Panel Data

Suppose we observe units $i=1,\dots,N$ over time periods $t=1,\dots,T$. We will assume individuals are sampled iid in the cross-section but allow for serial correlation.

Specify the linear model $Y_{it} = X_{it}'\beta + u_{it}$.

Have $N \cdot T$ observations. Could construct "pooled OLS" estimator
↳ Pooling across time.

$$\hat{\beta}^{OLS} = \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}'\right)^{-1} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}Y_{it}.$$

Not necessarily helpful if $E(V_{it} \mid X_{i1},\dots,X_{iT}) \neq 0$.

We will study small T large N asymptotics.

eg. $T=2$: Panel structure allows us to deal with certain types of endogeneity without an IV:

Let $u_{it} = \mu_i + v_{it}$, where μ_i can have arbitrary correlation with X_{it}.

$$Y_{it} = X_{it}'\beta + \mu_i + v_{it}.$$

First Difference: $Y_{i2} - Y_{i1} = (X_{i2}-X_{i1})'\beta + \underbrace{\cancel{\mu_i - \mu_i}}_{=0} + v_{i2} - v_{i1}$.

FD estimator of β: OLS applied to differenced model:

$$\hat{\beta}^{FD} = \left(\frac{1}{N}\sum_{i=1}^{N}(X_{i2}-X_{i1})(X_{i2}-X_{i1})'\right)^{-1} \frac{1}{N}\sum (X_{i2}-X_{i1})(y_{i2}-y_{i1}).$$

$$\hat{\beta}^{FD} - \beta = \left(\frac{1}{N}\sum_{i=1}^{N}(x_{i2}-x_{i1})(x_{i2}-x_{i1})'\right)^{-1}\frac{1}{N}\sum(x_{i2}-x_{i1})(v_{i2}-v_{i1})$$

$\hat{\beta}^{FD}$ is consistent for β by LLN provided $E\left((x_2-x_1)(v_2-v_1)\right)=0$.

$\Longleftrightarrow E(x_2v_2) + E(x_1v_1) - E(x_2v_1) - E(x_1v_2) = 0$.

Need not just contemporaneous exogeneity eg. $E(x_tU_t)=0$ or $E(U_t|x_t)=0$,

but <u>strict</u> exogeneity - Unobservables in other time periods are uncorrelated with today's regressors. Violated if x_2 decided partly on the basis of U_1.

General T: $Y_{it} = x_{it}'\beta + \mu_i + u_{it}$

Define $\Delta W_{it} = W_{it} - W_{it-1}$ for a panel $\{W_{it}\}_{i=1,\dots,N,\ t=1,\dots,T}$.

$$\Delta Y_{it} = \Delta x_{it}'\beta + \Delta u_{it}.$$

FD (1) $E(U_{it}|x_{i1},\dots,x_{iT}) = 0$. for all t. $\Rightarrow E(u_{it}x_{is})=0 \ \forall t,s$.

$\Rightarrow E(\Delta u_{it}|\Delta x_{it})=0$.

(2) $\sum_{t=2}^{T} E(\Delta x_{it}\Delta x_{it}')$ is invertible.

Why (2)? $\hat{\beta}^{FD} = \left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta x_{it}\Delta x_{it}'\right)^{-1}\frac{1}{N}\sum\sum\Delta x_{it}\Delta Y_{it}$

$$\hat{\beta}^{FD} - \beta = (\ \cdot\ )^{-1}\frac{1}{N}\sum\sum\Delta x_{it}\Delta u_{it}.$$

T is fixed, so $\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{t=2}^{T}\Delta X_{it}\Delta X_{it}'\right) \to^{p} E\left(\sum_{t=2}^{T}\Delta X_{it}\Delta X_{it}'\right)$

$= \sum_{t=2}^{T} E(\Delta X_{it}\Delta X_{it}')$.

must be invertible.

Under these conditions, $\hat{\beta}^{FD}$ is consistent

because $E\left(\sum_{t=2}^{T}\Delta X_{it}\Delta U_{it}\right) = 0$.

Within Estimator. Let $\dot{X}_{it} = X_{it} - \frac{1}{T}\sum_{t=1}^{T} X_{it}$

$= X_{it} - \bar{X}_i$

$\dot{Y}_{it} = Y_{it} - \frac{1}{T}\sum_{t=1}^{T} Y_{it} = Y_{it} - \bar{Y}_i$.

From $Y_{it} = X_{it}'\beta + \mu_i + U_{it}$, get

Note: $\mu_i - \frac{1}{T}\sum_{t=1}^{T}\mu_i = 0$.

$\dot{Y}_{it} = \dot{X}_{it}'\beta + \dot{U}_{it}$

FE (1) $E(U_{it} \mid X_{i1}, \ldots, X_{iT}) = 0$ for all $t = 1, \ldots, T$.

$\Rightarrow E(\dot{U}_{it} \mid \dot{X}_{it}) = 0$.

FE (2): $\sum_{t=1}^{T} E(\dot{X}_{it}\dot{X}_{it}')$ is invertible. (Again look at form of estimator).

$$\hat{\beta}^{FE} - \beta = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{X}_{it}\dot{X}_{it}'\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{X}_{it}\dot{U}_{it}\right),$$

Note: $\sum_t \dot{X}_{it}\dot{U}_{it} = \sum_t \dot{X}_{it}U_{it} - \bar{U}_i\sum_t \dot{X}_{it}$

$= \sum_t \dot{X}_{it}U_{it}$.

$\sum_t \dot{X}_{it} = \sum_t X_{it} - \sum_t \frac{1}{T}\sum X_{it}$.

$= T\bar{X}_i - T\bar{X}_i = 0$.

$$\ddot{Y}_{it} = \ddot{X}_{it}'\beta + \ddot{u}_{it}$$

Call $\ddot{X}_i = (\ddot{x}_{i1}, \ddot{x}_{i2}, \dots, \ddot{x}_{iT})'$ this is $T \times k$.

$U_i = (U_{i1}, \dots, U_{iT})' \in \mathbb{R}^T$ (or $T \times 1$).

So $\sum_{t=1}^{T} \ddot{x}_{it}\ddot{x}_{it}' = \ddot{X}_i'\ddot{X}_i$.

$$\sqrt{n}\,(\hat{\beta}^{FE} - \beta) = \underbrace{\left(\frac{1}{N}\sum_{i=1}^{N} \ddot{X}_i'\ddot{X}_i\right)^{-1}}_{\downarrow p \;\; \Sigma_X = E(\ddot{X}_i'\ddot{X}_i) \;\; (\text{as } N\to\infty)} \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{N} \ddot{X}_i U_i}_{\downarrow d \;\; N(0,\, E(\ddot{X}_i' U_i U_i' \ddot{X}_i))}$$

$$\uparrow \Omega.$$

$$\Rightarrow \sqrt{n}(\hat{\beta}^{FE} - \beta) \xrightarrow{d} N(0, \Sigma_X^{-1}\Omega\Sigma_X^{-1}).$$

Assuming homoskedasticity and no serial correlation gives.
$E(U_iU_i' \mid \ddot{X}_i) = \sigma^2 I_T$ then $Var = \sigma^2 E(\ddot{X}_i'\ddot{X}_i)^{-1}$.

More common approach is to use a generalization of heteroskedasticity robust standard errors.

$$\hat{V}^{fe} = \left(\frac{1}{N}\sum \ddot{X}_i'\ddot{X}_i\right)^{-1}\left(\frac{1}{N}\sum \ddot{X}_i'\hat{U}_i\hat{U}_i'\ddot{X}_i\right)\left(\frac{1}{N}\sum \ddot{X}_i'\ddot{X}_i\right)^{-1}.$$

$\xrightarrow{p} \Sigma_X^{-1}\Omega\Sigma_X^{-1}$, which is fully robust to heterosked. and serial correlation. Stata computes this if the cluster(id) option is specified in xtreg. This is exactly the cluster covariance estimator we computed in the DiD context allowing for arbitrary correlation across time but requiring zero correlation across individuals.

Clustering at individual level.

Fixed effects more efficient than FD under homoskedasticity and no serial correlation: Intuition: Taking 1st differences causes serial correlation in Δu_{it} since

$$E(\Delta u_{it}\Delta u_{it-1}) = E(u_{it}u_{it-1} - u_{it-1}^2 - u_{it}u_{it-2} + u_{it-1}u_{it-2})$$

(no serial corr. in u_{it}) $= -Var(u_{it-1})$.

On the other hand, if u_{it} follows a random walk:

$u_{it} = u_{it-1} + v_{it}$, where v_{it} is iid, then:

$\Delta u_{it} = v_{it}$, so FD becomes more efficient.

These results rely on homoskedasticity, so often best to use a robust standard error and pay less attention to efficiency concerns.

When $T=2$: $\hat{\beta}^{FD} = \hat{\beta}^{FE}$:

$$\ddot{x}_{it} = x_{it} - \left(\frac{x_{i1}+x_{i2}}{2}\right) = \frac{1}{2}(x_{it} - x_{i-t})$$

$-t$ means other time period.

So $$\sum_{t=1}^{2} \ddot{x}_{it}\ddot{x}_{it}' = \frac{1}{4}\left[(x_{i1}-x_{i2})(x_{i1}-x_{i2})' + (x_{i2}-x_{i1})(x_{i2}-x_{i1})'\right]$$

$$= \frac{1}{2}(x_{i2}-x_{i1})(x_{i2}-x_{i1})'.$$

$$\sum_{t=1}^{2} \ddot{x}_{it} \ddot{y}_{it} = \frac{1}{4}\left[(x_{i1}-x_{i2})(y_{i1}-y_{i2}) + (x_{i2}-x_{i1})(y_{i2}-y_{i1}) \right]$$

$$= \frac{1}{2}(x_{i2}-x_{i1})(y_{i2}-y_{i1}).$$

Put together into $\hat{\beta}^{FE} = \left(\frac{1}{N}\sum_{i=1}^{N}(x_{i2}-x_{i1})(x_{i2}-x_{i1})'\right)^{-1} \frac{1}{N}\sum_{i=1}^{N}(x_{i2}-x_{i1})(y_{i2}-y_{i1})$

$$= \hat{\beta}^{FD}.$$

Random Effects

Terminology fixed vs. random effects has nothing to do with whether μ_i is random (which it is because we are sampling individuals at random), but it is to with the relationship between μ_i and x_{it}. Fixed effects / FD causes us to lose N data points to get rid of μ_i, but RE approaches model the behaviour of μ_i and estimate the system via GLS to produce a more efficient estimator.

RE ① $E(\mu_i \mid x_{i1}, \dots, x_{iT}) = 0.$

All time invariant factors differenced out in FE approach are mean independent of x_{it} anyway at all time periods.

Together with FE①, error is $v_{it} = \mu_i + u_{it}$ now satisfies $E(v_{it} \mid x_{i1}, \dots, x_{iT}) = 0. \quad \forall t.$

$$x_i = (x_{i1}, \dots, x_{iT})'$$

Idea: Exploit serial correlation in V_{it} generated by common μ_i component in each time period.

RE ② (i) $Var(u_{it} | X_i) = \sigma_u^2$ (ii) $Var(\mu_i | X_i) = \sigma_\mu^2$.

(iii) $E(U_{it} U_{is} | X_i) = 0$ $\forall t \neq s$.

(iv) $E(U_{it} \mu_i | X_i) = 0$ $\forall t = 1, \dots, T$.

Then $Var(V_{it} | X_i) = E(\mu_i^2 + U_{it}^2 + 2\mu_i U_{it} | X_i)$

$= \sigma_\mu^2 + \sigma_u^2$.

$E(V_{it} V_{is} | X_i) = E(\mu_i^2 + U_{it} U_{is} + \mu_i U_{it} + \mu_i U_{is} | X_i)$

$= \sigma_\mu^2$.

T×T matrix of ones.

$$E(V_i V_i' | X) = \sigma_u^2 I_T + \sigma_\mu^2 \cdot 1_T 1_T' = \Omega.$$

Then use GLS: $\hat\beta^{GLS} = \hat\beta^{RE} = \left(\sum_{i=1}^n X_i' \Omega^{-1} X_i\right)^{-1} \sum_{i=1}^n X_i' \Omega^{-1} Y_i$.

Stack over t:

$\longrightarrow Y_i = X_i' \beta + V_i$

$\rightarrow \Omega^{-1/2} Y_i = \Omega^{-1/2} X_i' \beta + \Omega^{-1/2} V_i$.

Can estimate σ_u^2, σ_μ^2 and perform Feasible GLS:

OLS on $\hat{\Omega}^{-1/2} Y_i = \hat{\Omega}^{-1/2} X_i' \beta + \hat{\Omega}^{-1/2} V_i$.

① Dealing with $E(\mu_i | X_i) \neq 0$ with differencing is one of the attractive properties of having panel data. It means we can remove sources of endogeneity that would otherwise be caused by omitting time invariant regressors.

② Efficiency gains rely on homoskedasticity.

③ RE allows us to estimate coefficients on time-invariant covariates.

④ RE1, RE2 are enough to identify β in a single cross-section. Time dimension needed to estimate σ_u^2, σ_μ^2 separately because contemporaneous error covariance $Var(V_{it}|X_i) = \sigma_\mu^2 + \sigma_u^2$.

Hausman spec. test of RE assumptions notes that under RE assumptions (which we added to FE assumptions), both $\hat{\beta}^{RE}$ and $\hat{\beta}^{FE}$ are consistent, but only $\hat{\beta}^{RE}$ is efficient. Under H_1: RE is inconsistent.
Considering $\hat{\beta}^{RE}$ vs. $\hat{\beta}^{FE}$ on the basis of the outcome of such a test leads to an estimator whose distribution is quite non-normal. Pre-testing leads to undesirable statistical properties.
(Compare with checking for weak instruments as

a basis for using GMM).

Can handle time trends: $Y_{it} = X_{it}'\beta + \mu_i + \gamma_t + \varepsilon_{it}$:

Use double-demeaning estimator.

$$Y_{it} - \bar{Y}_i = (X_{it} - \bar{X}_i)'\beta + \gamma_t - \bar{\gamma} + \varepsilon_{it} - \bar{\varepsilon}_i$$

Cross-sectional avg. of Y_{it} $\longrightarrow$ $\bar{Y}_t = \bar{X}_t'\beta + \gamma_t + \bar{\mu} + \bar{\varepsilon}_t$

Cross-section + time avg. of Y_{it} $\longrightarrow$ $\bar{Y} = \bar{X}'\beta + \bar{\gamma} + \bar{\mu} + \bar{\varepsilon}$

$$\bar{Y}_t - \bar{Y} = (\bar{X}_t - \bar{X})'\beta + \gamma_t - \bar{\gamma} + \bar{\varepsilon}_t - \bar{\varepsilon}$$

$$(Y_{it} - \bar{Y}_i) - (\bar{Y}_t - \bar{Y}) = (X_{it} - \bar{X}_i - (\bar{X}_t - \bar{X}))'\beta + \varepsilon_{it} - \bar{\varepsilon}_i - (\bar{\varepsilon}_t - \bar{\varepsilon}).$$

$$\ddot{Y}_{it} = \ddot{X}_{it}'\beta + \ddot{\varepsilon}_{it} \qquad E(\varepsilon_{it} | X) = 0.$$

OLS estimator of β is consistent if $E(\varepsilon_{it} | X_{i1}, \dots, X_{iT}) = 0$ $\forall$ time periods $t = 1, \dots, T$ and $\sum_{t=1}^{T} E(\ddot{X}_{it}\ddot{X}_{it}')$ is invertible.

Now requires X to be both time-varying and unit-varying, but exogeneity condition is intuitively weaker because ε_{it} now contains only unobserved factor that vary with both i and t.

$$y_{it} = \alpha_t + \gamma_g + \pi D .$$

Common trend

Group specific mean of $y(0)$

GMM Estimation of Panel models

Write model as $Y_{it} = X_{it}'\beta + U_{it}$. Stack across time:

$$\underset{T\times 1}{Y_i} = X_i \beta + U_i \quad \text{where } X_i = \begin{pmatrix} x_{i1}' \\ \vdots \\ x_{iT}' \end{pmatrix}$$

Let $Z_i \in \mathbb{R}^{T\times L}$ be a matrix of instruments satisfying $E(Z_i' U_i) = 0$. This gives L moment conditions. For now assume none of these moments are redundant.

Again $C Z_i' Y_i = C Z_i' X_i \beta + C Z_i' U_i$

$$\beta = E(C Z_i' X_i)^{-1} E(C Z_i' Y_i) .$$

Optimal GMM: $C^* = E(Z_i' X_i)' \Omega^{-1}$, where

$$\Omega = E(Z_i' U_i U_i' Z_i)$$

Feasible optimal GMM estimator uses a consistent estimate of Ω to form:

$$\hat{\beta}^{FGMM} = \left[\left(\sum X_i' Z_i\right)\hat{\Omega}^{-1}\left(\sum Z_i' X_i\right)\right]^{-1} \sum X_i' Z_i \,\hat{\Omega}^{-1} \sum Z_i' Y_i$$

Need $\hat{\Omega} = \frac{1}{N} \sum Z_i' \hat{U}_i \hat{U}_i' Z_i$

Can find $\hat{U}_i$ using 2SLS instead of GMM as a first step.

Suppose we have homoskedasticity + no serial correlation.

$$E(U_i U_i' | Z_i) = \sigma^2 I \quad \Rightarrow \Omega = \sigma^2 E(Z_i' Z_i)$$

Replace $\hat{\Omega}^{-1}$ with $\left(\sigma^2 \frac{1}{N} \sum Z_i' Z_i\right)^{-1}$

This yields 2SLS estimator with $\left(\frac{1}{N} \sum Z_i' Z_i\right)^{-1}$ in place of $\hat{\Omega}^{-1}$.

Example: $Z_i = X_i$ (if there is no endogeneity concern).

$$\hat{\beta}^{FGMM} = \left(\left(\sum X_i' X_i\right)\hat{\Omega}^{-1}\left(\sum X_i' X_i\right)\right)^{-1} \sum X_i' X_i \,\hat{\Omega}^{-1} \sum X_i' Y_i$$

$(ABC)^{-1} = C^{-1} B^{-1} A^{-1}$

$$= \left(\sum X_i' X_i\right)^{-1} \hat{\Omega} \underbrace{\cancel{\left(\sum X_i' X_i\right)^{-1}\left(\sum X_i' X_i\right)}}_{I} \hat{\Omega}^{-1} \sum X_i Y_i$$

$$= \left(\sum X_i' X_i\right)^{-1} \sum X_i Y_i = \left(\sum_i \sum_t x_{it} x_{it}'\right)^{-1} \sum \sum x_{it} y_{it}$$

= Pooled OLS.

Note: We are using sums across time as moments, but nothing in the notation prevents us from using time moments separately. For example: β is identified by a single cross section at time t assuming $E(x_{it}u_{it}) = 0$:

$$x_{it} y_{it} = x_{it} x_{it}' \beta + x_{it} u_{it} \Rightarrow \beta = E(x_{it}x_{it}')^{-1} E(x_{it}y_{it})$$

To use only these moments we specify Z_i as

$$Z_i^t = \begin{bmatrix} 0 \\ x_{it}' \\ 0 \end{bmatrix} \quad U_i = \begin{bmatrix} u_{i1} \\ \vdots \\ \vdots \\ u_{iT} \end{bmatrix} \quad Z_i^{t\prime} U_i = \begin{bmatrix} 0 & x_{it} & 0 \end{bmatrix} \begin{bmatrix} u_{i1} \\ \vdots \\ \vdots \\ u_{iT} \end{bmatrix} = x_{it} u_{it}.$$

gives K moments since $K = \dim(x_{it})$

$$E(Z_i^{t\prime} U_i) = 0$$

$$\Leftrightarrow \quad E(x_{it} u_{it}) = 0. \quad K \text{ moment conditions.}$$

Typically we allow for presence of μ_i in error term and difference it out:

$$\dot{y}_{it} = \dot{x}_{it}' \beta + \dot{u}_{it}, \text{ stack over } T:$$

$$\dot{Y}_i = \dot{X}_i' \beta + \dot{U}_i.$$

$$E(\dot{U}_{it} \dot{x}_{it}) = 0.$$

$$Z_i = [\dot{z}_i^1, \dot{z}_i^2, \ldots, \dot{z}_i^T] \qquad E(Z_i' \dot{U}_i) = 0.$$

$$= \begin{bmatrix} \dot{x}_{i1}' & 0 & \cdots & 0 \\ 0 & \dot{x}_{i2}' & \cdots & 0 \\ \vdots & \vdots & \vdots & 0 \\ 0 & 0 & 0 & \dot{x}_{iT}' \end{bmatrix}$$

$$Z_i' \dot{U}_i = \begin{bmatrix} \dot{x}_{i1} & 0 & & 0 \\ 0 & \dot{x}_{i2} & \cdots & 0 \\ \vdots & \vdots & & 0 \\ 0 & 0 & & \dot{x}_{iT} \end{bmatrix} \begin{bmatrix} \dot{U}_1 \\ \vdots \\ \dot{U}_T \end{bmatrix} = \begin{bmatrix} \dot{x}_{i1} \dot{U}_1 \\ \dot{x}_{i2} \dot{U}_2 \\ \vdots \\ \dot{x}_{iT} \dot{U}_T \end{bmatrix} \quad KT \times 1$$

In principle we have KT moment conditions, but effectively we have $K(T-1)$ moment conditions because de-meaning / first differencing effectively reduces the number of time periods in the data by 1.

The 2SLS estimator using all the instruments in Z_i is just OLS (applied to within-transformed model) because the moments $E(\dot{X}_i' \dot{U}_i) = 0$ are included in this larger set of moments.

$$E(\dot{X}_i' \dot{U}_i) = \sum_{t=1}^{T} E(\dot{x}_{it} \dot{U}_{it}) = 0$$

So the fixed effects estimator is the 2SLS estimator with Z_i as the matrix of instruments.

$$\hat{\beta}^{FGMM} = \left\{ \left[\left(\sum \dot{X}_i' Z_i\right) \hat{\Omega}^{-1} \left(\sum Z_i' \dot{X}_i\right) \right] \right\}^{-1} \left(\sum \dot{X}_i' Z_i\right) \hat{\Omega}^{-1} \sum Z_i' \dot{Y}_i$$

① Relevance: $E(\dot{X}_i' Z_i)$ must have rank K.

② Exogeneity: $E(Z_i' \dot{U}_i) = 0$.

$$(U_{it} = \mu_i + v_{it}).$$

Exogeneity requires that $E(Z_i'(v_{it} - \bar{v}_i)) = 0$. A sufficient condition is that $E(Z_i' v_{is}) = 0 \quad \forall s = 1, \dots, T$.

First-Differencing allows us to avoid this:

$$Y_{it} - Y_{i(t-1)} = (X_{it} - X_{it-1})'\beta + v_{it} - v_{i(t-1)}.$$

ΔY_{it} – stack across time?

$$\tilde{Y}_i = \tilde{X}_i' \beta + \tilde{V}_i \leftarrow (T-1) \times 1.$$

Now $\tilde{v}_{it}$ only depends on 2 time periods, so we can use assumptions such as $E[Z_{is} \cdot v_{it}] = 0$ for $s < t$ for identification without requiring that future realizations of

Z are uncorrelated with today's error.

For example, if z_{it} is scalar, $\beta \in \mathbb{R}^K$, and we have 5 time periods, we could use:

$$E\left(\begin{pmatrix} 0 & z_{i1} & 0 & 0 \\ 0 & 0 & z_{i2} & 0 \\ 0 & 0 & z_{i1} & 0 \\ 0 & 0 & 0 & z_{i3} \\ 0 & 0 & 0 & z_{i2} \\ 0 & 0 & 0 & z_{i1} \end{pmatrix}\begin{pmatrix} \tilde{v}_{i2} \\ \tilde{v}_{i3} \\ \tilde{v}_{i4} \\ \tilde{v}_{i5} \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

This system is over-identified if $K<6$.

<u>Dynamic Models</u>: Panel Data allows us to model dynamic relationships such as

$$Y_{it} = \rho Y_{i(t-1)} + X_{it}'\beta + \mu_i + V_{it},$$

where we continue to assume either $E(V_{it} \mid X_{i1}, \dots, X_{iT}) = 0$

or $E(V_{it} X_{it}) = 0$.

Note: $Y_{i(t-1)}$ is generally correlated with μ_i $\because$

$$Y_{i(t-1)} = \rho Y_{i(t-2)} + X_{i(t-1)}'\beta + \mu_i + V_{i(t-1)}.$$

Suppose $T=3$: Fixed effects estimator will be the same as FD estimator:

After differencing out μ_i we get,

$$Y_{i3} - Y_{i2} = \rho(Y_{i2} - Y_{i1}) + (X_{i3} - X_{i2})'\beta + v_{i3} - v_{i2}.$$

Need: $E((Y_{i2} - Y_{i1})(v_{i3} - v_{i2})) = 0.$

(at a minimum, with no regressors)

$$= \underbrace{E(Y_{i2} v_{i3})}_{0} - E(Y_{i2} v_{i2})$$

$$- \underbrace{E(Y_{i1} v_{i3})}_{0} + \underbrace{E(Y_{i1} v_{i2})}_{0}$$

Typically we assume Dynamic Completeness:

$$E(v_{it} \mid Y_{i(t-1)}, Y_{i(t-2)}, \dots) = 0.$$

$$= - E(Y_{i2} v_{i2})$$

$$= - E((\rho Y_{i1} + \mu_i + v_{i2}) v_{i2})$$

$$= -\sigma_v^2$$

The problem arose because of the dynamic dependence of the outcome together with time-invariant factors μ_i. When getting rid of μ_i, because past outcomes affect future outcomes, we end-up with endogeneity from differenced outcomes being correlated with differenced errors.

One solution: Anderson-Hsiao (1988):

Dynamic completeness $\Rightarrow$ $Cov(Y_{i1}, v_{i3}-v_{i2}) = 0$.

So $E(Y_{i1}(v_{i3}-v_{i2})) = 0$.

Validity Condition.

$$E(Y_{i1}(Y_{i2}-Y_{i1})) = (\rho-1)E(Y_{i1}^2 + \cdots) \neq 0.$$

Relevance Condition. (Drop Covariates).

Construct IV estimator: (Do this for $T=3$):

$$\hat{\beta}^{AH} = \left(\sum_{i=1}^{N} Y_{i1}(Y_{i2}-Y_{i1})\right)^{-1}\left(\sum_{i=1}^{N} Y_{i1}(Y_{i3}-Y_{i2})\right)$$

$$(Z'X)^{-1} \qquad Z'Y$$

$$= \rho + \left(\sum Y_{i1}(Y_{i2}-Y_{i1})\right)^{-1}\left(\sum Y_{i1}(v_{i3}-v_{i2})\right)$$

Arellano-Bond (1991): For $T>3$, additional lags of Y could be used as instruments to provide additional moment conditions.

$$Y_{it} = \rho Y_{i(t-1)} + \mu_i + v_{it}$$

$$Y_{it} - Y_{i(t-1)} = \rho(Y_{i(t-1)} - Y_{i(t-2)}) + v_{it} - v_{i(t-1)}.$$

For period t there are $t-2$ valid instruments: $Y_{i1}, Y_{i2}, \ldots, Y_{it-2}$. We can specify the instrument moment conditions as:

$$Z_i' \tilde{V}_i = \begin{bmatrix} Y_{i1} & 0 & \cdots & 0 \\ 0 & Y_{i1} & 0 \cdots & 0 \\ 0 & Y_{i2} & 0 \cdots & 0 \\ \cdots & & & Y_{i1} \\ & & & \vdots \\ & & & Y_{i(T-2)} \end{bmatrix} \begin{bmatrix} \tilde{V}_{i3} \\ \vdots \\ \tilde{V}_{iT} \end{bmatrix}$$

Adding in X_{it} for $t \geq 3$ wouldn't change this a lot since we would be using $\tilde{X}_{it}$ as its own instrument (strict exogeneity).

Problem: Predictive power of larger lags diminishes, so may have an issue of many weak instruments. Advisable to use only first few lags for each time period. Usual finite sample criticisms of GMM vs 2SLS applies here.

To see where this problem of weak instruments comes from consider 1st stage:

$$Y_{i(t-1)} - Y_{i(t-2)} = \gamma \cdot Y_{i(t-2)} + \varepsilon_{it}$$

$$\gamma = \frac{E(y_{it-2} \cdot \Delta y_{it-1})}{E(y_{it-2}^2)} = \rho - 1 + \frac{E(y_{it-2}\mu_i)}{E(y_{i(t-2)}^2)}$$

Under a stationarity assumption: $E(y_{it-2}\mu_i) = \dfrac{\sigma_\mu^2}{1-\rho}$.

$$E(y_{it-2}^2) = \frac{\sigma_\mu^2}{(1-\rho)^2} + \frac{\sigma_v^2}{1-\rho^2}$$

$$\Rightarrow \gamma = (\rho - 1)\left(\frac{\frac{1-\rho}{1+\rho}}{\frac{1-\rho}{1+\rho} + \frac{\sigma_\mu^2}{\sigma_v^2}}\right) \approx 0 \quad \text{if either } \rho \approx 1 \text{ or } \frac{\sigma_\mu^2}{\sigma_v^2} \text{ is large.}$$

In this case, can improve performance by using differenced lagged dependent variables as instruments. (Blundell + Bond (1998), Arellano Bover (1995)).

Consider $y_{it} = \rho y_{i(t-1)} + \mu_i + v_{it}$.

Again: $\operatorname{Cov}(y_{i(t-1)}, \mu_i) \neq 0$.

Propose: $z_{it} = y_{i(t-1)} - y_{i(t-2)}$. Need z_{it} to be uncorrelated with $\mu_i + v_{it}$ but correlated with $y_{i(t-1)}$.

$\text{Cov}\left(Y_{i(t-1)}, Y_{i(t-1)} - Y_{i(t-2)}\right) \neq 0$, but

$E\left(Y_{i(t-1)} - Y_{i(t-2)} \mid V_{it}\right) = 0$, so validity depends on $E\left(\left(Y_{i(t-1)} - Y_{i(t-2)}\right)\mu_i\right) = 0$.

Blundell + Bond argue that a sufficient condition is

$$E\left(\left(Y_{i1} - \frac{\mu_i}{1-\rho}\right)\mu_i\right) = 0.$$

To see this, Recall: $\Delta Y_{it} = (\rho - 1) Y_{i(t-1)} + \mu_i + V_{it}$

So $\Delta Y_{i2} = (\rho-1) Y_{i1} + \mu_i + V_{i2}$. (1)

$= (\rho-1)\left(Y_{i1} - \frac{\mu_i}{1-\rho}\right) + V_{i2}$.

Alternatively: $\Delta Y_{i(t-1)} = \rho \Delta Y_{i(t-2)} + \Delta V_{it-1}$

$= \rho\left(\rho \Delta Y_{i(t-3)} + \Delta V_{i(t-2)}\right) + \Delta V_{it-1}$

$= \cdots$

$= \rho^{t-3} \Delta Y_{i2} + \sum_{j=0}^{t-3} \rho^j \Delta V_{i,t-1-j}$

Plug in (1): $\Delta Y_{i(t-1)} = \rho^{t-3}\left[(\rho-1)\left[Y_{i1} - \frac{\mu_i}{1-\rho}\right] + V_{i2}\right] + \sum_{j=0}^{t-3} \rho^j \Delta V_{i,t+j}$

So $E\left(\Delta Y_{i(t-1)} \mu_i\right) = \rho^{t-3}(\rho-1) E\left(\left(Y_{i1} - \frac{\mu_i}{1-\rho}\right)\mu_i\right) = 0$, where the first equality holds by assuming μ_i, v_{it} are indep. $\forall t$.

A sufficient condition for $E\left(\left(Y_{i1} - \frac{\mu_i}{1-\rho}\right)\mu_i\right) = 0$ is stationarity, since then

$E(Y_{it} | \mu_i) = \frac{\mu_i}{1-\rho}$, so the condition holds by the LIE.

The resulting moment conditions are:

$$E\left(\Delta Y_{i,t-1}\left(Y_{it} - \rho Y_{it-1}\right)\right) = 0.$$

Blundell-Bond estimator of ρ found using these moment conditions for GMM estimation.