UNOBSERVED EFFECTS LINEAR PANEL DATA MODELS, II

Econometric Analysis of Cross Section and Panel Data, 2e
MIT Press
Jeffrey M. Wooldridge

1. Equivalence Between GMM 3SLS and Standard Estimators
2. Chamberlain's Approach to UE Models
3. RE and FE Instrumental Variables Methods
4. Hausman and Taylor Models
5. First Differencing and IV
6. Measurement Error
7. Estimation under Sequential Exogeneity

1. EQUIVALENCE BETWEEN GMM 3SLS AND STANDARD ESTIMATORS

- Consider the standard UE model

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

which we write for all T time periods as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + c_i\mathbf{j}_T + \mathbf{u}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{v}_i.$$

- RE, FE, and FD still the most popular approaches to estimating $\boldsymbol{\beta}$ with strictly exogenous explanatory variables. Or, can use GLS versions of FE and FD.

• But what about the system IV procedures we discussed? We have lots of moment conditions. Suppose we impose RE.1. Then the explanatory variables are strictly exogenous with respect to the composite errors:

$$E(\mathbf{v}_i|\mathbf{x}_i) = \mathbf{0}$$

where $\mathbf{x}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT})$, as usual.

• Consequently, any system FGLS estimator using a constant variance-covariance matrix will be consistent.

- Alternatively, we can use all covariates across time as instruments in all time periods. Let $\mathbf{x}_i^o$ denote all nonredundant elements of $\mathbf{x}_i$, and define

$$\mathbf{Z}_i = \mathbf{I}_T \otimes \mathbf{x}_i^o.$$

- Now we have

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{v}_i \\ E(\mathbf{Z}_i'\mathbf{v}_i) &= \mathbf{0} \end{aligned}$$

and we can use GMM-3SLS or GMM with an unrestricted optimal weighting matrix.

- The GMM estimator using an optimal weighting matrix is generally (asymptotically) more efficient than RE or GLS with $\hat{\mathbf{\Omega}}$ unrestricted.
- There are many overidentifying restrictions in $E(\mathbf{Z}_i'\mathbf{v}_i) = \mathbf{0}$. Perhaps too many?
- If we impose system homoskedasticity then we do not improve over FGLS because of the following algebraic result: if we apply GMM-3SLS estimation with variance matrix $\hat{\mathbf{\Omega}}$ and IVs $\mathbf{Z}_i = \mathbf{I}_T \otimes \mathbf{x}_i^o$, we get the GLS estimator that uses $\hat{\mathbf{\Omega}}$ (for any structure of $\hat{\mathbf{\Omega}}$). [See Im, Ahn, Schmidt, and Wooldridge (1999, Journal of Econometrics.)]

• In the presence of system heteroskedasticity, is there a way to improve on RE without using the many overidentifying restrictions implied by $\mathbf{Z}_i = \mathbf{I}_T \otimes \mathbf{x}_i^o$?

• Yes. Let

$$\mathbf{P}_T = \mathbf{j}_T(\mathbf{j}_T'\mathbf{j}_T)^{-1}\mathbf{j}_T' = \mathbf{j}_T\mathbf{j}_T'/T = T^{-1}\begin{pmatrix} 1 & 1 & \vdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

$$\mathbf{Q}_T = \mathbf{I}_T - \mathbf{P}_T$$

- Let $\mathbf{W}_i$ be the submatrix of $\mathbf{X}_i$ that includes only time-varying variables, and consider the instruments

$$\mathbf{Z}_i = (\mathbf{P}_T\mathbf{X}_i, \mathbf{Q}_T\mathbf{W}_i) = \begin{pmatrix} \bar{\mathbf{x}}_i & \ddot{\mathbf{w}}_{i1} \\ \bar{\mathbf{x}}_i & \ddot{\mathbf{w}}_{i2} \\ \vdots & \vdots \\ \bar{\mathbf{x}}_i & \ddot{\mathbf{w}}_{iT} \end{pmatrix}$$

- Under Assumption RE.1, $E(\mathbf{Z}_i'\mathbf{v}_i) = \mathbf{0}$.

- If all elements of $\mathbf{X}_i$ are time-varying, $\mathbf{Z}_i$ has $2K$ columns, so there are K overidentifying restrictions.
- Algebraic Fact: If $\hat{\mathbf{\Omega}}$ is estimated so it has the RE structure, the GMM-3SLS using $\mathbf{Z}_i$ as instruments and $\hat{\mathbf{\Omega}}$ as the $T \times T$ variance-covariance matrix is identical to the RE estimator.

• Gives a different way to test overidentifying restrictions and also shows we can improve on RE without using too many overidentifying restrictions. If the true $\mathbf{\Omega}$ does not have the RE structure and system homoskedasticity holds, $E(\mathbf{v}_i\mathbf{v}_i'|\mathbf{x}_i) = \mathbf{\Omega}$, then the GMM 3SLS estimator that puts no restrictions on $\mathbf{\Omega}$ is more efficient than RE.

• If we do not restrict $E(\mathbf{v}_i\mathbf{v}_i'|\mathbf{x}_i)$ at all, then we can apply an optimal weighting matrix in GMM, using IVs $\mathbf{Z}_i = (\mathbf{P}_T\mathbf{X}_i, \mathbf{Q}_T\mathbf{W}_i)$, and we have a more efficient estimator than RE or GMM-3SLS.

• If we impose only Assumption FE.1, so that c_i and $\mathbf{x}_{it}$ can be arbitrarily correlated for all t, then we cannot use $\mathbf{P}_T\mathbf{X}_i$ as instruments. Assume all elements of $\mathbf{x}_{it}$ are time-varying.

• Define a "differencing" matrix as the $T \times (T-1)$ matrix

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & -1 & 1 \\ 0 & 0 & 0 & \cdots & 0 & -1 \end{pmatrix}$$

• Can show that the valid IVs in the original equation are

$$\mathbf{Z}_i = \mathbf{L} \otimes \mathbf{x}_i^o.$$

For $T = 4$,

$$\mathbf{L} \otimes \mathbf{x}_i^o = \begin{pmatrix} \mathbf{x}_i^o & \mathbf{0} & \mathbf{0} \\ -\mathbf{x}_i^o & \mathbf{x}_i^o & \mathbf{0} \\ \mathbf{0} & -\mathbf{x}_i^o & \mathbf{x}_i^o \\ \mathbf{0} & \mathbf{0} & -\mathbf{x}_i^o \end{pmatrix}$$

$$\mathbf{Z}_i'\mathbf{v}_i = \begin{pmatrix} \mathbf{x}_i^{o\prime} & -\mathbf{x}_i^{o\prime} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_i^{o} & -\mathbf{x}_i^{o\prime} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{x}_i^{o\prime} & -\mathbf{x}_i^{o\prime} \end{pmatrix} \begin{pmatrix} v_{i1} \\ v_{i2} \\ v_{i3} \\ v_{i4} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_i^{o\prime}(v_{i2} - v_{i1}) \\ \mathbf{x}_i^{o\prime}(v_{i3} - v_{i2}) \\ \mathbf{x}_i^{o\prime}(v_{i4} - v_{i3}) \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{x}_i^{o\prime}(u_{i2} - u_{i1}) \\ \mathbf{x}_i^{o\prime}(u_{i3} - u_{i2}) \\ \mathbf{x}_i^{o\prime}(u_{i4} - u_{i3}) \end{pmatrix} = (\mathbf{I}_{T-1} \otimes \mathbf{x}_i^{o})\Delta\mathbf{u}_i$$

because $v_{it} - v_{it-1} = (c_i + u_{it}) - (c_i + u_{i,t-1}) = u_{it} - u_{i,t-1}$.

• The moment conditions $E[(\mathbf{L} \otimes \mathbf{x}_i^o)'\mathbf{v}_i] = \mathbf{0}$ simply reproduce the usable moment conditions implied by FE.1 (or FD.1):

$$E[\mathbf{x}_{ir}'\Delta u_{it}] = \mathbf{0}, t = 2,\ldots,T, r = 1,\ldots,T.$$

• Algebraic Fact: If we use instruments $\mathbf{L} \otimes \mathbf{x}_i^o$ in GMM-3SLS and $\hat{\mathbf{\Omega}}$ has the RE structure, the GMM-3SLS estimator equals the FE estimator. In other words, if system homoskedasticity holds and the RE variance matrix structure is correct, we cannot improve on FE.

- What if we relax the RE variance structure? Then the GMM-3SLS estimator is the same as FEGLS (with any time period dropped) and FDGLS. In other words (and not surprisingly), under system homoskedasticity, GLS applied to an appropriately transformed system (FE or FD transformation) is efficient.
- See Im, Ahn, Schmidt, and Wooldridge (1999) for other algebraic equivalences.

2. CHAMBERLAIN'S APPROACH TO UE MODELS

- In the standard model

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

Chamberlain simply writes down a linear projection relating c_i to the entire history of the $\mathbf{x}_{it}$. Assume no aggregate time effects for notational simplicity (and no time-constant variables).

$$c_i = \psi + \mathbf{x}_{i1}\boldsymbol{\lambda}_1 + \mathbf{x}_{i2}\boldsymbol{\lambda}_2 + \ldots + \mathbf{x}_{iT}\boldsymbol{\lambda}_T + a_i$$
$$E(a_i) = 0, E(\mathbf{x}_i' a_i) = \mathbf{0},$$

- Assuming finite second moments, this specification is definitional.

Mundlak assumed $\boldsymbol{\lambda}_r = \boldsymbol{\xi}/T$ for $r = 1,\ldots,T$.

- Plugging in gives, for each t,

$$\begin{aligned}
y_{it} &= \mathbf{x}_{it}\boldsymbol{\beta} + \psi + \mathbf{x}_{i1}\boldsymbol{\lambda}_1 + \mathbf{x}_{i2}\boldsymbol{\lambda}_2 + \ldots + \mathbf{x}_{iT}\boldsymbol{\lambda}_T + a_i + u_{it} \\
&\equiv \mathbf{x}_{it}\boldsymbol{\beta} + \psi + \mathbf{x}_{i1}\boldsymbol{\lambda}_1 + \mathbf{x}_{i2}\boldsymbol{\lambda}_2 + \ldots + \mathbf{x}_{iT}\boldsymbol{\lambda}_T + r_{it} \\
&= \mathbf{x}_{it}\boldsymbol{\beta} + \psi + \mathbf{x}_i\boldsymbol{\lambda} + r_{it} \\
&\equiv \mathbf{w}_{it}\boldsymbol{\theta} + r_{it}
\end{aligned}$$

where $\mathbf{w}_{it} = (1, \mathbf{x}_{it}, \mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}) = (1, \mathbf{x}_{it}, \mathbf{x}_i)$.

- Write for all time periods as

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{x}_{i1} & \mathbf{x}_{i1} & \mathbf{x}_{i2} & \cdots & \mathbf{x}_{iT} \\ 1 & \mathbf{x}_{i2} & \mathbf{x}_{i1} & \mathbf{x}_{i2} & \cdots & \mathbf{x}_{iT} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & \mathbf{x}_{iT} & \mathbf{x}_{i1} & \mathbf{x}_{i2} & \cdots & \mathbf{x}_{iT} \end{pmatrix} \begin{pmatrix} \psi \\ \boldsymbol{\beta} \\ \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \\ \vdots \\ \boldsymbol{\lambda}_T \end{pmatrix} + \begin{pmatrix} r_{i1} \\ r_{i2} \\ \vdots \\ r_{iT} \end{pmatrix}$$

$$\mathbf{y}_i = \mathbf{W}_i \boldsymbol{\theta} + \mathbf{r}_i$$

- We can apply system OLS, FGLS, or method of moments procedures.

• Algebraic Fact: If we apply RE to

$$y_{it} = \psi + \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{x}_i\boldsymbol{\lambda} + a_i + u_{it},\ t = 1,\ldots,T,$$

the estimate of $\boldsymbol{\beta}$ is the FE estimate, just as when we use the seemingly more restrictive Mundlak version,

$$y_{it} = \psi + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i + u_{it}, t = 1,\ldots,T.$$

• In the Chamberlain equation, to account for system heteroskedasticity or a non RE unconditional variance matrix, we can use a GMM approach with IV matrix $\mathbf{Z}_i = \mathbf{I}_T \otimes \mathbf{x}_i$. If we use GMM-3SLS with variance matrix $\hat{\boldsymbol{\Omega}}$, this is identical to FGLS using the same $\hat{\boldsymbol{\Omega}}$.

3. RE AND FE INSTRUMENTAL VARIABLES METHODS

- We start with the usual unobserved effects model,

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

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.)

- Pooled 2SLS will be consistent if

$$
\begin{aligned}
E(\mathbf{z}_{it}'c_i) &= \mathbf{0} \\
E(\mathbf{z}_{it}'u_{it}) &= \mathbf{0},\ t = 1, \ldots, T.
\end{aligned}
$$

- In principle, this can be applied to models with lagged dependent variables, although in a model with only a lagged dependent variable, it would be hard to find a convincing instrument.
- Generally, assuming the instruments are uncorrelated with c_i is a strong assumption. If we are willing to make it, we probably are willing to assume strict exogeneity conditional on c_i. So, we can use an RE approach. Assumptions parallel those for exogenous $\mathbf{x}_{it}$. Let $\mathbf{z}_i = (\mathbf{z}_{i1}, \mathbf{z}_{i2}, \ldots, \mathbf{z}_{iT})$; some of these elements may be time-constant, and aggregate time variables act as their own IVs.

ASSUMPTION REIV.1:

$$\text{(a)}\ \ E(u_{it}|\mathbf{z}_i, c_i) = 0, t = 1, \ldots, T$$

$$\text{(b)}\ \ E(c_i|\mathbf{z}_i) = 0$$

- For simplicity, assumes that $\mathbf{x}_{it}$ contains an overall intercept (and probably a separate intercept in each time period), so we can take $E(c_i) = 0$.
- As usual, we could relax the assumptions to zero correlation without changing consistency.
- Define $\boldsymbol{\Omega} = Var(\mathbf{v}_i)$, 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: $\mathbf{\Omega}$ is nonsingular, and

$$\text{(a)}\ \ rank\, E(\mathbf{Z}_i'\mathbf{\Omega}^{-1}\mathbf{Z}_i) = L$$

$$\text{(b)}\ \ rank\, E(\mathbf{Z}_i'\mathbf{\Omega}^{-1}\mathbf{X}_i) = K$$

- This is just the usual rank condition for GIV estimation.

- The REIV estimator is just the GIV estimator where $\mathbf{\Omega}$ is assumed to have the RE form.

- Without further assumptions, fully robust inference is warranted, as usual.

ASSUMPTION REIV.3:

$$\text{(a)}\;\; E(\mathbf{u}_i\mathbf{u}_i'|\mathbf{z}_i, c_i) = \sigma_u^2\mathbf{I}_T$$

$$\text{(b)}\;\; E(c_i^2|\mathbf{z}_i) = \sigma_c^2$$

- 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. By similar reasoning for the usual RE estimator, the REIV estimator can be obtained as pooled 2SLS on the equation

$$y_{it} - \hat{\lambda}\bar{y}_i = (\mathbf{x}_{it} - \hat{\lambda}\bar{\mathbf{x}}_i)\boldsymbol{\beta} + error_{it}$$

using IVs $\mathbf{z}_{it} - \hat{\lambda}\bar{\mathbf{z}}_i$.

• As in the case of RE, the estimation in $\hat{\lambda}$ does not affect $\sqrt{N}$-inference.

• We can test the null that a set of variables is endogenous. Write the model as

$$y_{it} = \mathbf{z}_{it1}\boldsymbol{\delta}_1 + \mathbf{y}_{it2}\boldsymbol{\alpha}_1 + \mathbf{y}_{it3}\boldsymbol{\gamma}_1 + c_{i1} + u_{it1},$$

where $\mathbf{y}_{it2}$ and $\mathbf{y}_{it3}$ are the potential endogenous explanatory variables. Under H_0, we allow $\mathbf{y}_{it2}$ to be endogenous, and test the null that $\mathbf{y}_{it3}$ is exogenous. We maintain strict exogeneity of the instruments $\mathbf{z}_{it}$:

$$E(v_{it1}|\mathbf{z}_i) = 0, t = 1,\ldots,T$$

where $v_{it1} = c_{i1} + u_{it1}$.

• Reduced form for $\mathbf{y}_{it3}$:

$$\mathbf{y}_{it3} = \mathbf{z}_{it}\boldsymbol{\Pi}_3 + \mathbf{v}_{it3}.$$

- Estimate each reduced form by POLS or RE on each equation, and get the residuals, $\hat{\mathbf{v}}_{it3} = \mathbf{y}_{it3} - \mathbf{z}_{it}\hat{\mathbf{\Pi}}_3$. Then, estimate the augmented model

$$y_{it} = \mathbf{z}_{it1}\boldsymbol{\delta}_1 + \mathbf{y}_{it2}\boldsymbol{\alpha}_1 + \mathbf{y}_{it3}\boldsymbol{\gamma}_1 + \hat{\mathbf{v}}_{it3}\boldsymbol{\rho}_1 + error_{it}$$

by REIV and test $H_0 : \boldsymbol{\rho}_1 = \mathbf{0}$. If $\mathbf{y}_{it3}$ has dimension $1 \times J_1$, then the test has J_1 dfs. Because we are using RE, we are actually testing strict exogeneity of $\{\mathbf{y}_{it3} : t = 1,\ldots,T\}$:

$$H_0 : E(\mathbf{y}_{is3}' v_{it1}) = \mathbf{0}, \text{ all } s,t$$

- As usual, a fully robust test is attractive. Note that a test rejection is hard to intepret if $\{\mathbf{z}_{it}\}$ is not strictly exogenous.

- To test overidentifying restrictions, write

$$y_{it} = \mathbf{z}_{it1}\boldsymbol{\delta}_1 + \mathbf{y}_{it2}\boldsymbol{\alpha}_1 + c_{i1} + u_{it1},$$

where $\mathbf{z}_{it2}$ (the omitted exogenous variables) has dimension L_2 and $\mathbf{y}_{it2}$ has dimension G_1. The number of restrictions is $Q_1 = L_2 - G_1$. Write $\mathbf{z}_{it2} = (\mathbf{g}_{it2}, \mathbf{h}_{it2})$ where $\mathbf{g}_{it2}$ also has dimension G_1. Form the augmented model

$$y_{it} = \mathbf{z}_{it1}\boldsymbol{\delta}_1 + \mathbf{y}_{it2}\boldsymbol{\alpha}_1 + \mathbf{h}_{it2}\boldsymbol{\lambda}_1 + c_{i1} + u_{it1},$$

estimate by REIV, and test $H_0 : \boldsymbol{\lambda}_1 = \mathbf{0}$ (using a robust test).

• 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.

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

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

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

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

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

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

• 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:

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

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

• 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,

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

• Under FE.1, FE.2, and FE.3, the asymptotic variance matrix of $\hat{\boldsymbol{\beta}}_{FEIV}$ 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

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

• The tests for endogeneity and overidentification are based on the same equations but use the within transformation. So, for overidentification, estimate

$$y_{it} = \mathbf{z}_{it1}\boldsymbol{\delta}_1 + \mathbf{y}_{it2}\boldsymbol{\alpha}_1 + \mathbf{h}_{it2}\boldsymbol{\lambda}_1 + c_{i1} + u_{it1}$$

by FEIV and test $H_0 : \boldsymbol{\lambda}_1 = \mathbf{0}$.

• For endogeneity, estimate

$$\mathbf{y}_{it3} = \mathbf{z}_{it}\boldsymbol{\Pi}_3 + \mathbf{v}_{it3}$$

by FE and obtain the FE residuals, $\hat{\ddot{\mathbf{v}}}_{it3}$.

Then estimate

$$y_{it} = \mathbf{z}_{it1}\boldsymbol{\delta}_1 + \mathbf{y}_{it2}\boldsymbol{\alpha}_1 + \mathbf{y}_{it3}\boldsymbol{\gamma}_1 + \hat{\ddot{\mathbf{v}}}_{it3}\boldsymbol{\rho}_1 + error_{it}$$

by FEIV and test $H_0 : \boldsymbol{\rho}_1 = \mathbf{0}$. As with the other tests, the first-stage estimation does not affect the asymptotic distribution under the null.

More Specification Tests

- Can use a simple regression form of the Hausman test comparing REIV and FEIV. FEIV is equivalent to the REIV and pooled IV estimators that add the time average of the IVs, $\bar{\mathbf{z}}_i$, as regressors.
- Estimate

$$y_{it} = \alpha + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{z}}_i\boldsymbol{\xi} + a_i + u_{it}$$

by pooled 2SLS or REIV, using instruments $(1, \mathbf{z}_{it}, \bar{\mathbf{z}}_i)$. The estimator of $\boldsymbol{\beta}$ is the FEIV estimator.

- Test $H_0 : \boldsymbol{\xi} = \mathbf{0}$, preferably using a fully robust test. (xtivreg2 does not allow this with REIV). A rejection is evidence that the IVs are correlated with c_i, and should use FEIV.

• Other than the rank condition, the key condition for FEIV to be consistent is that the instruments, $\{\mathbf{z}_{it}\}$, are strictly exogenous with respect to $\{u_{it}\}$. With $T \geq 3$ time periods, this is easily tested – as in the usual FE case.

• The augmented model is

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

and we estimate it by FEIV, using instruments $(\mathbf{z}_{it}, \mathbf{z}_{i,t+1})$.

• Use a fully robust Wald test of $H_0 : \boldsymbol{\delta} = \mathbf{0}$. Can be selected about which leads to include.

EXAMPLE: Estimating Passenger Demand

```
. * First, use pooled IV, instrumenting lfare with concen

. ivreg lpassen ldist ldistsq y98 y99 y00 (lfare = concen), cluster(id)

Instrumental variables (2SLS) regression               Number of obs =    4596
                                                       F(  6,  1148) =   28.02
                                                       Prob > F      =  0.0000
                                                       R-squared     =       .
                                                       Root MSE      =  .95062

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
     lpassen |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfare |  -1.776549   .4753368    -3.74   0.000    -2.709175   -.8439226
       ldist |  -2.498972    .831401    -3.01   0.003    -4.130207   -.8677356
     ldistsq |   .2314932   .0705247     3.28   0.001     .0931215    .3698649
         y98 |   .0616171   .0131531     4.68   0.000     .0358103    .0874239
         y99 |   .1241675   .0183335     6.77   0.000     .0881967    .1601384
         y00 |   .2542695   .0458027     5.55   0.000      .164403    .3441359
       _cons |   21.21249   3.860659     5.49   0.000     13.63775    28.78722
------------------------------------------------------------------------------
Instrumented:  lfare
Instruments:   ldist ldistsq y98 y99 y00 concen
------------------------------------------------------------------------------
```

```
. xtivreg lpassen ldist ldistsq y98 y99 y00 (lfare = concen), re theta

G2SLS random-effects IV regression              Number of obs      =      4596
Group variable: id                              Number of groups   =      1149

R-sq:  within  = 0.4075                         Obs per group: min =         4
       between = 0.0542                                        avg =       4.0
       overall = 0.0641                                        max =         4

                                                Wald chi2(6)       =    231.10
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000
theta              = .91099494

------------------------------------------------------------------------------
     lpassen |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfare |  -.5078762    .229698    -2.21   0.027     -.958076   -.0576763
       ldist |  -1.504806   .6933147    -2.17   0.030    -2.863678   -.1459338
     ldistsq |   .1176013   .0546255     2.15   0.031     .0105373    .2246652
         y98 |   .0307363   .0086054     3.57   0.000     .0138699    .0476027
         y99 |   .0796548     .01038     7.67   0.000     .0593104    .0999992
         y00 |   .1325795   .0229831     5.77   0.000     .0875335    .1776255
       _cons |   13.29643   2.626949     5.06   0.000     8.147709    18.44516
-------------+----------------------------------------------------------------
```

```
     sigma_u |  .94920686
     sigma_e |  .16964171
         rho |  .96904799   (fraction of variance due to u_i)
------------------------------------------------------------------------------
Instrumented:   lfare
Instruments:    ldist ldistsq y98 y99 y00 concen
------------------------------------------------------------------------------

. * The quasi-time-demeaning parameter is quite large: .911 ("theta"), which
. * explains why REIV and pooled IV are so different.
```

```
. xtivreg lpassen y98 y99 y00 (lfare = concen), fe

Fixed-effects (within) IV regression           Number of obs      =      4596
Group variable: id                             Number of groups   =      1149

R-sq:  within  = 0.2265                        Obs per group: min =         4
       between = 0.0487                                       avg =       4.0
       overall = 0.0574                                       max =         4

                                               Wald chi2(4)       =  5.78e+06
corr(u_i, Xb)  = 0.0708                        Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     lpassen |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfare |  -.3015761   .2774005    -1.09   0.277    -.8452711     .242119
         y98 |   .0257147   .0097819     2.63   0.009     .0065426    .0448869
         y99 |   .0724166   .0120342     6.02   0.000       .04883    .0960031
         y00 |   .1127914   .0275332     4.10   0.000     .0588273    .1667556
       _cons |   7.501008   1.402758     5.35   0.000     4.751653    10.25036
-------------+----------------------------------------------------------------
     sigma_u |   .8493153
     sigma_e |  .16964171
         rho |  .96163479   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F  test that all u_i=0:     F(1148,3443) =    99.70       Prob > F    = 0.0000
------------------------------------------------------------------------------
Instrumented:   lfare
Instruments:    y98 y99 y00 concen
------------------------------------------------------------------------------
```

```
. * Obtain the Hausman test comparing RE versus FE

. egen concenb = mean(concen), by(id)

. xtivreg lpassen ldist ldistsq y98 y99 y00 concenb (lfare = concen), re theta

G2SLS random-effects IV regression              Number of obs      =      4596
Group variable: id                              Number of groups   =      1149

R-sq:  within  = 0.3188                         Obs per group: min =         4
       between = 0.0600                                        avg =       4.0
       overall = 0.0669                                        max =         4

                                                Wald chi2(7)       =    218.80
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000
theta              = .90084889

------------------------------------------------------------------------------
     lpassen |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfare |  -.3015761   .2764376    -1.09   0.275    -.8433838    .2402316
       ldist |  -1.148781   .6970189    -1.65   0.099    -2.514913    .2173514
     ldistsq |   .0772565   .0570609     1.35   0.176    -.0345808    .1890937
         y98 |   .0257147   .0097479     2.64   0.008     .0066092    .0448203
         y99 |   .0724165   .0119924     6.04   0.000     .0489118    .0959213
         y00 |   .1127914   .0274377     4.11   0.000     .0590146    .1665682
     concenb |  -.5933022   .1926313    -3.08   0.002    -.9708527   -.2157518
       _cons |    12.0578   2.735977     4.41   0.000     6.695384    17.42022
-------------+----------------------------------------------------------------
```

```
     sigma_u |  .85125514
     sigma_e |  .16964171
         rho |  .96180277   (fraction of variance due to u_i)
------------------------------------------------------------------------------
Instrumented:   lfare
Instruments:    ldist ldistsq y98 y99 y00 concenb concen
------------------------------------------------------------------------------
```

```
. ivreg lpassen ldist ldistsq y98 y99 y00 concenb (lfare = concen),
       cluster(id)

Instrumental variables (2SLS) regression               Number of obs =    4596
                                                       F(  7,  1148) =   20.28
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0649
                                                       Root MSE      =  .85549

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
     lpassen |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfare |  -.3015769   .6131465    -0.49   0.623     -1.50459    .9014366
       ldist |  -1.148781   .8809895    -1.30   0.193    -2.877312    .5797488
     ldistsq |   .0772566   .0811787     0.95   0.341    -.0820187    .2365319
         y98 |   .0257148   .0164291     1.57   0.118    -.0065196    .0579491
         y99 |   .0724166   .0251272     2.88   0.004     .0231163    .1217169
         y00 |   .1127915   .0620858     1.82   0.070    -.0090228    .2346058
     concenb |  -.5933019   .2963723    -2.00   0.046    -1.174794   -.0118099
       _cons |   12.05781   4.360868     2.77   0.006      3.50164    20.61397
------------------------------------------------------------------------------
Instrumented:  lfare
Instruments:   ldist ldistsq y98 y99 y00 concenb concen
------------------------------------------------------------------------------
```

```
. * Original form of the Hausman test breaks down, even with “sigmamore” or
. * “sigmaless” option. Thinks there are 4 df in test when there is only
. * one.

. qui xtivreg2 lpassen y98 y99 y00 (lfare = concen), fe

. estimates store b_feiv

. qui xtivreg lpassen y98 y99 y00 (lfare = concen), re

. estimates store b_reiv

. hausman b_feiv b_reiv

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     b_feiv       b_reiv       Difference          S.E.
-------------+----------------------------------------------------------------
       lfare |   -.3015761    -.6540984        .3525224                .
         y98 |    .0257147     .0342955       -.0085808                .
         y99 |    .0724166     .0847852       -.0123686                .
         y00 |    .1127914      .146605       -.0338136                .
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from xtivreg2
           B = inconsistent under Ha, efficient under Ho; obtained from xtivreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(4) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =    -1.47    chi2<0 ==> model fitted on these
                                        data fails to meet the asymptotic
                                        assumptions of the Hausman test;
                                        see suest for a generalized test
```

```
. * Using the same variance-covariance matrix solves problem of negative
. * statistic, but not incorrect df:

. hausman b_feiv b_reiv, sigmamore

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     b_feiv       b_reiv       Difference          S.E.
-------------+----------------------------------------------------------------
       lfare |   -.3015761    -.6540984        .3525224        1.465766
         y98 |    .0257147     .0342955       -.0085808        .0523018
         y99 |    .0724166     .0847852       -.0123686        .0640888
         y00 |    .1127914      .146605       -.0338136        .1457034
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from xtivreg2
          B = inconsistent under Ha, efficient under Ho; obtained from xtivreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(4) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        0.06
                Prob>chi2 =      0.9996
```

```
. * Now test whether instrument (concen) is strictly exogenous using FEIV:

. xtivreg lpassen y98 y99 concen_p1 (lfare = concen), fe

Fixed-effects (within) IV regression         Number of obs      =      3447
Group variable: id                           Number of groups   =      1149

R-sq:  within  = 0.4474                      Obs per group: min =         3
       between = 0.0496                                     avg =       3.0
       overall = 0.0564                                     max =         3

                                             Wald chi2(4)       =  7.64e+06
corr(u_i, Xb)  = -0.2111                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     lpassen |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfare | -.8520992     .15393    -5.54   0.000    -1.153796    -.550402
         y98 |  .0416985   .0064586     6.46   0.000     .0290398    .0543571
         y99 |  .0948286   .0074973    12.65   0.000     .0801343     .109523
   concen_p1 |  .1555725   .0482045     3.23   0.001     .0610935    .2500516
       _cons |  10.18819   .7852193    12.97   0.000     8.649187    11.72719
-------------+----------------------------------------------------------------
     sigma_u |  .8600387
     sigma_e | .12748791
         rho | .97849882   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F  test that all u_i=0:     F(1148,2294) =   128.42       Prob > F    = 0.0000
------------------------------------------------------------------------------
Instrumented:   lfare
Instruments:    y98 y99 concen_p1 concen
------------------------------------------------------------------------------
```

```
. xtivreg2 lpassen y98 y99 concen_p1 (lfare = concen), fe cluster(id)

FIXED EFFECTS ESTIMATION
------------------------
Number of groups =      1149                   Obs per group: min =          3
                                                              avg =        3.0
                                                              max =          3

IV (2SLS) estimation
--------------------

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity and clustering on id

Number of clusters (id) =         1149                Number of obs =     3447
                                                      F(  4,  1148) =    33.41
                                                      Prob > F      =   0.0000
Total (centered) SS     =  67.47207834                Centered R2   =   0.4474
Total (uncentered) SS   =  67.47207834                Uncentered R2 =   0.4474
Residual SS             =  37.28476721                Root MSE      =    .1274

------------------------------------------------------------------------------
             |               Robust
     lpassen |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfare |  -.8520992   .3211832    -2.65   0.008    -1.481607   -.2225917
         y98 |   .0416985   .0098066     4.25   0.000     .0224778    .0609192
         y99 |   .0948286    .014545     6.52   0.000      .066321    .1233363
   concen_p1 |   .1555725   .0814452     1.91   0.056    -.0040571    .3152021
------------------------------------------------------------------------------
```

```
Instrumented:         lfare
Included instruments: y98 y99 concen_p1
Excluded instruments: concen
------------------------------------------------------------------------------

. * Fully robust test gives a marginal rejection; down to three time periods
. * for the test.
```

```
. * What if we just use fixed effects without IV?

. xtreg lpassen lfare y98 y99 y00, fe cluster(id)

Fixed-effects (within) regression               Number of obs      =      4596
Group variable: id                              Number of groups   =      1149

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
     lpassen |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfare |  -1.155039   .1086574   -10.63   0.000    -1.368228   -.9418496
         y98 |   .0464889   .0049119     9.46   0.000     .0368516    .0561262
         y99 |   .1023612   .0063141    16.21   0.000     .0899727    .1147497
         y00 |   .1946548   .0097099    20.05   0.000     .1756036     .213706
       _cons |   11.81677     .55126    21.44   0.000     10.73518    12.89836
-------------+----------------------------------------------------------------
     sigma_u |  .89829067
     sigma_e |  .14295339
         rho |   .9753002   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

4. HAUSMAN AND TAYLOR MODELS

• The previous IV methods require us to find IVs from outside the model. This is often difficult (just as in the cross section case). Hausman and Taylor (1981) proposed assuming that certain variables are appropriately exogenous, and using these as IVs.

• Write the HT model as

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

where $\mathbf{w}_i$ are time-constant variables (including an intercept) and $\mathbf{x}_{it}$ are time-varying variables.

- Maintain the strict exogeneity assumption conditional on c_i, RE.1(a):

$$E(u_{it}|\mathbf{w}_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, c_i) = 0.$$

- If we want to estimate $\boldsymbol{\beta}$, can just use FE without further assumptions.

What about estimation of $\boldsymbol{\gamma}$?

- Suppose we *assume* that $\mathbf{w}_i$ is uncorrelated with c_i:

$$E(\mathbf{w}_i' c_i) = \mathbf{0}.$$

Now, $\bar{y}_i - \mathbf{w}_i\boldsymbol{\gamma} - \bar{\mathbf{x}}_i\boldsymbol{\beta} = c_i + \bar{u}_i$, and so

$$E[\mathbf{w}_i'(\bar{y}_i - \mathbf{w}_i\boldsymbol{\gamma} - \bar{\mathbf{x}}_i\boldsymbol{\beta})] = \mathbf{0}.$$

- Therefore,

$$E(\mathbf{w}_i'\mathbf{w}_i)\boldsymbol{\gamma} = E[\mathbf{w}_i'(\bar{y}_i - \bar{\mathbf{x}}_i\boldsymbol{\beta})].$$

So a consistent estimator of $\boldsymbol{\gamma}$ is

$$\hat{\boldsymbol{\gamma}} = \left(N^{-1}\sum_{i=1}^{N}\mathbf{w}_i'\mathbf{w}_i\right)^{-1}\left[N^{-1}\sum_{i=1}^{N}\mathbf{w}_i'(\bar{y}_i - \bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}_{FE})\right],$$

which is the OLS coefficients from the cross section regression

$$\bar{y}_i - \bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}_{FE} \text{ on } \mathbf{w}_i,\ i = 1,\ldots,N.$$

• In practice, $\mathbf{x}_{it}$ would contain a set of year dummies and $\mathbf{w}_i$ would contain an overall intercept (which allows the mean of $E(c_i)$ to be different from zero).

• Computing standard errors for $\hat{\boldsymbol{\gamma}}$ must account for estimation of $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}_{FE}$. Could stack the two sets of moment conditions and use system IV.

- More general assumptions. Partition $\mathbf{w}_i = (\mathbf{w}_{i1}, \mathbf{w}_{i2})$ and $\mathbf{x}_{it} = (\mathbf{x}_{it1}, \mathbf{x}_{it2})$ where $\mathbf{w}_{i1}$ is $1 \times J_1$, $\mathbf{w}_{i2}$ is $1 \times J_2$, $\mathbf{x}_{it1}$ is $1 \times K_1$, and $\mathbf{x}_{it2}$ is $1 \times K_2$. Then assume

$$
\begin{aligned}
E(\mathbf{w}_{i1}' c_i) &= \mathbf{0} \\
E(\mathbf{x}_{it1}' c_i) &= \mathbf{0}, \quad t = 1, \ldots, T.
\end{aligned}
$$

- Write the system as

$$\mathbf{y}_i = \mathbf{W}_i\boldsymbol{\gamma} + \mathbf{X}_i\boldsymbol{\beta} + \mathbf{v}_i$$

Let $\mathbf{Q}_T$ be the $T \times T$ demeaning matrix. Then $\mathbf{Q}_T\mathbf{X}_i$, the $T \times K$ matrix with rows $\ddot{\mathbf{x}}_{it}$, is a valid set of instruments because $\mathbf{Q}_T\mathbf{j}_T = \mathbf{0}$ and so

$$(\mathbf{Q}_T\mathbf{X}_i)'\mathbf{v}_i = (\mathbf{Q}_T\mathbf{X}_i)'\mathbf{u}_i$$

and

$$E[(\mathbf{Q}_T\mathbf{X}_i)'\mathbf{u}_i] = \mathbf{0}$$

under $E(u_{it}|\mathbf{x}_i, c_i) = 0$.

• Under the assumptions given, $\mathbf{w}_{i1}$ can act as its own IVs. But we need instruments for $\mathbf{w}_{i2}$. If we define $\mathbf{x}_{i1}^{o} = (\mathbf{x}_{i11}, \mathbf{x}_{i21}, \ldots, \mathbf{x}_{iT1})$ then $E(\mathbf{x}_{i1}^{o\prime} c_i) = \mathbf{0}$; technically, we can use all of $\mathbf{x}_{i1}^{o}$ as IVs in each time period. Then, the matrix of IVs is

$$[\mathbf{Q}_T \mathbf{X}_i, \mathbf{j}_T \otimes (\mathbf{w}_{i1}, \mathbf{x}_{i1}^{o})].$$

• But, it is probably misleading to think all of $\mathbf{x}_{i1}^{o}$ has explanatory power for the time-constant $\mathbf{w}_{i2}$. For example, under certain assumptions – i.i.d. is sufficient, but not necessary – $L(\mathbf{w}_{i2} | \mathbf{x}_{i1}^{o}) = L(\mathbf{w}_{i2} | \bar{\mathbf{x}}_{i1})$ where $\bar{\mathbf{x}}_{i1}$ is the $1 \times K_2$ vector of time averages.

• If $K_1 \geq J_2$, might only use $\bar{\mathbf{x}}_{i1}$. (If $K_1 < J_2$ the entire strategy is questionable.) In other words, at time t use IVs

$$(\ddot{\mathbf{x}}_{it}, \mathbf{w}_{i1}, \bar{\mathbf{x}}_{i1})$$

and use REIV. (Again, should make inference robust. Original HT paper assumes RE variance-covariance assumptions and system homoskedasticity.)

• Given the instruments, can use GMM-3SLS with unrestricted $\hat{\mathbf{\Omega}}$ or even optimal GMM with general weighting matrix.

- The bottom line is that, if we have enough time-varying variables that we think are uncorrelated with c_i (as well as being strictly exogenous with respect to $\{u_{it}\}$), then we can allow a subset of the time-constant variables to be correlated with c_i.
- In HT example, $\mathbf{w}_{i1} = (nonwhite_i, union_i)$ and $w_{i2} = educ_i$ (which did not change over time). $\mathbf{x}_{it1}$ contains experience, an indicator for bad health, and an indicator for being unemployed the previous year.
- In fact, in the HT example they effectively take $\mathbf{x}_{it} = \mathbf{x}_{it1}$ because the only other element in $\mathbf{x}_{it}$ is a time dummy, properly treated as uncorrelated with c_i. (The panel was $T = 2$ and several years apart.)

5. FIRST DIFFERENCING AND IV

- Not suprisingly, can use FDIV, too. Useful to allow a general set of IVs:

$$
\begin{aligned}
\Delta y_{it} &= \Delta \mathbf{x}_{it}\boldsymbol{\beta} + \Delta u_{it},\ t = 2,\ldots,T \\
E(\mathbf{w}_{it}'\Delta u_{it}) &= \mathbf{0},\ t = 2,\ldots,T.
\end{aligned}
$$

- Choose as instruments

$$
\mathbf{W}_i = \begin{pmatrix} \mathbf{w}_{i2} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{w}_{i3} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{w}_{iT} \end{pmatrix}.
$$

• As we discussed with system IV, if each $\mathbf{w}_{it}$ has, say, dimension $L \geq K$, then we can take

$$\mathbf{W}_i = \begin{pmatrix} \mathbf{w}_{i2} \\ \mathbf{w}_{i3} \\ \vdots \\ \mathbf{w}_{iT} \end{pmatrix}.$$

• Then, a pooled 2SLS approach is possible. Fully robust inference is straightforward.

- If $\mathbf{W}_i$ has the general diagonal form, use full GMM. As an intial estimator, can use a variation of pooled IV: (1) For each $t = 2, \ldots T$, regress $\Delta\mathbf{x}_{it}$ on $\mathbf{w}_{it}$, $i = 1, \ldots, N$ and obtain the fitted values, $\widehat{\Delta\mathbf{x}}_{it}$. (2) Estimate the equation $\Delta y_{it} = \Delta\mathbf{x}_{it}\boldsymbol{\beta} + \Delta u_{it}$ using IVs $\widehat{\Delta\mathbf{x}}_{it}$ (which are all necessarily $1 \times K$). Again, use robust inference. (Use $\widehat{\Delta\mathbf{x}}_{it}$ as IVs, not regressors.)
- Levitt (1996) estimates a state-level crime equation in FD form:

$$\Delta\log(crime_{it}) = \eta_{t1} + \alpha_1\Delta\log(prison_{it}) + \Delta\mathbf{z}_{it1}\boldsymbol{\delta}_1 + \Delta u_{it1}$$

with IVs dummies for whether final decisions were reached on prison overcrowding litigation.

```
. * First-stage regression (with year dummies suppressed in output):

. reg gpris final1 final2 gpolpc gincpc cag0_14 cag15_17 cag18_24 cag25_34
      cunem cblack cmetro y81 y82 y83 y84 y85 y86 y87 y88 y89 y90 y91 y92 y93,
      cluster(state)

Linear regression                                      Number of obs =     714
                                                       F( 24,    50) =    9.27
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1522
                                                       Root MSE      =  .06237

                                 (Std. Err. adjusted for 51 clusters in state)
------------------------------------------------------------------------------
             |               Robust
       gpris |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      final1 |   -.077488   .0164372    -4.71   0.000    -.1105032   -.0444729
      final2 |  -.0529558   .0160327    -3.30   0.002    -.0851585   -.0207531
      gpolpc |  -.0286921   .0305312    -0.94   0.352    -.0900159    .0326316
      gincpc |   .2095521   .1597362     1.31   0.196    -.1112875    .5303918
     cag0_14 |   2.617307   2.029707     1.29   0.203     -1.45948    6.694094
    cag15_17 |  -1.608738   4.138375    -0.39   0.699    -9.920908    6.703433
    cag18_24 |   .9533678   1.640538     0.58   0.564    -2.341749    4.248485
    cag25_34 |  -1.031684   1.945366    -0.53   0.598    -4.939067      2.8757
       cunem |   .1616595    .280673     0.58   0.567    -.4020888    .7254077
      cblack |  -.0044763   .0266392    -0.17   0.867    -.0579828    .0490301
      cmetro |  -1.418389   .7425213    -1.91   0.062    -2.909787    .0730092
------------------------------------------------------------------------------
```

```
. ivreg gcriv gpolpc gincpc cag0_14 cag15_17 cag18_24 cag25_34 cunem cblack
         cmetro y81 y82 y83 y84 y85 y86 y87 y88 y89 y90 y91 y92 y93
         (gpris = final1 final2)

------------------------------------------------------------------------------
       gcriv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       gpris |  -1.031956   .3699628    -2.79   0.005    -1.758344   -.3055684
      gpolpc |    .035315   .0674989     0.52   0.601    -.0972128    .1678428
      gincpc |   .9101992   .2143266     4.25   0.000     .4893885     1.33101
     cag0_14 |   3.379384   2.634893     1.28   0.200    -1.793985    8.552753
    cag15_17 |   3.549945   5.766302     0.62   0.538    -7.771659    14.87155
    cag18_24 |   3.358348   2.680839     1.25   0.211    -1.905233    8.621929
    cag25_34 |   2.319993   2.706345     0.86   0.392    -2.993667    7.633652
       cunem |   .5236958   .4785632     1.09   0.274     -.415919     1.46331
      cblack |  -.0158476   .0401044    -0.40   0.693    -.0945889    .0628937
      cmetro |   -.591517   1.298252    -0.46   0.649    -3.140516    1.957482
------------------------------------------------------------------------------
Instrumented:  gpris
Instruments:   gpolpc gincpc cag0_14 cag15_17 cag18_24 cag25_34 cunem cblack
               cmetro y81 y82 y83 y84 y85 y86 y87 y88 y89 y90 y91 y92 y93
               final1 final2
------------------------------------------------------------------------------

. * Again, different year intercepts are suppressed.
```

```
. ivreg gcriv gpolpc gincpc cag0_14 cag15_17 cag18_24 cag25_34 cunem cblack
   cmetro y81 y82 y83 y84 y85 y86 y87 y88 y89 y90 y91 y92 y93
  (gpris = final1 final2), cluster(state)

                                 (Std. Err. adjusted for 51 clusters in state)
------------------------------------------------------------------------------
             |               Robust
       gcriv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       gpris |  -1.031956   .2134723    -4.83   0.000    -1.460728   -.6031845
      gpolpc |    .035315   .0549931     0.64   0.524    -.0751419    .1457719
      gincpc |   .9101992   .3375487     2.70   0.010     .2322128    1.588186
     cag0_14 |   3.379384   2.445851     1.38   0.173    -1.533252     8.29202
    cag15_17 |   3.549945   5.458091     0.65   0.518    -7.412954    14.51284
    cag18_24 |   3.358348   3.246766     1.03   0.306    -3.162973    9.879669
    cag25_34 |   2.319993   3.248509     0.71   0.478     -4.20483    8.844815
       cunem |   .5236958   .4749941     1.10   0.276    -.4303579    1.477749
      cblack |  -.0158476   .0306832    -0.52   0.608    -.0774766    .0457815
      cmetro |   -.591517   1.277895    -0.46   0.645    -3.158245    1.975211
------------------------------------------------------------------------------
Instrumented:  gpris
Instruments:   gpolpc gincpc cag0_14 cag15_17 cag18_24 cag25_34 cunem cblack
               cmetro y81 y82 y83 y84 y85 y86 y87 y88 y89 y90 y91 y92 y93
               final1 final2
------------------------------------------------------------------------------
```

```
. reg gcriv gpris gpolpc gincpc cag0_14 cag15_17 cag18_24 cag25_34 cunem
      cblack cmetro y81 y82 y83 y84 y85 y86 y87 y88 y89 y90 y91 y92 y93,
      cluster(state)

                                 (Std. Err. adjusted for 51 clusters in state)
------------------------------------------------------------------------------
             |               Robust
       gcriv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       gpris |  -.1808974   .0487909    -3.71   0.001    -.2788967    -.082898
      gpolpc |   .0514239    .047601     1.08   0.285    -.0441854    .1470333
      gincpc |   .7383676   .2457843     3.00   0.004     .2446953     1.23204
     cag0_14 |    .989306    1.86767     0.53   0.599    -2.762019    4.740631
    cag15_17 |    4.98384   4.758174     1.05   0.300    -4.573234    14.54091
    cag18_24 |   2.412758    3.33858     0.72   0.473    -4.292978    9.118493
    cag25_34 |   2.879946    2.61131     1.10   0.275    -2.365025    8.124917
       cunem |     .41126   .3824013     1.08   0.287    -.3568156    1.179335
      cblack |  -.0147435   .0147599    -1.00   0.323    -.0443896    .0149027
      cmetro |   .5383056   1.112491     0.48   0.631    -1.696199     2.77281
------------------------------------------------------------------------------

. * Without instrumenting for grpis, the estimated prison effect is much smaller.
```

6. MEASUREMENT ERROR

- Consider a simple UE model with measurement error:

$$
\begin{aligned}
y_{it} &= \beta x_{it}^* + c_i + u_{it},\ t = 1, \ldots, T \\
x_{it} &= x_{it}^* + r_{it},\ r = 1, \ldots, T.
\end{aligned}
$$

- Maintain a strict exogeneity assumption conditional on c_i:

$$E(y_{it}|\mathbf{x}_i^*, \mathbf{x}_i, c_i) = E(y_{it}|x_{it}^*, c_i) = \beta x_{it}^* + c_i$$

or

$$E(u_{it}|\mathbf{x}_i^*, \mathbf{x}_i, c_i) = 0.$$

- Substitute the observed measure, x_{it}, for x_{it}^*: $x_{it}^* = x_{it} - r_{it}$, so

$$y_{it} = \beta x_{it} + c_i + u_{it} - \beta r_{it}$$

- Let $\hat{\beta}_{POLS}$ be the pooled OLS estimator from y_{it} on x_{it} across t and i. If we make the classical errors-in-variables (CEV) assumption

$$Cov(x_{it}^*, r_{it}) = 0$$

and assume constant variances and covariances across t,

$$\begin{aligned}\text{plim}(\hat{\beta}_{POLS}) &= \beta + \frac{Cov(x_{it}, c_i + u_{it} - \beta r_{it})}{Var(x_{it})} \\ &= \beta + \frac{Cov(x_{it}, c_i) - \beta\sigma_r^2}{\sigma_{x^*}^2 + \sigma_r^2}.\end{aligned}$$

• Without an unobserved effect, or if $Cov(x_{it}, c_i) = 0$, we get the standard attenuation bias:

$$\text{plim}(\hat{\beta}_{POLS}) = \beta\left(\frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_r^2}\right)$$

• If $\beta > 0$ and $Cov(x_{it}, c_i) > 0$, the presence of c_i can *help* reduced the "bias."

- What about removing c_i? Suppose $T = 2$, and assume

$$Cov(x_{is}^*, r_{it}) = 0, \text{ all } s, t.$$

- Let $\hat{\beta}_{FD}$ be the usual FD estimator. Then

$$\begin{aligned}
\text{plim}(\hat{\beta}_{FD}) &= \beta + \frac{Cov(\Delta x_{it}, \Delta u_{it} - \beta \Delta r_{it})}{Var(\Delta x_{it})} \\
&= \beta - \beta \frac{Cov(\Delta x_{it}, \Delta r_{it})}{Var(\Delta x_{it})}.
\end{aligned}$$

- Now

$$Var(\Delta x_{it}) = Var(\Delta x^*_{it} + \Delta r_{it}) = Var(\Delta x^*_{it}) + Var(\Delta r_{it})$$
$$= 2\sigma^2_{x^*}(1 - \rho_{x^*}) + 2\sigma^2_r(1 - \rho_r)$$

where $\rho_{x^*} = Corr(x^*_{it}, x^*_{i,t-1})$ and $\rho_r = Corr(r_{it}, r_{i,t-1})$. Also,

$$Cov(\Delta x_{it}, \Delta r_{it}) = Var(\Delta r_{it}) = 2\sigma^2_r(1 - \rho_r).$$

Therefore,

$$\text{plim}(\hat{\beta}_{FD}) = \beta\left(\frac{\sigma_{x^*}^2(1-\rho_{x^*})}{\sigma_{x^*}^2(1-\rho_{x^*}) + \sigma_r^2(1-\rho_r)}\right).$$

- The attenuation bias depends on the amount of serial correlation in $\{x_{it}^*\}$ and $\{r_{it}\}$. The usual formula as $\rho_{x^*} = \rho_r = 0$.
- The attenuation is worse the larger is $(1-\rho_r)/(1-\rho_{x^*})$. As $\rho_{x^*} \to 1$, $\text{plim}(\hat{\beta}_{FD})$ heads to zero regardless of β.

• Can solve the problem by have strictly exogenous instrumental variables.

$$\begin{aligned} y_{it} &= \mathbf{z}_{it}\boldsymbol{\gamma} + \delta w_{it}^* + c_i + u_{it} \\ E(u_{it}|\mathbf{z}_i, \mathbf{w}_i^*, \mathbf{w}_i, \mathbf{h}_i, c_i) &= 0 \end{aligned}$$

where $\{\mathbf{h}_{it} : t = 1, \ldots, T\}$ are the strictly exogenous instruments. Note that $\{w_{it}^*\}$ and $\{w_{it}\}$ are strictly exogenous, too.

$$\begin{aligned} y_{it} &= \mathbf{z}_{it}\boldsymbol{\gamma} + \delta w_{it} + c_i + (u_{it} - \delta r_{it}) \\ w_{it} &= w_{it}^* + r_{it} \end{aligned}$$

- Now we assume

$$E(r_{it}|\mathbf{z}_i, \mathbf{w}_i^*, \mathbf{h}_i, c_i) = 0,\ t = 1, \ldots, T.$$

- Must also ensure $\{\mathbf{h}_{it}\}$ has sufficient correlation with $\{w_{it}\}$. Could include a second measure of w_{it}^*. Estimate

$$w_{it} = \mathbf{z}_{it}\boldsymbol{\xi} + \mathbf{h}_{it}\boldsymbol{\pi} + d_i + e_{it}$$

by fixed effects, and reject $H_0 : \boldsymbol{\pi} = \mathbf{0}$.

- Or, we can use FDIV.

- The above methods assume nothing about the serial correlation properties in the measurement error, $\{r_{it}\}$. Suppose

$$E(r_{it}r_{is}) = 0, \text{ all } t \neq s.$$

- If we assume

$$E(r_{it}|\mathbf{z}_i, \mathbf{w}_i^*, c_i) = 0, \ t = 1, \ldots, T.$$

then the addition of no serial correlation means w_{is} is uncorrelated with r_{it} for all $s \neq t$. But w_{it} is correlated with r_{it}.

- FD is useful now because strict exogeneity cannot hold, but other forms do:

$$\Delta y_{it} = \Delta \mathbf{z}_{it}\boldsymbol{\gamma} + \Delta w_{it} + \Delta u_{it} - \beta \Delta r_{it}$$

Valid IVs for Δw_{it} in this equation are, say, $w_{i,t-2}, w_{i,t-3}$, and even $w_{i,t+1}$.

- Of course, as we add lags to the IV list, we lose time periods.

7. ESTIMATION UNDER SEQUENTIAL EXOGENEITY

- We now consider IV estimation of the model

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

under the sequential exogeneity assumption

$$E(u_{it}|\mathbf{x}_{it},\mathbf{x}_{i,t-1},\ldots,\mathbf{x}_{i1},c_i) = 0,\ t = 1,\ldots,T.$$

- Actually, for consistency, we can get by with the weaker form

$$Cov(\mathbf{x}_{is},u_{it}) = 0, \text{all } s \leq t.$$

• This leads to simple moment conditions after first differencing:

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

$$E(\mathbf{x}_{is}'\Delta u_{it}) = \mathbf{0},\ s = 1,\ldots,t-1;\ t = 2,\ldots,T.$$

• Therefore, at time t, the available instruments in the FD equation are in the vector $\mathbf{x}_{i,t-1}^{o} \equiv (\mathbf{x}_{i1},\mathbf{x}_{i2},\ldots,\mathbf{x}_{i,t-1})$, $t = 2,\ldots,T$.

- The matrix of instruments is

$$\mathbf{W}_i = diag(\mathbf{x}_{i1}^o, \mathbf{x}_{i2}^o, \ldots, \mathbf{x}_{i,T-1}^o),$$

which has $T - 1$ rows. Fairly routine to apply GMM estimation with an optimal weight matrix. With even moderate T there are lots of overidentifying restrictions (at least nominally).

• A simple strategy mentioned earlier is available: Estimate a reduced form for $\Delta \mathbf{x}_{it}$ separately for each t. So, at time t, run the regression $\Delta \mathbf{x}_{it}$ on $\mathbf{x}_{i,t-1}^{o}$, $i = 1, \ldots, N$, and obtain the fitted values, $\widehat{\Delta \mathbf{x}}_{it}$. Then, estimate the FD equation

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

by pooled IV using instruments (not regressors) $\widehat{\Delta \mathbf{x}}_{it}$.

• Should probably be done even if using full GMM to confirm that the IVs are sufficiently correlated with $\Delta \mathbf{x}_{it}$.

• Any of the IV approaches can suffer from a weak instrument problem when $\Delta \mathbf{x}_{it}$ has little correlation with $\mathbf{x}^o_{i,t-1}$. In particular, if

$$\begin{aligned} \mathbf{x}_{it} &= \boldsymbol{\omega}_t + \mathbf{x}_{i,t-1} + \mathbf{r}_{it} \\ E(\mathbf{r}_{it} | \mathbf{x}_{i,t-1}, \mathbf{x}_{i,t-2}, \ldots, \mathbf{x}_{i0}) &= \mathbf{0} \end{aligned}$$

then $E(\Delta \mathbf{x}_{it} | \mathbf{x}^o_{i,t-1}) = E(\Delta \mathbf{x}_{it}) = \boldsymbol{\omega}_t$, and IV fails when a full set of year intercepts is included in the equation.

• If we add some assumptions, we can get more moment conditions.

- Suppose we assume dynamic completeness in the mean:

$$E(u_{it}|\mathbf{x}_{it}, y_{i,t-1}\mathbf{x}_{i,t-1}, \ldots, y_{i1}, \mathbf{x}_{i1}, c_i) = 0.$$

- This condition rules out serial correlation in $\{u_{it}\}$, which is often too restrictive when $\mathbf{x}_{it}$ does not include $y_{i,t-1}$. (If $y_{i,t-1} \in \mathbf{x}_{it}$, there is no difference between sequential exogenenity and dynamic completeness.)

• Dynamic completeness means many more moment conditions are available. Using linear functions only, for $t = 3, \ldots, T$,

$$E[\Delta u_{i,t-1}(c_i + u_{it})] = E[(\Delta y_{i,t-1} - \Delta \mathbf{x}_{i,t-1}\boldsymbol{\beta})'(y_{it} - \mathbf{x}_{it}\boldsymbol{\beta})] = \mathbf{0}.$$

• Drawback: We often do not want to assume dynamic completeness. Plus, the extra conditions are nonlinear in parameters.

• Arellano and Bover (1995) suggested instead the restrictions

$$Cov(\Delta \mathbf{x}_{it}', c_i) = 0, \;\; t = 2, \ldots, T$$

which allows the "level" of the sequence $\{\mathbf{x}_{it} : t = 1, \ldots, T\}$ to be correlated with c_i but not the changes. Holds if

$$\mathbf{x}_{it} = \boldsymbol{\omega}_t + \mathbf{h}_i + \mathbf{r}_{it}$$

where $\{\mathbf{r}_{it} : t = 1, 2, \ldots T\}$ are uncorrelated with c_i. Allows $\mathbf{h}_i$ and c_i to be arbitrarily correlated.

- Would fail if, say,

$$\mathbf{x}_{it} = \boldsymbol{\omega}_t + \mathbf{h}_i + \mathbf{g}_i t + \mathbf{r}_{it}$$

where $\mathbf{g}_i$ and c_i are correlated.

- To use the Arellano and Bover moment conditions, need to let $\alpha = E(c_i)$ to allow a nonzero mean. Then

$$E[\Delta\mathbf{x}_{it}'((c_i - \alpha) + u_{it})] = \mathbf{0},\ t = 2,\ldots,T.$$

• In terms of the parameters and observable data, we have the moment conditions

$$E[\Delta \mathbf{x}_{it}'(y_{it} - \alpha - \mathbf{x}_{it}\boldsymbol{\beta})] = \mathbf{0},\ t = 2, \ldots, T.$$

We can use these along with the moment conditions in the FD equation that are implied by sequential exogeneity. Note that all moment conditions are linear in $\boldsymbol{\beta}$.

• Because we are mixing moment conditions in FD and levels, if $\mathbf{x}_{it}$ includes year dummies (it should) then these must be differenced in $\Delta \mathbf{x}_{it}$.

• The full set of moment conditions:

$$\begin{pmatrix} E[\mathbf{x}_{i1}'(\Delta y_{i2} - \Delta \mathbf{x}_{i2}\boldsymbol{\beta})] \\ \vdots \\ E[\mathbf{x}_{i,T-1}^{o\prime}(\Delta y_{iT} - \Delta \mathbf{x}_{iT}\boldsymbol{\beta})] \\ E[\Delta \mathbf{x}_{i2}'(y_{i2} - \alpha - \mathbf{x}_{i2}\boldsymbol{\beta})] \\ \vdots \\ E[\Delta \mathbf{x}_{iT}'(y_{iT} - \alpha - \mathbf{x}_{it}\boldsymbol{\beta})] \end{pmatrix} = \mathbf{0}.$$

• Use GMM with a general weighting matrix to allow arbitrary correlation across all time periods/equations. (Now called "system" GMM.)

- Simple AR(1) model:

$$y_{it} = \rho y_{i,t-1} + c_i + u_{it}, t = 1, \ldots, T.$$

- Typically, the minimal assumptions imposed are

$$E(y_{is} u_{it}) = 0,\ s = 0, \ldots, t-1,\ t = 1, \ldots, T,$$

(implied by dynamic completeness) so, for $t = 2, \ldots, T$,

$$E[y_{is}(\Delta y_{it} - \rho \Delta y_{i,t-1}) = 0, s \leq t-2.$$

• Again, can suffer from weak instruments when ρ is close to unity.

Blundell and Bond (1998) showed that if the condition

$$Cov(\Delta y_{i1}, c_i) = Cov(y_{i1} - y_{i0}, c_i) = 0$$

is added to $E(u_{it}|y_{i,t-1}, \ldots, y_{i0}, c_i) = 0$ then the Arellano and Bover extra moment conditions hold:

$$E[\Delta y_{i,t-1}(y_{it} - \alpha - \rho y_{i,t-1})] = 0$$

• The condition

$$Cov(\Delta y_{i1}, c_i) = 0$$

can be intepreted as a restriction on the initial condition, y_{i0}. For $|\rho| < 1$, write y_{i0} as a deviation from its steady state: $y_{i0} = c_i/(1-\rho) + r_{i0}$. Then the extra condition is

$$Cov(r_{i0}, c_i) = 0;$$

the deviation of y_{i0} from its steady state is uncorrelated with the SS.

• Potential problem: As ρ approaches one, how realistic is it to assume there is a steady state?

• Extensions of the AR(1) model:

$$y_{it} = \rho y_{i,t-1} + \mathbf{z}_{it}\boldsymbol{\gamma} + c_i + u_{it}, \quad t = 1, \ldots, T$$

and use FD:

$$\Delta y_{it} = \rho \Delta y_{i,t-1} + \Delta \mathbf{z}_{it}\boldsymbol{\gamma} + \Delta u_{it}, \quad t = 2, \ldots, T.$$

• Can use $\Delta \mathbf{z}_{it}$ as own IVs if they are strictly exogenous, $y_{i,t-h}$, $h \geq 2$, and can still add moment conditions in levels.

• Because $y_{i,t-1}$ is included as a control, $\mathbf{z}_{it}$ (perhaps program assignment) is allowed to be correlated with $y_{i,t-1}$ as well as with c_i.

• If $\{\mathbf{z}_{it}\}$ is not strictly exogenous, can use $\{\mathbf{z}_{i,t-1},\ldots,\mathbf{z}_{i1}\}$ as IVs, along with $\{y_{i,t-2},\ldots,y_{i0}\}$ in the FD equation at time t.

• And, we still might use, for $t = 2,\ldots,T,$ the Arellano-Bover moments:

$$\begin{aligned} E[\Delta y_{i,t-1}(y_{it} - \alpha - \rho y_{i,t-1} - \mathbf{z}_{it}\boldsymbol{\gamma})] &= 0 \\ E[\Delta \mathbf{z}_{it}'(y_{it} - \alpha - \rho y_{i,t-1} - \mathbf{z}_{it}\boldsymbol{\gamma})] &= \mathbf{0} \end{aligned}$$

• As usual, time dummies act as their own IVs.

AR(1) Model for Airfare Example

• Use the original Arellano and Bond moment conditions (that is, in the differenced equation only). Put a "d" to indicate the first differences (or changes).

```
. gen dlfare = d.lfare
(1149 missing values generated)

. gen dlfare_1 = l.dlfare
(2298 missing values generated)

. gen dconcen = d.concen
(1149 missing values generated)
```

```
. reg dlfare dlfare_1 dconcen y99 y00, cluster(id)

Linear regression                                      Number of obs =    2298
                                                       F(  3,  1148) =   36.38
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0651
                                                       Root MSE      =   .1168

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
      dlfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    dlfare_1 |  -.1264673   .0267104    -4.73   0.000    -.1788739   -.0740606
     dconcen |   .0762671   .0527226     1.45   0.148    -.0271763    .1797106
         y99 |  -.0473536   .0050308    -9.41   0.000    -.0572241    -.037483
         y00 |  (dropped)
       _cons |   .0624434   .0032977    18.94   0.000     .0559732    .0689136
------------------------------------------------------------------------------

. * Pooled OLS on the differenced equation is very misleading. FE on the levels
. * does better, but is still downward biased.
```

```
. xtreg lfare lfare_1 concen y99 y00, fe cluster(id)

Fixed-effects (within) regression               Number of obs      =      3447
Group variable: id                              Number of groups   =      1149

R-sq:  within  = 0.1605                         Obs per group: min =         3
       between = 0.8863                                        avg =       3.0
       overall = 0.5014                                        max =         3

                                                F(4,1148)          =     98.81
corr(u_i, Xb)  = 0.6597                         Prob > F           =    0.0000

                                   (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lfare_1 |   .0773594   .0318913     2.43   0.015     .0147877    .1399311
      concen |   .0579086   .0533893     1.08   0.278    -.0468428    .1626601
         y99 |   .0098236   .0037176     2.64   0.008     .0025296    .0171177
         y00 |   .0700164   .0043967    15.92   0.000       .06139    .0786428
       _cons |   4.653928   .1628858    28.57   0.000     4.334341    4.973515
-------------+----------------------------------------------------------------
     sigma_u |  .38891209
     sigma_e |  .09055856
         rho |  .94856899   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
. * Try FDIV, generating instruments using first-stage regressions.

. gen lfare_2 = l2.lfare
(2298 missing values generated)

. gen lfare_3 = l3.lfare
(3447 missing values generated)

. reg dlfare_1 lfare_2 dconcen if y99

      Source |       SS       df       MS              Number of obs =    1149
-------------+------------------------------           F(  2,  1146) =  110.84
       Model |  3.63569369     2  1.81784684           Prob > F      =  0.0000
    Residual |  18.7948202  1146  .016400367           R-squared     =  0.1621
-------------+------------------------------           Adj R-squared =  0.1606
       Total |  22.4305139  1148  .019538775           Root MSE      =  .12806

------------------------------------------------------------------------------
   dlfare_1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lfare_2 |  -.1221207   .0082417   -14.82   0.000    -.1382913   -.1059502
     dconcen |  -.1754244   .0544243    -3.22   0.001    -.2822069    -.068642
       _cons |   .6389637   .0417491    15.30   0.000     .5570504    .7208769
------------------------------------------------------------------------------

. predict dlfare_1h99
(option xb assumed; fitted values)
(2298 missing values generated)
```

```
. reg dlfare_1 lfare_2 lfare_3 dconcen if y00

      Source |       SS       df       MS              Number of obs =    1149
-------------+------------------------------           F(  3,  1145) =   11.93
       Model |  .524236952     3  .174745651           Prob > F      =  0.0000
    Residual |  16.7684066  1145  .014644897           R-squared     =  0.0303
-------------+------------------------------           Adj R-squared =  0.0278
       Total |  17.2926436  1148  .015063278           Root MSE      =  .12102

------------------------------------------------------------------------------
    dlfare_1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lfare_2 |  -.1027683   .0278186    -3.69   0.000    -.1573495   -.0481871
     lfare_3 |   .0744738    .025707     2.90   0.004     .0240356     .124912
     dconcen |  -.1971475   .0483136    -4.08   0.000    -.2919407   -.1023543
       _cons |    .155675   .0429415     3.63   0.000     .0714222    .2399278
------------------------------------------------------------------------------

. * No evidence of weak instruments in either time period.
```

```
. predict dlfare_1h00
(option xb assumed; fitted values)
(3447 missing values generated)

. gen dlfare_1h = dlfare_1h99 if y99
(3447 missing values generated)

. replace dlfare_1h = dlfare_1h00 if y00
(1149 real changes made)
```

```
. ivreg dlfare dconcen y00 (dlfare_1 = dlfare_1h), cluster(id)

Instrumental variables (2SLS) regression               Number of obs =    2298
                                                       F(  3,  1148) =   24.03
                                                       Prob > F      =  0.0000
                                                       R-squared     =       .
                                                       Root MSE      =  .12529

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
      dlfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    dlfare_1 |   .2190128   .0619844     3.53   0.000     .0973973    .3406283
     dconcen |   .1262854    .056415     2.24   0.025     .0155974    .2369735
         y00 |    .051385    .006324     8.13   0.000     .0389771    .0637929
       _cons |   .0075111   .0042639     1.76   0.078    -.0008549    .0158771
------------------------------------------------------------------------------
Instrumented:  dlfare_1
Instruments:   dconcen y00 dlfare_1h
------------------------------------------------------------------------------

. * With FDIV, both the lag and the concen variable are positive and
. * statistically significant.
```

```
. * Now use the Arellano and Bond generalized method of moments approach.

. xtabond lfare concen y99 y00

Arellano-Bond dynamic panel-data estimation  Number of obs         =      2298
Group variable: id                           Number of groups      =      1149
Time variable: year
                                             Obs per group:    min =         2
                                                               avg =         2
                                                               max =         2

Number of instruments =      7               Wald chi2(4)          =    441.62
                                             Prob > chi2           =    0.0000
One-step results
------------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfare |
         L1. |   .3326355   .0548124     6.07   0.000     .2252051    .4400659
      concen |   .1519406   .0399507     3.80   0.000     .0736386    .2302425
         y99 |   .0051715   .0041216     1.25   0.210    -.0029066    .0132496
         y00 |   .0629313   .0043475    14.48   0.000     .0544103    .0714523
       _cons |   3.304619   .2820506    11.72   0.000      2.75181    3.857428
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/.).lfare
        Standard: D.concen D.y99 D.y00
Instruments for level equation
        Standard: _cons
```