New Developments in Econometrics
Lecture 3: Linear Panel Data Models I

Jeff Wooldridge

Cemmap Lectures, UCL, June 2009

1. Overview of the Basic Model
2. New Insights Into Old Estimators
3. Behavior of Estimators without Strict Exogeneity
4. IV Estimation under Sequential Exogeneity

1. Overview of the Basic Model

- Unless stated otherwise, the methods discussed in these slides are for the case with a large cross section and small time series, although some approximations are based on T increasing.
- For a generic i in the population,

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

where η_t is a separate time period intercept (parameters that can be estimated), $\mathbf{x}_{it}$ is a $1 \times K$ vector of explanatory variables, c_i is the time-constant unobserved effect, and the $\{u_{it} : t = 1, \ldots, T\}$ are idiosyncratic errors. We view the c_i as random draws along with the observed variables.

• An attractive assumption (but, of course, not universally applicable) is *contemporaneous exogeneity conditional on* c_i:

$$E(u_{it}|\mathbf{x}_{it}, c_i) = 0,\ t = 1, \ldots, T. \tag{2}$$

This equation defines $\boldsymbol{\beta}$ in the sense that under (1) and (2),

$$E(y_{it}|\mathbf{x}_{it}, c_i) = \eta_t + \mathbf{x}_{it}\boldsymbol{\beta} + c_i, \tag{3}$$

so the β_j are partial effects holding c_i fixed.

• Unfortunately, $\boldsymbol{\beta}$ is not identified only under (2). If we add the strong assumption $Cov(\mathbf{x}_{it}, c_i) = \mathbf{0}$, then $\boldsymbol{\beta}$ is identified.

• Allow any correlation between $\mathbf{x}_{it}$ and c_i by assuming *strict exogeneity conditional on* c_i :

$$E(u_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}, c_i) = 0, t = 1, \ldots, T, \tag{4}$$

which can be expressed as

$$E(y_{it}|\mathbf{x}_i, c_i) = E(y_{it}|\mathbf{x}_{it}, c_i) = \eta_t + \mathbf{x}_{it}\boldsymbol{\beta} + c_i. \tag{5}$$

If $\{\mathbf{x}_{it} : t = 1, \ldots, T\}$ has suitable time variation, $\boldsymbol{\beta}$ can be consistently estimated by fixed effects (FE) or first differencing (FD), or generalized least squares (GLS) or generalized method of moments (GMM) versions of them.

• The fixed effects estimator uses the deviations from time averages to remove c_i (put time dummies in $\mathbf{x}_{it}$):

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{u}_{it}, \quad t = 1, \ldots, T, \tag{6}$$

where $\ddot{y}_{it} = y_{it} - T^{-1}\sum_{r=1}^{T} y_{ir}$, and so on. FE is pooled OLS on (6).

• Canned packages provide proper standard errors and inference (with the proper "degrees-of-freedom" adjustment). But the "usual" inference assumes homoskedasticity and serial independence in $\{u_{it}\}$.

• Make inference fully robust to heteroksedasticity and serial dependence. With large N and small T, there is little excuse not to compute "cluster robust" standard errors.

• Treating the c_i as parameters to estimate can lead to trouble even in the linear model: an attempt to "cluster" the standard errors to allow arbitrary serial correlation leads to meaningless standard errors for the $\hat{c}_i = \bar{y}_i - \bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}$.

• An alternative way to remove c_i is to first difference:

$$\Delta y_{it} = \Delta \mathbf{x}_{it}\boldsymbol{\beta} + \Delta u_{it}, t = 2, \ldots, T. \quad (7)$$

The FD estimator is pooled OLS on the first differences.

• FD also requires a kind of strict exogeneity, namely, that u_{it} is uncorrelated with $\mathbf{x}_{i,t-1}, \mathbf{x}_{it}$, and $\mathbf{x}_{i,t+1}$. Failure of strict exogeneity will cause different inconsistencies in FE and FD when $T > 2$.

- Should make inference robust to serial correlation and heteroskedasticity in the differenced errors, $e_{it} \equiv u_{it} - u_{i,t-1}$. For example, if $\{u_{it}\}$ is uncorrelated, $Corr(e_{it}, e_{i,t+1}) = -.5$.
- In unbalanced cases, FD requires that data exists in adjacent time periods. FE does not.
- Even with FE and FD, have to worry about violations of strict exogeneity: strict exogeneity is always violated if $\mathbf{x}_{it}$ contains lagged dependent variables, but also if changes in u_{it} cause changes in $\mathbf{x}_{i,t+1}$ ("feedback effect").

- *Sequential exogeneity condition on* c_i:

$$E(u_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{it}, c_i) = 0, \; t = 1, \ldots, T \tag{8}$$

or, maintaining the linear model,

$$E(y_{it}|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{it}, c_i) = E(y_{it}|\mathbf{x}_{it}, c_i). \tag{9}$$

Allows for lagged dependent variables and other feedback.

- Sequential exogeneity imposes a certain form of correct dynamics, but does not rule out feedback from shocks to y_{it} to $\mathbf{x}_{i,t+1}$.
- If $\mathbf{x}_{it}$ contains $y_{i,t-1}$ (and perhaps other variables $\mathbf{z}_{it}$ and lags), sequential exogeneity rules out serial correlation in $\{u_{it}\}$; otherwise, it
- If, say, $\mathbf{x}_{it} = (\mathbf{z}_{it}, \mathbf{z}_{i,t-1}, \ldots, \mathbf{z}_{i,t-Q})$ then sequential exogeneity implies correct distributed lag dynamics, but allows shocks u_{it} to be correlated with $\mathbf{z}_{i,t+1}$, and $\{u_{it}\}$ can be serially correlated.
- Generally, $\boldsymbol{\beta}$ is identified under sequential exogeneity. (More later.)

- The key “random effects” assumption is

$$E(c_i|\mathbf{x}_i) = E(c_i). \tag{10}$$

- RE leaves c_i in the error term and accounts for the serial correlation in the composite error, $c_i + u_{it}$, via generalized least squares. The nominal assumption is homoskedasticity and serial independence in $\{u_{it}\}$. But RE inference can also be made fully robust to violations of this assumption.
- Can show RE can be computed as a pooled OLS estimator on quasi-time-demeaned data:

$$y_{it} - \lambda \bar{y}_i = (\mathbf{x}_{it} - \lambda \bar{\mathbf{x}}_i)\boldsymbol{\beta} + v_{it} - \lambda \bar{v}_i \tag{11}$$

where $v_{it} = c_i + u_{it}$ and

$$\lambda = 1 - \left[\frac{1}{1 + T(\sigma_c^2/\sigma_u^2)} \right]^{1/2}, \tag{12}$$

• RE can be close to FE with large T or when σ_c^2/σ_u^2 is large, or when $E(c_i|\mathbf{x}_i) = E(c_i)$.

• RE has some advantages over FE and FD: (a) RE allows inclusion of time-constant variables; (b) It can it substantially more efficient than FE. But, it is inconsistent without $E(c_i|\mathbf{x}_i) = E(c_i)$ (or at least zero correlation).

• It is useful to define two *correlated random effects* (CRE) assumptions. The first just defines c_i as a linear function of all of the covariates (linear projection):

$$L(c_i|\mathbf{x}_i) = \psi + \mathbf{x}_i\boldsymbol{\xi} \tag{13}$$

where $\mathbf{x}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT})$. This is often called the *Chamberlain device*, after Chamberlain (1982).

• Mundlak (1978) used a restricted version

$$E(c_i|\mathbf{x}_i) = \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi}, \tag{14}$$

where $\bar{\mathbf{x}}_i = T^{-1}\sum_{t=1}^{T}\mathbf{x}_{it}$.

- Writing $c_i = \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i$ we get the Mundlak equation

$$y_{it} = \eta_t + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i + u_{it}, \tag{15}$$

and we can apple pooled OLS or RE because $E(a_i + u_{it}|\mathbf{x}_i) = 0$. Both equal the FE estimator of $\boldsymbol{\beta}$.

- Can add time-constant covariates, say $\mathbf{z}_i\boldsymbol{\gamma}$, to (15). POLS and RE estimates of $\boldsymbol{\beta}$ are still fixed effects.

- Assumptions such as $D(c_i|\mathbf{x}_i) = D(c_i|\bar{\mathbf{x}}_i)$ are very convenient for nonlinear models.

• The Mundlak equation makes it easy to compute a fully robust Hausman test comparing RE and FE. Separate covariates into aggregate time effects, time-constant variables, and variables that change across i and t:

$$y_{it} = \mathbf{g}_t\boldsymbol{\eta} + \mathbf{z}_i\boldsymbol{\gamma} + \mathbf{w}_{it}\boldsymbol{\delta} + c_i + u_{it}. \tag{16}$$

We cannot estimate $\boldsymbol{\gamma}$ by FE, so it is not part of the Hausman test comparing RE and FE. Less clear is that $\boldsymbol{\eta}$, the coefficients on the time effects, cannot be included, either. (RE and FE estimation only with aggregate time effects are identical.) We can only compare $\hat{\boldsymbol{\delta}}_{FE}$ and $\hat{\boldsymbol{\delta}}_{RE}$ (M parameters).

- Convenient test:

$$y_{it} \text{ on } \mathbf{g}_t, \mathbf{z}_i, \mathbf{w}_{it}, \bar{\mathbf{w}}_i, \; t = 1, \ldots, T; \; i = 1, \ldots, N, \tag{17}$$

which makes it clear there are M restrictions to test. Pooled OLS or RE, fully robust inference!

- Regression (17) can also be used to estimate coefficients on $\mathbf{z}_i$ while allowing correlation between c_i and $\bar{\mathbf{w}}_i$. (We should use caution in interpreting the coefficients on $\mathbf{z}_i$).

- Be wary of canned Hausman test procedures that directly compare estimates: the df are often wrong (the aggregate time variables are counted) and the tests are nonrobust. Can get negative statistics, too.

EXAMPLE: For $N = 1,149$ U.S. air routes and the years 1997 through 2000, y_{it} is $\log(fare_{it})$ and the key explanatory variable is $concen_{it}$, the concentration ratio for route i. Other covariates are year dummies and the time-constant variables $\log(dist_i)$ and $[\log(dist_i)]^2$. Called AIRFARE.DTA.

```
. sum fare concen dist

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        fare |      4596    178.7968    74.88151         37        522
      concen |      4596    .6101149     .196435      .1605          1
        dist |      4596     989.745    611.8315         95       2724
```

```
. reg lfare concen ldist ldistsq y98 y99 y00

      Source |       SS       df       MS              Number of obs =    4596
-------------+------------------------------           F(  6,  4589) =  523.
       Model |  355.453858     6  59.2423096           Prob > F      =  0.0000
    Residual |  519.640516  4589  .113236112           R-squared     =  0.4062
-------------+------------------------------           Adj R-squared =  0.4054
       Total |  875.094374  4595  .190444913           Root MSE      =  .33651

------------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
      concen |   .3601203   .0300691    11.98   0.000     .3011705    .4190702
       ldist |  -.9016004    .128273    -7.03   0.000    -1.153077   -.6501235
     ldistsq |   .1030196   .0097255    10.59   0.000     .0839529    .1220863
         y98 |   .0211244   .0140419     1.50   0.133    -.0064046    .0486533
         y99 |   .0378496   .0140413     2.70   0.007      .010322    .0653772
         y00 |     .09987   .0140432     7.11   0.000     .0723385    .1274015
       _cons |   6.209258   .4206247    14.76   0.000     5.384631    7.033884
------------------------------------------------------------------------------
```

```
. reg lfare concen ldist ldistsq y98 y99 y00, cluster(id)

                                   (Std. Err. adjusted for 1149 clusters in id
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
      concen |   .3601203    .058556     6.15   0.000     .2452315    .4750092
       ldist |  -.9016004   .2719464    -3.32   0.001    -1.435168   -.3680328
     ldistsq |   .1030196   .0201602     5.11   0.000     .0634647    .1425745
         y98 |   .0211244   .0041474     5.09   0.000     .0129871    .0292617
         y99 |   .0378496   .0051795     7.31   0.000     .0276872     .048012
         y00 |     .09987   .0056469    17.69   0.000     .0887906    .1109493
       _cons |   6.209258   .9117551     6.81   0.000     4.420364    7.998151
------------------------------------------------------------------------------
```

```
. xtreg lfare concen ldist ldistsq y98 y99 y00, re

Random-effects GLS regression                   Number of obs      =      4596
Group variable: id                              Number of groups   =      1149

R-sq:  within  = 0.1348                         Obs per group: min =
       between = 0.4176                                        avg =       4
       overall = 0.4030                                        max =

Random effects u_i ~Gaussian                   Wald chi2(6)       =   1360.42
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

----------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval
-------------+--------------------------------------------------------------
      concen |   .2089935   .0265297     7.88   0.000     .1569962    .2609907
       ldist |  -.8520921   .2464836    -3.46   0.001    -1.335191   -.3689931
     ldistsq |   .0974604   .0186358     5.23   0.000     .0609348     .133986
         y98 |   .0224743   .0044544     5.05   0.000     .0137438    .0312047
         y99 |   .0366898   .0044528     8.24   0.000     .0279626    .0454171
         y00 |    .098212   .0044576    22.03   0.000     .0894752    .1069487
       _cons |   6.222005   .8099666     7.68   0.000       4.6345     7.80951
-------------+--------------------------------------------------------------
     sigma_u |  .31933841
     sigma_e |  .10651186
         rho |  .89988885   (fraction of variance due to u_i)
----------------------------------------------------------------------------
```

```
. xtreg lfare concen ldist ldistsq y98 y99 y00, re cluster(id)

                                  (Std. Err. adjusted for 1149 clusters in id
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval
-------------+----------------------------------------------------------------
      concen |   .2089935   .0422459     4.95   0.000      .126193    .2917939
       ldist |  -.8520921   .2720902    -3.13   0.002    -1.385379   -.3188051
     ldistsq |   .0974604   .0201417     4.84   0.000     .0579833    .1369375
         y98 |   .0224743   .0041461     5.42   0.000      .014348    .0306005
         y99 |   .0366898   .0051318     7.15   0.000     .0266317     .046748
         y00 |    .098212   .0055241    17.78   0.000     .0873849     .109039
       _cons |   6.222005   .9144067     6.80   0.000     4.429801    8.014209
-------------+----------------------------------------------------------------
     sigma_u |  .31933841
     sigma_e |  .10651186
         rho |  .89988885   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
. xtreg lfare concen ldist ldistsq y98 y99 y00, fe

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

R-sq:  within  = 0.1352                         Obs per group: min =
       between = 0.0576                                        avg =       4
       overall = 0.0083                                        max =

                                                F(4,3443)          =    134.
corr(u_i, Xb)  = -0.2033                        Prob > F           =    0.0000

------------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
      concen |    .168859   .0294101     5.74   0.000     .1111959     .226522
       ldist |  (dropped)
     ldistsq |  (dropped)
         y98 |   .0228328   .0044515     5.13   0.000     .0141048    .0315607
         y99 |   .0363819   .0044495     8.18   0.000     .0276579    .0451058
         y00 |   .0977717   .0044555    21.94   0.000      .089036    .1065073
       _cons |   4.953331   .0182869   270.87   0.000     4.917476    4.989185
-------------+----------------------------------------------------------------
     sigma_u |  .43389176
     sigma_e |  .10651186
         rho |  .94316439   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(1148, 3443) =    36.90          Prob > F = 0.0000
```

```
. xtreg lfare concen ldist ldistsq y98 y99 y00, fe cluster(id)

                                  (Std. Err. adjusted for 1149 clusters in id
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
      concen |    .168859   .0494587     3.41   0.001     .0718194    .2658985
       ldist |  (dropped)
     ldistsq |  (dropped)
         y98 |   .0228328    .004163     5.48   0.000     .0146649    .0310007
         y99 |   .0363819   .0051275     7.10   0.000     .0263215    .0464422
         y00 |   .0977717   .0055054    17.76   0.000     .0869698    .1085735
       _cons |   4.953331   .0296765   166.91   0.000     4.895104    5.011557
-------------+----------------------------------------------------------------
     sigma_u |  .43389176
     sigma_e |  .10651186
         rho |  .94316439   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
. egen concenbar = mean(concen), by(id)

. xtreg lfare concen concenbar ldist ldistsq y98 y99 y00, re cluster(id)

                                  (Std. Err. adjusted for 1149 clusters in id
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval
-------------+----------------------------------------------------------------
      concen |    .168859   .0494749     3.41   0.001       .07189    .2658279
   concenbar |   .2136346   .0816403     2.62   0.009     .0536227    .3736466
       ldist |  -.9089297   .2721637    -3.34   0.001    -1.442361   -.3754987
     ldistsq |   .1038426   .0201911     5.14   0.000     .0642688    .1434164
         y98 |   .0228328   .0041643     5.48   0.000     .0146708    .0309947
         y99 |   .0363819   .0051292     7.09   0.000     .0263289    .0464349
         y00 |   .0977717   .0055072    17.75   0.000     .0869777    .1085656
       _cons |   6.207889   .9118109     6.81   0.000     4.420773    7.995006
-------------+----------------------------------------------------------------
     sigma_u |  .31933841
     sigma_e |  .10651186
         rho |  .89988885   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

- Can combine IV methods with unobserved effects. Allow contemporaneous endogeneity (correlation between $\mathbf{x}_{it}$ and u_{it}, in addition to correlation between $\mathbf{x}_{it}$ and c_i).
- Let $\ddot{\mathbf{z}}_{it} = \mathbf{z}_{it} - \bar{\mathbf{z}}_i$ be time-demeaned instruments. Apply IV methods to

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{u}_{it},\ t = 1, \ldots, T, \tag{18}$$

such as pooled 2SLS. (Again, use fully robust inference: xtivreg2 in Stata.) Can also apply generalized method of moments to the demeaned equation to enhance efficiency.

2. New Insights Into Old Estimators

• Consider an extension of the usual model to allow for unit-specific slopes,

$$y_{it} = c_i + \mathbf{x}_{it}\mathbf{b}_i + u_{it} \tag{19}$$

$$E(u_{it}|\mathbf{x}_i, c_i, \mathbf{b}_i) = 0, t = 1, \ldots, T, \tag{20}$$

where $\mathbf{b}_i$ is $K \times 1$. We act as if $\mathbf{b}_i$ is constant for all i but think c_i might be correlated with $\mathbf{x}_{it}$; we apply usual FE estimator. When does the usual FE estimator consistently estimate the population average effect, $\boldsymbol{\beta} = E(\mathbf{b}_i)$?

• A sufficient condition for consistency of the FE estimator, along with $E(u_{it}|\mathbf{x}_i, c_i) = 0$ and the usual FE rank condition, is

$$E(\mathbf{b}_i|\ddot{\mathbf{x}}_{it}) = E(\mathbf{b}_i) = \boldsymbol{\beta}, \quad t = 1, \ldots, T \tag{21}$$

where $\ddot{\mathbf{x}}_{it}$ are the time-demeaned covariates. Allows the slopes, $\mathbf{b}_i$, to be correlated with the regressors $\mathbf{x}_{it}$ through permanent components. For example, if $\mathbf{x}_{it} = \mathbf{f}_i + \mathbf{r}_{it}, t = 1, \ldots, T$. Then (21) holds if $E(\mathbf{b}_i|\mathbf{r}_{i1}, \mathbf{r}_{i2}, \ldots, \mathbf{r}_{iT}) = E(\mathbf{b}_i)$.

• Extends to a more general class of FE estimators. Write

$$y_{it} = \mathbf{w}_t \mathbf{a}_i + \mathbf{x}_{it} \mathbf{b}_i + u_{it}, \quad t = 1, \ldots, T \tag{22}$$

where $\mathbf{w}_t$ is a set of deterministic functions of time. Now FE refers to sweeping away $\mathbf{a}_i$ by netting out $\mathbf{w}_t$ from $\mathbf{x}_{it}$.

• In the random trend model, $\mathbf{w}_t = (1, t)$, and now the elements of $\ddot{\mathbf{x}}_{it}$ have had unit-specific linear trends removed. If $\mathbf{x}_{it} = \mathbf{f}_i + \mathbf{h}_i t + \mathbf{r}_{it}$, then $\mathbf{b}_i$ can be arbitrarily correlated with $(\mathbf{f}_i, \mathbf{h}_i)$.

• Generally, need $\dim(\mathbf{w}_t) < T$.

- Can apply to models with time-varying factor loads, η_t :

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \eta_t c_i + u_{it}, t = 1, \ldots, T. \tag{23}$$

Sufficient for consistency of FE estimator that ignores the η_t is

$$Cov(\ddot{\mathbf{x}}_{it}, c_i) = \mathbf{0}, t = 1, \ldots, T. \tag{24}$$

- Can use nonlinear GMM to estimate $\boldsymbol{\beta}$ along with the η_t; estimates are often similar to FE.

• FE-IV methods also have some robustness in the general random slopes model

$$y_{it} = \mathbf{w}_t \mathbf{a}_i + \mathbf{x}_{it} \mathbf{b}_i + u_{it}, \quad t = 1, \ldots, T.$$

The slopes $\mathbf{b}_i$ can be correlated with $\mathbf{x}_{it}$.

• Assume the instruments are strictly exogenous:

$$E(u_{it} | \mathbf{z}_i, \mathbf{a}_i, \mathbf{b}_i) = 0. \tag{25}$$

• Along with

$$E(\mathbf{b}_i|\ddot{\mathbf{z}}_{it}) = E(\mathbf{b}_i) = \boldsymbol{\beta}, \quad t = 1, \ldots, T. \tag{26}$$

also assume

$$Cov(\ddot{\mathbf{x}}_{it}, \mathbf{b}_i|\ddot{\mathbf{z}}_{it}) = Cov(\ddot{\mathbf{x}}_{it}, \mathbf{b}_i), t = 1, \ldots, T. \tag{27}$$

The $K \times K$ matrix unconditional covariance matrix, $Cov(\ddot{\mathbf{x}}_{it}, \mathbf{b}_i)$, is unrestricted. The *conditional* covariance cannot depend on the $\ddot{\mathbf{z}}_{it}$. Then, FEIV is consistent for $\boldsymbol{\beta} = E(\mathbf{b}_i)$ provided a full set of time dummies is included.

• Assumption (27) can hold for continuous $\mathbf{x}_{it}$ but is unrealistic when endogenous elements of $\mathbf{x}_{it}$ are discrete.

• Simple test to see whether the slopes $\mathbf{b}_i$ depend on observed factors, say, $\mathbf{w}_i$ (which do not change over time) and $\bar{\mathbf{x}}_i$ (the time averages of time-varying covariates):

$$y_{it} = \eta_t + \mathbf{x}_{it}\mathbf{b}_i + c_i + u_{it} \tag{28}$$

$$c_i = \alpha + (\mathbf{h}_i - \boldsymbol{\mu}_{\mathbf{h}})\boldsymbol{\gamma} + a_i \tag{29}$$

$$\mathbf{b}_i = \boldsymbol{\beta} + \boldsymbol{\Pi}(\mathbf{h}_i - \boldsymbol{\mu}_{\mathbf{h}})' + \mathbf{d}_i \tag{30}$$

where $\mathbf{h}_i = (\bar{\mathbf{x}}_i, \mathbf{w}_i)$ (a row vector) and $E(a_i|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},\mathbf{w}_i) = 0$, $E(\mathbf{d}_i|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},\mathbf{w}_i) = \mathbf{0}$.

- After some algebra,

$$
\begin{aligned}
y_{it} = {} & \alpha_t + (\mathbf{h}_i - \boldsymbol{\mu}_{\mathbf{h}})\boldsymbol{\gamma} + \mathbf{x}_{it}\boldsymbol{\beta} + [(\mathbf{h}_i - \boldsymbol{\mu}_{\mathbf{h}}) \otimes \mathbf{x}_{it}]\boldsymbol{\pi} \\
& + a_i + \mathbf{x}_{it}\mathbf{d}_i + u_{it},
\end{aligned} \tag{31}
$$

which just says interact elements of $\mathbf{h}_i - \boldsymbol{\mu}_{\mathbf{h}}$ with elements of $\mathbf{x}_{it}$. (Can be selective, of course.)

- In practice, replace $\boldsymbol{\mu}_{\mathbf{h}}$ with $\bar{\mathbf{h}}$ (sample average) and use pooled OLS.

- Can even estimate (31) by FE, which removes $(\mathbf{h}_i - \boldsymbol{\mu}_{\mathbf{h}})$ but not the interactions. The interactions might be significant even though the estimates of $\boldsymbol{\beta}$ (population averaged effect) might be similar to FE on the equation without the interactions, $[(\mathbf{h}_i - \boldsymbol{\mu}_{\mathbf{h}}) \otimes \mathbf{x}_{it}]$.
- Can use random effects, too, although it would ignore $\mathbf{x}_{it}\mathbf{d}_i$ (so, at a minimum, inference should be made fully robust).
- A GLS estimator that accounts for $\mathbf{x}_{it}\mathbf{d}_i$ is possible (but may not be worth it if want mean effects).

EXAMPLE: Airfare Application.

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

. sum concenb ldist ldistsq

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     concenb |      4596    .6101149    .1888741      .1862      .9997
       ldist |      4596    6.696482    .6593177   4.553877   7.909857
     ldistsq |      4596    45.27747    8.726898   20.73779   62.56583

. gen cbconcen = (concenb - .61)*concen
. gen ldconcen = (ldist - 6.696)*concen
. gen ldsqconcen = (ldistsq - 45.277)*concen
```

```
. xtreg lfare concen concenb cbconcen ldconcen ldsqconcen ldist ldistsq
      y98 y99 y00, re cluster(id)

                                  (Std. Err. adjusted for 1149 clusters in id
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval
-------------+----------------------------------------------------------------
      concen |   .1682492   .0496695     3.39   0.001      .0708988    .2655996
     concenb |    .157291   .2085049     0.75   0.451     -.2513711     .565953
    cbconcen |   .0635453   .3033809     0.21   0.834     -.5310704    .6581609
    ldconcen |  -.2994869   .9930725    -0.30   0.763     -2.245873    1.646899
  ldsqconcen |   .0112477   .0746874     0.15   0.880      -.135137    .1576324
       ldist |  -.4394368   .6713288    -0.65   0.513     -1.755217    .8763435
     ldistsq |   .0752147   .0494201     1.52   0.128     -.0216469    .1720764
         y98 |   .0229684   .0041542     5.53   0.000      .0148262    .0311105
         y99 |   .0358549   .0051298     6.99   0.000      .0258007    .0459091
         y00 |   .0976256    .005461    17.88   0.000      .0869221     .108329
       _cons |   4.382552   2.272566     1.93   0.054     -.0715953    8.836699
-------------+----------------------------------------------------------------
. * Nonrobust se for concen is .0295.
. test cbconcen ldconcen ldsqconcen

 ( 1)  cbconcen = 0
 ( 2)  ldconcen = 0
 ( 3)  ldsqconcen = 0

           chi2(  3) =    5.47
         Prob > chi2 =    0.1407
```

3. Behavior of Estimators without Strict Exogeneity

• Both the FE and FD estimators are inconsistent (with fixed T, $N \to \infty$) without the strict exogeneity assumption. But inconsistencies (as function of T) can differ.

• If we maintain $E(u_{it}|\mathbf{x}_{it}, c_i) = 0$ and assume $\{(\mathbf{x}_{it}, u_{it}) : t = 1, \ldots, T\}$ is "weakly dependent", can show

$$\text{plim}_{N\to\infty}\, \hat{\boldsymbol{\beta}}_{FE} = \boldsymbol{\beta} + O(T^{-1}) \tag{32}$$

$$\text{plim}_{N\to\infty}\, \hat{\boldsymbol{\beta}}_{FD} = \boldsymbol{\beta} + O(1). \tag{33}$$

• Still holds if $\{\mathbf{x}_{it} : t = 1, \ldots, T\}$ has unit roots as long as $\{u_{it}\}$ is I(0) and contemporaneous exogeneity holds.

• Important caveat: if $\{u_{it}\}$ is I(1) – so that the time series "model" is a spurious regression (y_{it} and $\mathbf{x}_{it}$ are not *cointegrated*), then (32) is no longer true. FD is attractive because it eliminates any unit roots.

• Same conclusions hold for instrumental variables versions: FE has bias of order T^{-1} if instruments are contemporaneously exogenous and $\{u_{it}\}$ is weakly dependent.

• Simple test for lack of strict exogeneity in covariates:

$$y_{it} = \eta_t + \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{w}_{i,t+1}\boldsymbol{\delta} + c_i + e_{it} \tag{34}$$

Estimate the equation by fixed effects and test $H_0 : \boldsymbol{\delta} = \mathbf{0}$.

```
. sort id year
. gen concenp1 = concen[_n+1] if year < 2000
. xtreg lfare concen concenp1 y98 y99 y00, fe cluster(id)

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

R-sq:  within  = 0.0558                         Obs per group: min =
       between = 0.0535                                        avg =        3
       overall = 0.0347                                        max =

                                                F(4,1148)          =      25.
corr(u_i, Xb)  = -0.2949                        Prob > F           =    0.0000

                                  (Std. Err. adjusted for 1149 clusters in id
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
      concen |   .2983988    .054797     5.45   0.000     .1908854    .4059122
    concenp1 |  -.0659259   .0467578    -1.41   0.159    -.1576663    .0258145
         y98 |   .0205809   .0042341     4.86   0.000     .0122735    .0288883
         y99 |   .0360638   .0050754     7.11   0.000     .0261058    .0460218
         y00 |  (dropped)
       _cons |   4.914953   .0478488   102.72   0.000     4.821072    5.008834
-------------+----------------------------------------------------------------
```

- Test contemporaneous endogeneity of a subset of certain regressors.

$$y_{it1} = \mathbf{z}_{it1}\boldsymbol{\delta}_1 + \mathbf{y}_{it2}\boldsymbol{\alpha}_1 + \mathbf{y}_{it3}\boldsymbol{\gamma}_1 + c_{i1} + u_{it1}, \tag{35}$$

where, in an FE environment, we want to test $H_0 : E(\mathbf{y}_{it3}' u_{it1}) = \mathbf{0}$.

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

$$\mathbf{y}_{it3} = \mathbf{z}_{it}\boldsymbol{\Pi}_3 + \mathbf{c}_{i3} + \mathbf{v}_{it3}. \tag{36}$$

• Obtain FE residuals, $\hat{\mathbf{v}}_{it3} = \mathbf{y}_{it3} - \mathbf{z}_{it}\hat{\mathbf{\Pi}}_3$ ($\hat{\mathbf{\Pi}}_3$ FE estimates). Estimate

$$y_{it1} = \mathbf{z}_{it1}\boldsymbol{\delta}_1 + \mathbf{y}_{it2}\boldsymbol{\alpha}_1 + \mathbf{y}_{it3}\boldsymbol{\gamma}_1 + \hat{\mathbf{v}}_{it3}\boldsymbol{\rho}_1 + error_{it1} \quad (37)$$

by Fixed Effects Instrumental Variables, using instruments $(\mathbf{z}_{it}, \mathbf{y}_{it3}, \hat{\mathbf{v}}_{it3})$. To test the null that $\mathbf{y}_{it3}$ is exogenous, use (robust) test that $\boldsymbol{\rho}_1 = \mathbf{0}$; need not adjust for the first-step estimation.

Airfare Example: Passenger Demand Function

```
. xtreg lpassen lfare y98 y99 y00, fe cluster(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
-------------+----------------------------------------------------------------
```

```
. qui areg lfare concen y98 y99 y00, absorb(id)

. predict v2h, resid

. xtreg lpassen lfare y98 y99 y00 v2h, fe cluster(id)

------------------------------------------------------------------------------
             |               Robust
     lpassen |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
       lfare |   -.301576   .4829734    -0.62   0.532    -1.249185    .6460335
         y98 |   .0257147   .0131382     1.96   0.051    -.0000628    .0514923
         y99 |   .0724165   .0197133     3.67   0.000     .0337385    .1110946
         y00 |   .1127914    .048597     2.32   0.020     .0174425    .2081403
         v2h |  -.8616344   .5278388    -1.63   0.103    -1.897271    .1740025
       _cons |   7.501007   2.441322     3.07   0.002     2.711055    12.29096
-------------+----------------------------------------------------------------
```

4. IV Estimation under Sequential Exogeneity

We now consider IV estimation of the model

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it},\ t = 1,\ldots,T, \tag{38}$$

under sequential exogeneity assumptions; the weakest form is

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

This leads to simple moment conditions after first differencing:

$$E(\mathbf{x}_{is}'\Delta u_{it}) = \mathbf{0},\ s = 1,\ldots,t-1;\ t = 2,\ldots,T. \tag{39}$$

Therefore, at time t, the available instruments in the FD equation are in the vector $\mathbf{x}_{it}^{o} \equiv (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{it})$. The matrix of instruments is

$$\mathbf{W}_i = diag(\mathbf{x}_{i1}^{o}, \mathbf{x}_{i2}^{o}, \ldots, \mathbf{x}_{i,T-1}^{o}), \tag{40}$$

which has $T - 1$ rows. Routine to apply GMM estimation.

• Simple strategy: 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

$$\Delta y_{it} = \Delta\mathbf{x}_{it}\boldsymbol{\beta} + \Delta u_{it},\ t = 2, \ldots, T \tag{41}$$

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

• Can suffer from a weak instrument problem when $\Delta \mathbf{x}_{it}$ has little correlation with $\mathbf{x}_{i,t-1}^{o}$. But at least we have the $T-1$ RF regressions to study the weak instrument problem.

• If we assume

$$E(u_{it}|\mathbf{x}_{it}, y_{i,t-1}\mathbf{x}_{i,t-1}, \ldots, y_{i1}, \mathbf{x}_{i1}, c_i) = 0, \tag{42}$$

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

$$E[(\Delta y_{i,t-1} - \Delta \mathbf{x}_{i,t-1}\boldsymbol{\beta})'(y_{it} - \mathbf{x}_{it}\boldsymbol{\beta})] = \mathbf{0}. \tag{43}$$

• Drawback: we often do not want to assume (42). Plus, the conditions in (43) are nonlinear in parameters.

- Arellano and Bover (1995) suggested instead the restrictions

$$Cov(\Delta \mathbf{x}_{it}', c_i) = 0, \ \ t = 2, \ldots, T, \tag{44}$$

which imply linear moment conditions in the levels equation,

$$E[\Delta \mathbf{x}_{it}'(y_{it} - \alpha - \mathbf{x}_{it}\boldsymbol{\beta})] = \mathbf{0}, \ t = 2, \ldots, T. \tag{45}$$

- Simple AR(1) model:

$$y_{it} = \rho y_{i,t-1} + c_i + u_{it}, t = 1, \ldots, T. \tag{46}$$

- Typically, the minimal assumptions imposed are

$$E(y_{is}u_{it}) = 0,\ s = 0, \ldots, t-1,\ t = 1, \ldots, T, \tag{47}$$

so for $t = 2, \ldots, T$,

$$E[y_{is}(\Delta y_{it} - \rho \Delta y_{i,t-1}) = 0, s \leq t-2. \tag{48}$$

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

- Blundell and Bond (1998) showed that if the condition

$$Cov(\Delta y_{i1}, c_i) = Cov(y_{i1} - y_{i0}, c_i) = 0 \tag{49}$$

is added to $E(u_{it}|y_{i,t-1}, \ldots, y_{i0}, c_i) = 0$ then

$$E[\Delta y_{i,t-1}(y_{it} - \alpha - \rho y_{i,t-1})] = 0 \tag{50}$$

which can be added to the usual moment conditions (38). We have two sets of moments linear in the parameters.

• The condition $Cov(\Delta y_{i1}, c_i) = 0$ can be intepreted as a restriction on the initial condition, y_{i0}. Write y_{i0} as a deviation from its steady state, $c_i/(1-\rho)$ (obtained for $|\rho| < 1$ by recursive subsitution and then taking the limit), as $y_{i0} = c_i/(1-\rho) + r_{i0}$. Then $(1-\rho)y_{i0} + c_i = (1-\rho)r_{i0}$, and so (49) reduces to

$$Cov(r_{i0}, c_i) = 0. \tag{51}$$

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

• Of course, if $\rho = 1$ there is no steady state.

- Extensions of the AR(1) model, such as

$$y_{it} = \rho y_{i,t-1} + \mathbf{z}_{it}\boldsymbol{\gamma} + c_i + u_{it}, \quad t = 1, \ldots, T \tag{52}$$

and use FD:

$$\Delta y_{it} = \rho \Delta y_{i,t-1} + \Delta \mathbf{z}_{it}\boldsymbol{\gamma} + \Delta u_{it}, \quad t = 1, \ldots, T. \tag{53}$$

- Airfare Example: Dynamic model with a route-specific effect.

Dep. Var.	*lfare*			
	(1)	(2)	(3)	(4)
Expl. Var.	Pooled OLS (FD)	FE	Pooled IV (FD)	Arellano-Bond
$lfare_{-1}$	$-.126$ (.027)	$.077$ (.032)	$.219$ (.062)	$.333$ (.055)
concen	$.076$ (.053)	$.058$ (.053)	$.126$ (.056)	$.152$ (.040)
N	$1,149$	$1,149$	$1,149$	$1,149$