it} + (1-\lambda)\overline{\mathbf{x}}_i\right] \\
&= \sum_{t=1}^{T} \left[\ddot{\mathbf{x}}_{it}'\ddot{\mathbf{x}}_{it} + (1-\lambda)^2\overline{\mathbf{x}}_i'\overline{\mathbf{x}}_i\right] \\
&= \ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i + (1-\lambda)^2 T\overline{\mathbf{x}}_i'\overline{\mathbf{x}}_i \\
E\left(\breve{\mathbf{X}}_i'\breve{\mathbf{X}}_i\right) - E\left(\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i\right) &= (1-\lambda)^2 T E\left(\overline{\mathbf{x}}_i'\overline{\mathbf{x}}_i\right)
\end{aligned}$$

which is positive semidefinite.

Hausman Test

Testing the Key RE Assumption

- Recall the key RE assumption is $\operatorname{Cov}(\mathbf{x}_{it}, c_i) = 0$. With lots of good time-constant controls ("observed heterogeneity") might be able to make this condition roughly true.
- Caution: Usual Hausman test maintains RE. 3 - second moment assumptions - yet has no systematic power for detecting violations from this assumption.

Hausman Test

- The original form of the Hausman statistic can be computed as follows. Let $\hat{\boldsymbol{\delta}}_{RE}$ denote the vector of random effects estimates without the coefficients on time-constant variables or aggregate time variables, and let $\hat{\boldsymbol{\delta}}_{FE}$ denote the corresponding fixed effects estimates; let these each be $M \times 1$ vectors. Then

$$H = \left(\hat{\boldsymbol{\delta}}_{FE} - \hat{\boldsymbol{\delta}}_{RE}\right)' \left[\hat{\text{Avar}}\left(\hat{\boldsymbol{\delta}}_{FE}\right) - \hat{\text{Avar}}\left(\hat{\boldsymbol{\delta}}_{RE}\right)\right]^{-1} \left(\hat{\boldsymbol{\delta}}_{FE} - \hat{\boldsymbol{\delta}}_{RE}\right)$$

  is distributed asymptotically as χ^2_M under Assumptions RE.1-RE.3.

- The usual estimators of $\text{Avar}\,(\boldsymbol{\delta}_{FE})$ and $\text{Avar}\,(\boldsymbol{\delta}_{RE})$ can be used in the equation above, but if different estimates of σ_u^2 are used, the matrix $\hat{\text{Avar}}\left(\hat{\boldsymbol{\delta}}_{FE}\right) - \hat{\text{Avar}}\left(\hat{\boldsymbol{\delta}}_{RE}\right)$ need not be positive definite. Thus it is best to use either the fixed effects estimate or the random effects estimate of σ_u^2 in both places.

Ashenfelter and Krueger (1994)

Return to Schooling Redux

Ashenfelter and Krueger (1994)

A general setup specifies wage rates as consisting of an unobservable component that varies by family μ_i, observable components that vary by family, $\mathbf{X}_i$, observable components that vary across individuals, $\mathbf{Z}_{1i}$ and $\mathbf{Z}_{2i}$, and unobservable individual components (ε_{1i} and ε_{2i}). This implies

$$y_{1i} = \boldsymbol{\alpha}\mathbf{X}_i + \boldsymbol{\beta}\mathbf{Z}_{1i} + \mu_i + \varepsilon_{1i} \tag{1}$$

and

$$y_{2i} = \boldsymbol{\alpha}\mathbf{X}_i + \boldsymbol{\beta}\mathbf{Z}_{2i} + \mu_i + \varepsilon_{2i} \tag{2}$$

where we assume that the equations are identical for the two twins. A general representation for the correlation between the family effect and the observables is

$$\mu_i = \gamma\mathbf{Z}_{1i} + \gamma\mathbf{Z}_{2i} + \delta\mathbf{X}_i + \omega_i \tag{3}$$

Ashenfelter and Krueger (1994)

$$y_{1i} = [\boldsymbol{\alpha} + \boldsymbol{\delta}]\,\mathbf{X}_i + [\boldsymbol{\beta} + \boldsymbol{\gamma}]\,\mathbf{Z}_{1i} + \boldsymbol{\gamma}\mathbf{Z}_{2i} + \varepsilon'_{1i} \tag{4}$$

$$y_{2i} = [\boldsymbol{\alpha} + \boldsymbol{\delta}]\,\mathbf{X}_i + \boldsymbol{\gamma}\mathbf{Z}_{1i} + [\boldsymbol{\beta} + \boldsymbol{\gamma}]\,\mathbf{Z}_{2i} + \varepsilon'_{2i} \tag{5}$$

where $\varepsilon'_{1i} = \omega_i + \varepsilon_{1i}$ and $\varepsilon'_{2i} = \omega_i + \varepsilon_{2i}$. Although equations (4) and (5) may be fitted by ordinary least squares (OLS), generalized least squares (GLS) is the optimal estimator for these equations because of the cross-equation restrictions on the coefficients. (Generalized least squares also provides the appropriate estimates of standard errors for the estimated coefficients.)

Ashenfelter and Krueger (1994)

$$y_{1i} - y_{2i} = \boldsymbol{\beta} \left(\mathbf{Z}_{1i} - \mathbf{Z}_{2i} \right) + \varepsilon_{1i} - \varepsilon_{2i} \tag{6}$$

In (6) the individual effect μ_i has been removed. The least-squares estimator for this equation is called the "fixed-effects" estimator. In equations (4) and (5) the selection effect is estimated explicitly and then subtracted to obtain the structural estimate of the return to schooling. In (6) the selection effect is eliminated by differencing.

Ashenfelter and Krueger (1994)

In the presence of selection effects, however, the ordinary least-squares estimator will be biased even in the absence of measurement error (because of the omitted sibling's schooling variable).
The fixed-effects estimator eliminates this selection (or "omitted variable") bias, but it does so at the expense of introducing far greater measurement-error bias. In the presence of classical measurement error (see Zvi Griliches, 1979), the probability limit of the fixed-effects estimator, $\hat{\beta}_{\mathrm{FE}}$, is

$$\beta_{\mathrm{FE}}\left(1-\frac{\mathrm{Var}(\nu)}{[\mathrm{Var}(\nu)+\mathrm{Var}(S)]\,(1-\rho_{\mathrm{s}})}\right)$$

Ashenfelter and Krueger (1994)

A straightforward consistent estimator for equation (4), (5), or (6), assuming classical measurement error, may be obtained by the method of instrumental variables using the independent measures of the schooling variables as instruments. For example, we may fit

$$y_{1i} - y_{2i} = \beta \left(S_1^1 - S_2^2 \right) + \varepsilon_{1i} - \varepsilon_{2i} = \beta \Delta S' + \Delta \varepsilon \tag{7}$$

using $\Delta S'' = \left(S_1^2 - S_2^1 \right)$ as an instrument for $\Delta S'$.

Ashenfelter and Krueger (1994)

Table 3—Ordinary Least-Squares (OLS), Generalized Least-Squares (GLS), Instrumental-Variables (IV), and Fixed-Effects Estimates of Log Wage Equations for Identical Twins[a]

Variable	OLS (i)	GLS (ii)	GLS (iii)	IV[a] (iv)	First difference (v)	First difference by IV (vi)
Own education	0.084	0.087	0.088	0.116	0.092	0.167
	(0.014)	(0.015)	(0.015)	(0.030)	(0.024)	(0.043)
Sibling's education	—	—	−0.007	−0.037	—	—
			(0.015)	(0.029)		
Age	0.088	0.090	0.090	0.088	—	—
	(0.019)	(0.023)	(0.023)	(0.019)		
Age squared (÷100)	−0.087	−0.089	−0.090	−0.087	—	—
	(0.023)	(0.028)	(0.029)	(0.024)		
Male	0.204	0.204	0.206	0.206	—	—
	(0.063)	(0.077)	(0.077)	(0.064)		
White	−0.410	−0.417	−0.424	−0.428	—	—
	(0.127)	(0.143)	(0.144)	(0.128)		
Sample size:	298	298	298	298	149	149
R^2:	0.260	0.219	0.219	—	0.092	—

Notes: Each equation also includes an intercept term. Numbers in parentheses are estimated standard errors.

[a]Own education and sibling's education are instrumented for using each sibling's report of the other sibling's education as instruments.

The Generalized IV Estimator

Derivation of the GIV Estimator and Its Asymptotic Properties

- Rather than estimating β using the moment conditions $E\left[\mathbf{Z}_i'\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right)\right] = \mathbf{0}$, an alternative is to transform the moment conditions in a way analogous to generalized least squares. Let $\Omega = E\left(\mathbf{u}_i\mathbf{u}_i'\right)$ as before, and assume Ω is positive definite. Consider the transformed equation

$$\Omega^{-1/2}\mathbf{y}_i = \Omega^{-1/2}\mathbf{X}_i\beta + \Omega^{-1/2}\mathbf{u}_i.$$

- Apply the system 2SLS estimator to the equation above with instruments $\Omega^{-1/2}\mathbf{Z}_i$.

$$\hat{\boldsymbol{\beta}}_{GV} = \left[\left(\sum_{i=1}^{N}\mathbf{X}_i'\Omega^{-1}\mathbf{Z}_i\right)\left(\sum_{i=1}^{N}\mathbf{Z}_i'\Omega^{-1}\mathbf{Z}_i\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{Z}_i'\Omega^{-1}\mathbf{X}_i\right)\right]^{-1}$$
$$\cdot\left(\sum_{i=1}^{N}\mathbf{X}_i'\Omega^{-1}\mathbf{Z}_i\right)\left(\sum_{i=1}^{N}\mathbf{Z}_i'\Omega^{-1}\mathbf{Z}_i\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{Z}_i'\Omega^{-1}\mathbf{y}_i\right)$$

The Generalized IV Estimator

Assumption GIV. 1 (Exogeneity): $E\left(\mathbf{Z}_i \otimes \mathbf{u}_i\right) = \mathbf{0}$.

- GIV. 1 implies
$$E\left(\mathbf{Z}_i'\boldsymbol{\Omega}^{-1}\mathbf{u}_i\right) = \mathbf{0}$$
- Assumption GIV. 1 is identical to the consistency condition for GLS when $\mathbf{Z}_i = \mathbf{X}_i$.

Naturally, we also need a rank condition:
Assumption GIV. 2 (Rank Condition): (a) $\operatorname{rank} E\left(\mathbf{Z}_i'\boldsymbol{\Omega}^{-1}\mathbf{Z}_i\right) = L$; $\operatorname{rank} E\left(\mathbf{Z}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i\right) = K$.

- When Ω is replaced with a consistent estimator, $\hat{\Omega}$, we obtain the **generalized instrumental variables (GIV) estimator**.

RE and FE Instrumental Variable Methods

RE and FE Instrumental Variable Methods

We start with the usual unobserved effects model,

$$y_{it} = \mathbf{x}_{it}\beta + c_i + u_{it}, t = 1, \ldots, T,$$

but now we think some elements of $\mathbf{x}_{it}$ are correlated with u_{it} (or maybe even with u_{ir} for $r \neq t$). Let $\mathbf{z}_{it}$ be a set of $1 \times L$ (possible) instrumental variables, $L \geq K$. (Intercept in $\mathbf{x}_{it}$ so $E(c_i) = 0$ can be assumed.)

REIV

Assumption REIV.1:

$$(a) E\left(u_{it} \mid \mathbf{z}_{it}, c_i\right)=0, t=1, \ldots, T$$
$$(b) E\left(c_i \mid \mathbf{z}_{it}\right)=0$$

- For simplicity, assumes that $\mathbf{x}_{it}$ contains an overall intercept (and probably a separate intercept in each time period), so we can take $E\left(c_i\right)=0$.
- As usual, we could relax the assumptions to zero correlation without changing consistency.
- Define $\Omega=\operatorname{Var}\left(\mathbf{v}_i\right)$, where $\mathbf{v}_i=c_i \mathbf{j}_T+\mathbf{u}_i$.
- Let $\mathbf{X}_i$ be $T \times K$ and $\mathbf{Z}_i$ be $T \times L$.

Assumption REIV.2: Ω is nonsingular, and

$$(a) \operatorname{rank} E\left(\mathbf{Z}_i^{\prime} \Omega^{-1} \mathbf{Z}_i\right)=L$$
$$(b) \operatorname{rank} E\left(\mathbf{Z}_i^{\prime} \Omega^{-1} \mathbf{X}_i\right)=K$$

- This is just the usual rank condition for GIV estimation.
- The REIV estimator is just the GIV estimator where Ω is assumed to have the RE form.
- Without further assumptions, fully robust inference is warranted, as usual.

REIV

Assumption REIV.3:

$$(a) E\left(\mathbf{u}_i \mathbf{u}_i' \mid \mathbf{z}_i, c_i\right) = \sigma_u^2 \mathbf{I}_T$$
$$(b) E\left(c_i^2 \mid \mathbf{z}_i\right) = \sigma_c^2$$

- Under REIV.3, the nonrobust variance matrix estimator is valid:

$$\left[\left(\sum_{i=1}^{N} \mathbf{X}_i' \hat{\mathbf{\Omega}}^{-1} \mathbf{Z}_i\right)\left(\sum_{i=1}^{N} \mathbf{Z}_i' \hat{\mathbf{\Omega}}^{-1} \mathbf{Z}_i\right)^{-1}\left(\sum_{i=1}^{N} \mathbf{Z}_i' \hat{\mathbf{\Omega}}^{-1} \mathbf{X}_i\right)\right]^{-1}$$

  where $\hat{\mathbf{\Omega}}$ has the RE structure.

- In Stata, the command "xtivreg" with the "re" option produces this estimator and the nonrobust variance matrix estimator.
- The REIV estimator is also called the random effects 2SLS estimator.

REIV

With REIV, can have time-constant explanatory variables and time-constant instruments. With lots of good controls, or an exogenous intervention in an initial time period, the analysis can be convincing. But time-constant IVs in panel data are often unconvincing.

FEIV

- A more robust analysis uses fixed effects and instrumental variables (FEIV). This requires time-varying instruments.

Assumption FEIV.1: Same as REIV.1(a):

$$E\left(u_{it} \mid \mathbf{z}_i, c_i\right) = 0, t = 1, \ldots, T.$$

- Now apply pooled 2SLS to the time-demeaned equation:

$$y_{it} - \bar{y}_i = \left(\mathbf{x}_{it} - \overline{\mathbf{x}}_i\right)\beta + \left(u_{it} - \bar{u}_i\right)$$

  using instruments $(\mathbf{z}_{it} - \overline{\mathbf{z}}_i)$.

- This can be very convincing: the IVs can be arbitrarily correlated with c_i as long as there is exogenous time variation in the instruments.

Assumption FEIV.2:

$$(a)\,\text{rank}\,E\left(\ddot{\mathbf{Z}}_i'\ddot{\mathbf{Z}}_i\right) = L$$

$$(b)\,\text{rank}\,E\left(\ddot{\mathbf{Z}}_i'\ddot{\mathbf{X}}_i\right) = K$$

FEIV

- As usual, make inference fully robust to serial correlation and heteroskedasticity in, unless the following assumption holds:

Assumption FEIV.3: Same as REIV.3(a), that is,

$$E\left(\mathbf{u}_i \mathbf{u}_i' \mid \mathbf{z}_i, c_i\right) = \sigma_u^2 \mathbf{I}_T$$

- Under FE.1, FE.2, and FE.3, the asymptotic variance matrix of $\hat{\beta}_{FEIV}$ is estimated as

$$\hat{\sigma}_u^2 \left[\left(\sum_{i=1}^{N} \ddot{\mathbf{X}}_i' \ddot{\mathbf{Z}}_i\right)\left(\sum_{i=1}^{N} \ddot{\mathbf{Z}}_i' \ddot{\mathbf{Z}}_i\right)^{-1}\left(\sum_{i=1}^{N} \ddot{\mathbf{Z}}_i' \ddot{\mathbf{X}}_i\right)\right]^{-1}$$

where

$$\hat{\sigma}_u^2 = [N(T-1)-K]^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} \widehat{\ddot{u}}_{it}^2\right)$$

Moving Ahead

Putting panel data into a dynamic and linear GMM framework - relax strict exogeneity