Panel Data Models with Heterogeneity and Endogeneity

Jeff Wooldridge
Michigan State University

Programme Evaluation for Policy Analysis
Institute for Fiscal Studies
June 2012

1. Introduction
2. General Setup and Quantities of Interest
3. Assumptions with Neglected Heterogeneity
4. Models with Heterogeneity and Endogeneity
5. Estimating Some Popular Models

1. Introduction

- When panel data models contain unobserved heterogeneity and omitted time-varying variables, control function methods can be used to account for both problems.
- Under fairly week assumptions can obtain consistent, asymptotically normal estimators of average structural functions – provided suitable instruments are available.
- Other issues with panel data: How to treat dynamics? Models with lagged dependent variables are hard to estimate when heterogeneity and other sources of endogeneity are present.

• Approaches to handling unobserved heterogeneity:

1. Treat as parameters to estimate. Can work well with large T but with small T can have incidental parameters problem. Bias adjustments are available for parameters and average partial effects. Usually weak dependence or even independence is assumed across the time dimension.

2. Remove heterogeneity to obtain an estimating equation. Works for simple linear models and a few nonlinear models (via conditional MLE or a quasi-MLE variant). Cannot be done in general. Also, may not be able to identify interesting partial effects.

- Correlated Random Effects: Mundlak/Chamberlain. Requires some restrictions on distribution of heterogeneity, although these can be nonparametric. Applies generally, does not impose restrictions on dependence over time, allows estimation of average partial effects. Can be easily combined with CF methods for endogeneity.
- Can try to establish bounds rather than estimate parameters or APEs. Chernozhukov, Fernández-Val, Hahn, and Newey (2009) is a recent example.

2. General Setup and Quantities of Interest

- Static, unobserved effects probit model for panel data with an omitted time-varying variable r_{it}:

$$P(y_{it} = 1|\mathbf{x}_{it}, c_i, r_{it}) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta} + c_i + r_{it}),\ t = 1,\ldots,T. \tag{1}$$

What are the quantities of interest for most purposes?

(i) The element of $\boldsymbol{\beta}$, the β_j. These give the directions of the partial effects of the covariates on the response probability. For any two continuous covariates, the ratio of coefficients, β_j/β_h, is identical to the ratio of partial effects (and the ratio does not depend on the covariates or unobserved heterogeneity, c_i).

(ii) The magnitudes of the partial effects. These depend not only on the value of the covariates, say $\mathbf{x}_t$, but also on the value of the unobserved heterogeneity. In the continuous covariate case,

$$\frac{\partial P(y_t = 1|\mathbf{x}_t, c, r_t)}{\partial x_{tj}} = \beta_j \phi(\mathbf{x}_t \boldsymbol{\beta} + c + r_t). \tag{2}$$

• Questions: (a) Assuming we can estimate $\boldsymbol{\beta}$, what should we do about the unobservables (c, r_t)? (b) If we can only estimate $\boldsymbol{\beta}$ up-to-scale, can we still learn something useful about magnitudes of partial effects? (c) What kinds of assumptions do we need to estimate partial effects?

• Let $\{(\mathbf{x}_{it}, y_{it}) : t = 1, \ldots, T\}$ be a random draw from the cross section. Suppose we are interested in

$$E(y_{it}|\mathbf{x}_{it}, \mathbf{c}_i, \mathbf{r}_{it}) = m_t(\mathbf{x}_{it}, \mathbf{c}_i, \mathbf{r}_{it}). \tag{3}$$

$\mathbf{c}_i$ can be a vector of unobserved heterogeneity, $\mathbf{r}_{it}$ a vector of omitted time-varying variables.

• Partial effects: if x_{tj} is continuous, then

$$\theta_j(\mathbf{x}_t, \mathbf{c}, \mathbf{r}_t) \equiv \frac{\partial m_t(\mathbf{x}_t, \mathbf{c}, \mathbf{r}_t)}{\partial x_{tj}}, \tag{4}$$

or discrete changes.

• How do we account for unobserved $(\mathbf{c}_i, \mathbf{r}_{it})$? If we know enough about the distribution of $(\mathbf{c}_i, \mathbf{r}_{it})$ we can insert meaningful values for $(\mathbf{c}, \mathbf{r}_t)$. For example, if $\boldsymbol{\mu}_{\mathbf{c}} = E(\mathbf{c}_i)$, $\boldsymbol{\mu}_{\mathbf{r}_t} = E(\mathbf{r}_{it})$ then we can compute the partial effect at the average (PEA),

$$PEA_j(\mathbf{x}_t) = \theta_j(\mathbf{x}_t, \boldsymbol{\mu}_{\mathbf{c}}, \boldsymbol{\mu}_{\mathbf{r}_t}). \tag{5}$$

Of course, we need to estimate the function m_t and $(\boldsymbol{\mu}_{\mathbf{c}}, \boldsymbol{\mu}_{\mathbf{r}_t})$. If we can estimate the distribution of $(\mathbf{c}_i, \mathbf{r}_{it})$, or features in addition to its mean, we can insert different quantiles, or a certain number of standard deviations from the mean.

• Alternatively, we can obtain the average partial effect (APE) (or population average effect) by averaging across the distribution of $\mathbf{c}_i$:

$$APE(\mathbf{x}_t) = E_{(\mathbf{c}_i,\mathbf{r}_{it})}[\theta_j(\mathbf{x}_t, \mathbf{c}_i, \mathbf{r}_{it})]. \tag{6}$$

The difference between (5) and (6) can be nontrivial. In some leading cases, (6) is identified while (5) is not. (6) is closely related to the notion of the average structural function (ASF) (Blundell and Powell (2003)). The ASF is defined as

$$ASF_t(\mathbf{x}_t) = E_{(\mathbf{c}_i,\mathbf{r}_{it})}[m_t(\mathbf{x}_t, \mathbf{c}_i, \mathbf{r}_{it})]. \tag{7}$$

• Passing the derivative through the expectation in (7) gives the APE.

3. Assumptions with Neglected Heterogeneity

Exogeneity of Covariates

- Cannot get by with just specifying a model for the contemporaneous conditional distribution, $D(y_{it}|\mathbf{x}_{it},\mathbf{c}_i)$.
- The most useful definition of strict exogeneity for nonlinear panel data models is

$$D(y_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},\mathbf{c}_i) = D(y_{it}|\mathbf{x}_{it},\mathbf{c}_i). \tag{8}$$

Chamberlain (1984) labeled (8) *strict exogeneity conditional on the unobserved effects* $\mathbf{c}_i$. Conditional mean version:

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

- The *sequential exogeneity* assumption is

$$D(y_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{it},\mathbf{c}_i) = D(y_{it}|\mathbf{x}_{it},\mathbf{c}_i). \tag{10}$$

Much more difficult to allow sequential exogeneity in in nonlinear models. (Most progress has been made for lagged dependent variables or specific functional forms, such as exponential.)

- Neither strict nor sequential exogeneity allows for contemporaneous endogeneity of one or more elements of $\mathbf{x}_{it}$, where, say, x_{itj} is correlated with unobserved, time-varying unobservables that affect y_{it}.

Conditional Independence

• In linear models, serial dependence of idiosyncratic shocks is easily dealt with, either by "cluster robust" inference or Generalized Least Squares extensions of Fixed Effects and First Differencing. With strictly exogenous covariates, serial correlation never results in inconsistent estimation, even if improperly modeled. The situation is different with most nonlinear models estimated by MLE.

• *Conditional independence* (CI) (under strict exogeneity):

$$D(y_{i1},\ldots,y_{iT}|\mathbf{x}_i,\mathbf{c}_i) = \prod_{t=1}^{T} D(y_{it}|\mathbf{x}_{it},\mathbf{c}_i). \tag{11}$$

• In a parametric context, the CI assumption reduces our task to specifying a model for $D(y_{it}|\mathbf{x}_{it}, \mathbf{c}_i)$, and then determining how to treat the unobserved heterogeneity, $\mathbf{c}_i$.

• In random effects and correlated random frameworks (next section), CI plays a critical role in being able to estimate the “structural” parameters and the parameters in the distribution of $\mathbf{c}_i$ (and therefore, in estimating PEAs). In a broad class of popular models, CI plays no essential role in estimating APEs.

Assumptions about the Unobserved Heterogeneity

Random Effects

- Generally stated, the key RE assumption is

$$D(\mathbf{c}_i|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}) = D(\mathbf{c}_i). \tag{12}$$

Under (12), the APEs are actually nonparametrically identified from

$$E(y_{it}|\mathbf{x}_{it} = \mathbf{x}_t). \tag{13}$$

- In some leading cases (RE probit and RE Tobit with heterogeneity normally distributed), if we want PEs for different values of **c**, we must assume more: strict exogeneity, conditional independence, and (12) with a parametric distribution for $D(\mathbf{c}_i)$.

Correlated Random Effects

A CRE framework allows dependence between $\mathbf{c}_i$ and $\mathbf{x}_i$, but restricted in some way. In a parametric setting, we specify a distribution for $D(\mathbf{c}_i|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT})$, as in Chamberlain (1980,1982), and much work since. Distributional assumptions that lead to simple estimation – homoskedastic normal with a linear conditional mean — can be restrictive.

- Possible to drop parametric assumptions and just assume

$$D(c_i|\mathbf{x}_i) = D(c_i|\bar{\mathbf{x}}_i), \tag{14}$$

without restricting $D(c_i|\bar{\mathbf{x}}_i)$. Altonji and Matzkin (2005, Econometrica).

- Other functions of $\{\mathbf{x}_{it} : t = 1, \ldots, T\}$ are possible.

• APEs are identified very generally. For example, under (14), a consistent estimate of the average structural function is

$$\widehat{ASF}(\mathbf{x}_t) = N^{-1} \sum_{i=1}^{N} q_t(\mathbf{x}_t, \bar{\mathbf{x}}_i), \tag{15}$$

where $q_t(\mathbf{x}_{it}, \bar{\mathbf{x}}_i) = E(y_{it}|\mathbf{x}_{it}, \bar{\mathbf{x}}_i)$.

• Need a random sample $\{\bar{\mathbf{x}}_i : i = 1, \ldots, N\}$ for the averaging out to work.

Fixed Effects

• The label "fixed effects" is used differently by different researchers. One view: $\mathbf{c}_i,\ i = 1,\ldots,N$ are parameters to be estimated. Usually leads to an "incidental parameters problem."

• Second meaning of "fixed effects": $D(\mathbf{c}_i|\mathbf{x}_i)$ is unrestricted and we look for objective functions that do not depend on $\mathbf{c}_i$ but still identify the population parameters. Leads to "conditional MLE" if we can find "sufficient statistics" $\mathbf{s}_i$ such that

$$D(y_{i1},\ldots,y_{iT}|\mathbf{x}_i,\mathbf{c}_i,\mathbf{s}_i) = D(y_{i1},\ldots,y_{iT}|\mathbf{x}_i,\mathbf{s}_i). \tag{16}$$

• Conditional Independence is usually maintained.

• Key point: PEAs and APEs are generally unidentified.

4. Models with Heterogeneity and Endogeneity

- Let y_{it1} be a scalar response, $\mathbf{y}_{it2}$ a vector of endogenous variables, $\mathbf{z}_{it1}$ exogenous variables, and we have

$$E(y_{it1}|\mathbf{y}_{it2}, \mathbf{z}_{it1}, \mathbf{c}_{i1}, \mathbf{r}_{it1}) = m_{t1}(\mathbf{y}_{it2}, \mathbf{z}_{it1}, \mathbf{c}_{i1}, \mathbf{r}_{it1}) \tag{17}$$

- $\mathbf{y}_{it2}$ is allowed to be correlated with $\mathbf{r}_{it1}$ (as well as with $\mathbf{c}_{i1}$).
- The vector of exogenous variables $\{\mathbf{z}_{it} : t = 1, \ldots, T\}$ with $\mathbf{z}_{it1} \subset \mathbf{z}_{it}$ are strictly exogenous in the sense that

$$E(y_{it}|\mathbf{y}_{it2}, \mathbf{z}_{i}, \mathbf{c}_{i1}, \mathbf{r}_{it1}) = E(y_{it}|\mathbf{y}_{it2}, \mathbf{z}_{it1}, \mathbf{c}_{i1}, \mathbf{r}_{it1}) \tag{18}$$

$$D(\mathbf{r}_{it1}|\mathbf{z}_{i}, \mathbf{c}_{i1}) = D(\mathbf{r}_{it1}) \tag{19}$$

- Sometimes we can eliminate $\mathbf{c}_i$ and obtain an equation that can be estimated by IV (linear, exponential). Generally not possible.
- Now a CRE approach involves modeling $D(\mathbf{c}_{i1}|\mathbf{z}_i)$.
- Generally, we need to model how $\mathbf{y}_{it2}$ is related to $\mathbf{r}_{it1}$.
- Control Function methods are convenient for allowing both.
- Suppose y_{it2} is a scalar and

$$\begin{aligned} y_{it2} &= m_{it2}(\mathbf{z}_{it}, \bar{\mathbf{z}}_i, \boldsymbol{\delta}_2) + v_{it2} \\ E(v_{it2}|\mathbf{z}_i) &= 0 \\ D(r_{it1}|v_{it2}, \mathbf{z}_i) &= D(r_{it1}|v_{it2}) \end{aligned} \tag{20}$$

• With suitable time-variation in the instruments, the assumptions in (20) allow identification of the ASF if we assume a model for

$$D(\mathbf{c}_{i1}|\mathbf{z}_i, v_{it2})$$

Generally, we can estimate

$$E(y_{it1}|y_{it2}, \mathbf{z}_i, v_{it2}) = E(y_{it1}|y_{it2}, \mathbf{z}_{it1}, \bar{\mathbf{z}}_i, v_{it2}) \equiv g_{t1}(y_{it2}, \mathbf{z}_{it1}, \bar{\mathbf{z}}_i, v_{it2}) \qquad (21)$$

• The ASF is now obtained by averaging out $(\bar{\mathbf{z}}_i, v_{it2})$:

$$ASF(y_{t2}, \mathbf{z}_{t1}) = E_{(\bar{\mathbf{z}}_i, v_{it2})}[g_{t1}(y_{t2}, \mathbf{z}_{t1}, \bar{\mathbf{z}}_i, v_{it2})]$$

• Most of this can be fully nonparametric (Altonji and Matzkin, 2005; Blundell and Powell, 2003) although some restriction is needed on $D(\mathbf{c}_{i1}|\mathbf{z}_i, v_{it2})$, such as

$$D(\mathbf{c}_{i1}|\mathbf{z}_i, v_{it2}) = D(\mathbf{c}_{i1}|\bar{\mathbf{z}}_i, v_{it2})$$

• With T sufficiently large we can add other features of $\{\mathbf{z}_{it} : t = 1, \ldots, T\}$ to $\bar{\mathbf{z}}_i$.

5. Estimating Some Popular Models

Linear Model with Endogeneity

- Simplest model is

$$y_{it1} = \alpha_1 y_{it2} + \mathbf{z}_{it1}\boldsymbol{\delta}_1 + c_{i1} + u_{it1} \equiv \mathbf{x}_{it1}\boldsymbol{\beta}_1 + c_{i1} + u_{it1} \qquad (22)$$

$$E(u_{it1}|\mathbf{z}_i, c_{i1}) = 0$$

- The fixed effects 2SLS estimator is common. Deviate variables from time averages to remove c_{i1} then apply IV:

$$\ddot{y}_{it1} = \ddot{\mathbf{x}}_{it1}\boldsymbol{\beta}_1 + \ddot{u}_{it1}$$

$$\ddot{\mathbf{z}}_{it} = \mathbf{z}_{it} - \bar{\mathbf{z}}_i$$

• Easy to make inference robust to serial correlation and heteroskedasticity in $\{u_{it1}\}$. ("Cluster-robust inference.")

• Test for (strict) exogeneity of $\{y_{it2}\}$:

(i) Estimate the reduced form of y_{it2} by usual fixed effects:

$$y_{it2} = \mathbf{z}_{it}\boldsymbol{\delta}_1 + c_{i2} + u_{it2}$$

Get the FE residuals, $\hat{\ddot{u}}_{it2} = \ddot{y}_{it2} - \ddot{\mathbf{z}}_{it}\hat{\boldsymbol{\delta}}_1$.

• Estimate the augment equation

$$y_{it1} = \alpha_1 y_{it2} + \mathbf{z}_{it1}\boldsymbol{\delta}_1 + \rho_1 \hat{\ddot{u}}_{it2} + c_{i1} + error_{it} \quad (23)$$

by FE and use a cluster-robust test of $H_0 : \rho_1 = 0$.

• The random effects IV approach assumes c_{i1} is uncorrelated with $\mathbf{z}_i$, and nominally imposes serial independence on $\{u_{it1}\}$.

• Simple way to test the null whether REIV is sufficient. (Robust Hausman test comparing REIV and FEIV.)

Estimate

$$y_{it1} = \eta_1 + \mathbf{x}_{it1}\boldsymbol{\beta}_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + a_{i1} + u_{it1} \tag{24}$$

by REIV, using instruments $(1, \mathbf{z}_{it}, \bar{\mathbf{z}}_i)$. The estimator of $\boldsymbol{\beta}_1$ is the FEIV estimator.

• Test $H_0 : \boldsymbol{\xi}_1 = \mathbf{0}$, preferably using a fully robust test. 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

$$y_{it1} = \mathbf{x}_{it1}\boldsymbol{\beta}_1 + \mathbf{z}_{i,t+1}\boldsymbol{\psi}_1 + c_{i1} + u_{it1}, t = 1,\ldots,T-1$$

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

• Use a fully robust Wald test of $H_0 : \boldsymbol{\psi}_1 = \mathbf{0}$. Can be selective about which leads to include.

Example: Estimating a Passenger Demand Function for Air Travel

$N = 1,149$, $T = 4$.

- Uses route concentration for largest carrier as IV for log(*fare*).

```
. use airfare

. * Reduced form for lfare; concen is the IV.

. 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)
------------------------------------------------------------------------------
```

```
. 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").
```

```
. xtivreg2 lpassen ldist ldistsq y98 y99 y00 (lfare = concen), fe cluster(id)
Warning - collinearities detected
Vars dropped:  ldist ldistsq

FIXED EFFECTS ESTIMATION
------------------------
Number of groups =      1149                   Obs per group: min =          4
                                                              avg =        4.0
                                                              max =          4

Number of clusters (id) = 1149                      Number of obs =       4596
                                                    F(  4,  1148) =      26.07
                                                    Prob > F      =     0.0000
Total (centered) SS     =  128.0991685              Centered R2   =     0.2265
Total (uncentered) SS   =  128.0991685              Uncentered R2 =     0.2265
Residual SS             =   99.0837238              Root MSE      =      .1695

------------------------------------------------------------------------------
             |               Robust
     lpassen |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfare |  -.3015761   .6124127    -0.49   0.622    -1.501883    .8987307
         y98 |   .0257147   .0164094     1.57   0.117    -.0064471    .0578766
         y99 |   .0724166   .0250971     2.89   0.004     .0232272    .1216059
         y00 |   .1127914   .0620115     1.82   0.069    -.0087488    .2343316
------------------------------------------------------------------------------
Instrumented:         lfare
Included instruments: y98 y99 y00
Excluded instruments: concen
------------------------------------------------------------------------------
```

```
. 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

                                                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
------------------------------------------------------------------------------
```

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

. 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

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
------------------------------------------------------------------------------
```

```
. * 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

R-sq:  within  = 0.4507                         Obs per group: min =         4
       between = 0.0487                                        avg =       4.0
       overall = 0.0574                                        max =         4

                                                F(4,1148)          =    121.85
corr(u_i, Xb)  = -0.3249                        Prob > F           =    0.0000

                                 (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)
------------------------------------------------------------------------------
```

```
. * Test formally for endogeneity of lfare in FE:

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

 . predict u2h, 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
         u2h |  -.8616344   .5278388    -1.63   0.103    -1.897271    .1740025
       _cons |   7.501007   2.441322     3.07   0.002     2.711055    12.29096
-------------+----------------------------------------------------------------

. * p-value is about .10, so not strong evidence even though FE and
. * FEIV estimatoestimates are uite different.
```

• Turns out that the FE2SLS estimator is robust to random coefficients on $\mathbf{x}_{it1}$, but one should include a full set of time dummies. (Murtazashvili and Wooldridge, 2005).

• Can model random coefficients and use a CF approach.

$$
\begin{aligned}
y_{it1} &= c_{i1} + \mathbf{x}_{it1}\mathbf{b}_{i1} + u_{it1} \\
y_{it2} &= \eta_2 + \mathbf{z}_{it}\boldsymbol{\delta}_2 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_2 + v_{it2}
\end{aligned}
$$

• Assume $E(c_{i1}|\mathbf{z}_i, v_{it2})$ and $E(\mathbf{b}_{i1}|\mathbf{z}_i, v_{it2})$ are linear in $(\bar{\mathbf{z}}_i, v_{it2})$ and $E(u_{it1}|\mathbf{z}_i, v_{it2})$ is linear in v_{it2}, can show

$$
\begin{aligned}
E(y_{it1}|\mathbf{z}_i, v_{it2}) &= \tau_1 + \mathbf{x}_{it1}\boldsymbol{\beta}_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + \rho_1 v_{it2} \\
&\quad + [(\bar{\mathbf{z}}_i - \boldsymbol{\mu}_{\bar{\mathbf{z}}}) \otimes \mathbf{x}_{it1}]\boldsymbol{\omega}_1 + v_{it2}\mathbf{x}_{it1}\boldsymbol{\zeta}_1
\end{aligned} \tag{25}
$$

(1) Regress y_{it2} on 1, $\mathbf{z}_{it}$, $\bar{\mathbf{z}}_i$ and obtain residuals $\hat{v}_{it2}$.

(2) Regress

$$y_{it1} \text{ on } 1, \mathbf{x}_{it1}, \bar{\mathbf{z}}_i, \hat{v}_{it2}, [(\bar{\mathbf{z}}_i - \bar{\mathbf{z}}) \otimes \mathbf{x}_{it1}], \hat{v}_{it2}\mathbf{x}_{it1}$$

- Probably include time dummies in both stages.

Binary and Fractional Response

• Unobserved effects (UE) "probit" model – exogenous variables. For a binary or fractional y_{it},

$$E(y_{it}|\mathbf{x}_{it}, c_i) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta} + c_i),\ t = 1, \ldots, T. \tag{26}$$

Assume strict exogeneity (conditional on c_i) and Chamberlain-Mundlak device:

$$c_i = \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i,\ a_i|\mathbf{x}_i \sim Normal(0, \sigma_a^2). \tag{27}$$

• In binary response case under serial independence, all parameters are identified and MLE (Stata: xtprobit) can be used. Just add the time averages $\bar{\mathbf{x}}_i$ as an additional set of regressors. Then $\hat{\mu}_c = \hat{\psi} + \bar{\mathbf{x}}\hat{\boldsymbol{\xi}}$ and $\hat{\sigma}_c^2 \equiv \hat{\boldsymbol{\xi}}' \left[N^{-1} \sum_{i=1}^{N} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}) \right] \hat{\boldsymbol{\xi}} + \hat{\sigma}_a^2$. Can evaluate PEs at, say, $\hat{\mu}_c \pm k\hat{\sigma}_c$.

• Only under restrictive assumptions does c_i have an unconditional normal distribution, although it becomes more reasonable as T gets large.

• Simple to test $H_0 : \boldsymbol{\xi} = \mathbf{0}$ as null that c_i, $\bar{\mathbf{x}}_i$ are independent.

- The APEs are identified from the ASF, estimated as

$$\widehat{ASF}(\mathbf{x}_t) = N^{-1} \sum_{i=1}^{N} \Phi(\mathbf{x}_t \hat{\boldsymbol{\beta}}_a + \hat{\psi}_a + \bar{\mathbf{x}}_i \hat{\boldsymbol{\xi}}_a) \tag{28}$$

where, for example, $\hat{\boldsymbol{\beta}}_a = \hat{\boldsymbol{\beta}}/(1 + \hat{\sigma}_a^2)^{1/2}$.

- For binary or fractional response, APEs are identified without the conditional serial independence assumption. Use pooled Bernoulli quasi-MLE (Stata: glm) or generalized estimating equations (Stata: xtgee) to estimate scaled coefficients based on

$$E(y_{it}|\mathbf{x}_i) = \Phi(\mathbf{x}_{it}\boldsymbol{\beta}_a + \psi_a + \bar{\mathbf{x}}_i \boldsymbol{\xi}_a). \tag{29}$$

(Time dummies have been supressed for simplicity.)

• A more radical suggestion, but in the spirit of Altonji and Matzkin (2005), is to just use a flexible model for $E(y_{it}|\mathbf{x}_{it},\bar{\mathbf{x}}_i)$ directly, say,

$$\begin{aligned} E(y_{it}|\mathbf{x}_{it},\bar{\mathbf{x}}_i) &= \Phi[\theta_t + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\gamma} + \\ &\quad (\bar{\mathbf{x}}_i \otimes \bar{\mathbf{x}}_i)\boldsymbol{\delta} + (\mathbf{x}_{it} \otimes \bar{\mathbf{x}}_i)\boldsymbol{\eta}]. \end{aligned} \tag{30}$$

Just average out over $\bar{\mathbf{x}}_i$ to get APEs.

• If we have a binary response, start with

$$P(y_{it} = 1|\mathbf{x}_{it},c_i) = \Lambda(\mathbf{x}_{it}\boldsymbol{\beta} + c_i), \tag{31}$$

and assume CI, we can estimate $\boldsymbol{\beta}$ by FE logit without restricting $D(c_i|\mathbf{x}_i)$.

• In any nonlinear model using the Mundlak assumption $D(c_i|\mathbf{x}_i) = D(c_i|\bar{\mathbf{x}}_i)$, if $T \geq 3$ can include lead values, $\mathbf{w}_{i,t+1}$, to simply test strict exogeneity.

• Example: Married Women's Labor Force Participation: $N = 5,663$, $T = 5$ (four-month intervals).

• Following results include a full set of time period dummies (not reported).

• The APEs are directly comparable across models, and can be compared with the linear model coefficients.

LFP	(1)	(2)		(3)		(4)		(5)
Model	Linear	Probit		CRE Probit		CRE Probit		FE Logit
Est. Method	FE	Pooled MLE		Pooled MLE		MLE		MLE
	Coef.	Coef.	APE	Coef.	APE	Coef.	APE	Coef.
kids	$-.0389$	$-.199$	$-.0660$	$-.117$	$-.0389$	$-.317$	$-.0403$	$-.644$
	$(.0092)$	$(.015)$	$(.0048)$	$(.027)$	$(.0085)$	$(.062)$	$(.0104)$	$(.125)$
lhinc	$-.0089$	$-.211$	$-.0701$	$-.029$	$-.0095$	$-.078$	$-.0099$	$-.184$
	$(.0046)$	$(.024)$	$(.0079)$	$(.014)$	$(.0048)$	$(.041)$	$(.0055)$	$(.083)$
$\overline{kids}$	—	—	—	$-.086$	—	$-.210$	—	—
	—	—	—	$(.031)$	—	$(.071)$	—	—
$\overline{lhinc}$	—	—	—	$-.250$	—	$-.646$	—	—
	—	—	—	$(.035)$	—	$(.079)$	—	—

Probit with Endogenous Explanatory Variables

- Represent endogeneity as an omitted, time-varying variable, in addition to unobserved heterogeneity:

$$
\begin{aligned}
P(y_{it1} = 1 | y_{it2}, \mathbf{z}_i, c_{i1}, v_{it1}) &= P(y_{it1} = 1 | y_{it2}, \mathbf{z}_{it1}, c_{i1}, r_{it1}) \\
&= \Phi(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + c_{i1} + r_{it1})
\end{aligned}
$$

- Elements of $\mathbf{z}_{it}$ are assumed strictly exogenous, and we have at least one exclusion restriction: $\mathbf{z}_{it} = (\mathbf{z}_{it1}, \mathbf{z}_{it2})$.

• Papke and Wooldridge (2008, Journal of Econometrics): Use a Chamberlain-Mundlak approach, but only relating the heterogeneity to all strictly exogenous variables:

$$c_{i1} = \psi_1 + \bar{\mathbf{z}}_i \boldsymbol{\xi}_1 + a_{i1},\ D(a_{i1}|\mathbf{z}_i) = D(a_{i1}).$$

• Even before we specify $D(a_{i1})$, this is restrictive because it assumes, in particular, $E(c_i|\mathbf{z}_i)$ is linear in $\bar{\mathbf{z}}_i$ and that $Var(c_i|\mathbf{z}_i)$ is constant. Using nonparametrics can get by with less, such as $D(c_{i1}|\mathbf{z}_i) = D(c_{i1}|\bar{\mathbf{z}}_i)$.

- Only need

$$E(y_{it1}|y_{it2}, \mathbf{z}_i, c_{i1}, v_{it1}) = \Phi(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + c_{i1} + v_{it1}), \tag{32}$$

so applies to fractional response.

- Need to obtain an estimating equation. First, note that

$$\begin{aligned} E(y_{it1}|y_{it2}, \mathbf{z}_i, a_{i1}, r_{it1}) &= \Phi(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + \psi_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + a_{i1} + r_{it1}) \\ &\equiv \Phi(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + \psi_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + v_{it1}). \end{aligned} \tag{33}$$

- Assume a linear reduced form for y_{it2}:

$$\begin{aligned} y_{it2} &= \psi_2 + \mathbf{z}_{it}\delta_2 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_2 + v_{it2}, t = 1, \ldots, T \\ D(v_{it2}|\mathbf{z}_i) &= D(v_{it2}) \end{aligned} \tag{34}$$

(and we might allow for time-varying coefficients).

- Next, assume

$$v_{it1}|(\mathbf{z}_i, v_{it2}) \sim Normal(\eta_1 v_{it2}, \kappa_1^2), t = 1, \ldots, T.$$

[Easy to allow η_1 to change over time; just have time dummies interact with v_{it2}.]

- Assumptions effectively rule out discreteness in y_{it2}.

• Write

$$v_{it1} = \eta_1 v_{it2} + e_{it1}$$

where e_{it1} is independent of $(\mathbf{z}_i, v_{it2})$ (and, therefore, of y_{it2}) and normally distributed. Again, using a standard mixing property of the normal distribution,

$$E(y_{it1}|y_{it2}, \mathbf{z}_i, v_{it2}) = \Phi(\mathbf{x}_{it1}\boldsymbol{\beta}_{\kappa 1} + \psi_{\kappa 1} + \bar{\mathbf{z}}_i\boldsymbol{\xi}_{\kappa 1} + \eta_{\kappa 1} v_{it2}) \tag{35}$$

where the "κ" denotes division by $(1 + \kappa_1^2)^{1/2}$.

• Identification comes off of the exclusion of the time-varying exogenous variables $\mathbf{z}_{it2}$.

• Two step procedure (Papke and Wooldridge, 2008):

(1) Estimate the reduced form for y_{it2} (pooled or for each t separately). Obtain the residuals, $\hat{v}_{it2}$.

(2) Use the probit QMLE to estimate $\boldsymbol{\beta}_{\kappa 1}, \psi_{\kappa 1}, \boldsymbol{\xi}_{\kappa 1}$ and $\eta_{\kappa 1}$.

• How do we interpret the scaled estimates? They give directions of effects. Conveniently, they also index the APEs. For given y_2 and $\mathbf{z}_1$, average out $\bar{\mathbf{z}}_i$ and $\hat{v}_{it2}$ (for each t):

$$\hat{\alpha}_{\kappa 1} \cdot \left[N^{-1} \sum_{i=1}^{N} \phi(\hat{\alpha}_{\kappa 1} y_{t2} + \mathbf{z}_{t1}\hat{\boldsymbol{\delta}}_{\kappa 1} + \hat{\psi}_{\kappa 1} + \bar{\mathbf{z}}_i \hat{\boldsymbol{\xi}}_{\kappa 1} + \hat{\eta}_{\kappa 1} \hat{v}_{it2}) \right].$$

- Application: Effects of Spending on Test Pass Rates
- $N = 501$ school districts, $T = 7$ time periods.
- Once pre-policy spending is controlled for, instrument spending with the “foundation grant.”
- Initial spending takes the place of the time average of IVs.

```
. * First, linear model:

. ivreg math4 lunch alunch lenroll alenroll y96-y01 lexppp94 le94y96-le94y01
       (lavgrexp = lfound lfndy96-lfndy01), cluster(distid)

Instrumental variables (2SLS) regression               Number of obs =    3507
                                                       F( 18,   500) =  107.05
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4134
                                                       Root MSE      =  .11635

                              (Std. Err. adjusted for 501 clusters in distid)
------------------------------------------------------------------------------
             |               Robust
       math4 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |   .5545247   .2205466     2.51   0.012     .1212123     .987837
       lunch |  -.0621991   .0742948    -0.84   0.403    -.2081675    .0837693
      alunch |  -.4207815   .0758344    -5.55   0.000    -.5697749   -.2717882
     lenroll |   .0463616   .0696215     0.67   0.506    -.0904253    .1831484
    alenroll |   -.049052    .070249    -0.70   0.485    -.1870716    .0889676
         y96 |  -1.085453   .2736479    -3.97   0.000    -1.623095   -.5478119
...
         y01 |   -.704579   .7310773    -0.96   0.336    -2.140941    .7317831
    lexppp94 |  -.4343213   .2189488    -1.98   0.048    -.8644944   -.0041482
     le94y96 |   .1253255   .0318181     3.94   0.000     .0628119    .1878392
...
     le94y01 |   .0865874   .0816732     1.06   0.290    -.0738776    .2470524
       _cons |   -.334823   .2593105    -1.29   0.197    -.8442955    .1746496
------------------------------------------------------------------------------
Instrumented:  lavgrexp
Instruments:   lunch alunch lenroll alenroll y96 y97 y98 y99 y00 y01
               lexppp94 le94y96 le94y97 le94y98 le94y99 le94y00 le94y01
               lfound lfndy96 lfndy97 lfndy98 lfndy99 lfndy00 lfndy01
------------------------------------------------------------------------------
```

```
. * Get reduced form residuals for fractional probit:

. reg lavgrexp lfound lfndy96-lfndy01 lunch alunch lenroll alenroll y96-y01
       lexppp94 le94y96-le94y01, cluster(distid)

Linear regression                                      Number of obs =    3507
                                                       F( 24,   500) = 1174.57
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9327
                                                       Root MSE      =  .03987

                               (Std. Err. adjusted for 501 clusters in distid)
------------------------------------------------------------------------------
             |               Robust
    lavgrexp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lfound |   .2447063   .0417034     5.87   0.000     .1627709    .3266417
     lfndy96 |   .0053951   .0254713     0.21   0.832     -.044649    .0554391
     lfndy97 |  -.0059551   .0401705    -0.15   0.882    -.0848789    .0729687
     lfndy98 |   .0045356   .0510673     0.09   0.929    -.0957972    .1048685
     lfndy99 |   .0920788   .0493854     1.86   0.063    -.0049497    .1891074
     lfndy00 |   .1364484   .0490355     2.78   0.006     .0401074    .2327894
     lfndy01 |   .2364039   .0555885     4.25   0.000      .127188    .3456198
       ...
       _cons |   .1632959   .0996687     1.64   0.102    -.0325251     .359117
------------------------------------------------------------------------------

. predict v2hat, resid
(1503 missing values generated)
```

```
. glm math4 lavgrexp v2hat lunch alunch lenroll alenroll y96-y01 lexppp94
      le94y96-le94y01, fa(bin) link(probit) cluster(distid)
note: math4 has non-integer values

Generalized linear models                          No. of obs      =      3507
Optimization     : ML                              Residual df     =      3487
                                                   Scale parameter =         1
Deviance         =  236.0659249                    (1/df) Deviance =  .0676989
Pearson          =  223.3709371                    (1/df) Pearson  =  .0640582

Variance function: V(u) = u*(1-u/1)                [Binomial]
Link function    : g(u) = invnorm(u)               [Probit]

                              (Std. Err. adjusted for 501 clusters in distid)
------------------------------------------------------------------------------
             |               Robust
       math4 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |   1.731039   .6541194     2.65   0.008     .4489886    3.013089
       v2hat |  -1.378126    .720843    -1.91   0.056    -2.790952    .0347007
       lunch |  -.2980214   .2125498    -1.40   0.161    -.7146114    .1185686
      alunch |  -1.114775   .2188037    -5.09   0.000    -1.543623    -.685928
     lenroll |   .2856761    .197511     1.45   0.148    -.1014383    .6727905
    alenroll |  -.2909903   .1988745    -1.46   0.143    -.6807771    .0987966
         ...
       _cons |  -2.455592   .7329693    -3.35   0.001    -3.892185   -1.018998
------------------------------------------------------------------------------
```

```
. margeff

Average partial effects after glm
      y  = Pr(math4)

------------------------------------------------------------------------------
    variable |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |   .5830163   .2203345     2.65   0.008     .1511686    1.014864
       v2hat |  -.4641533    .242971    -1.91   0.056    -.9403678    .0120611
       lunch |  -.1003741   .0716361    -1.40   0.161    -.2407782      .04003
      alunch |  -.3754579   .0734083    -5.11   0.000    -.5193355   -.2315803
     lenroll |   .0962161   .0665257     1.45   0.148    -.0341719    .2266041
    alenroll |  -.0980059   .0669786    -1.46   0.143    -.2292817    .0332698
         ...
------------------------------------------------------------------------------

. * These standard errors do not account for the first-stage estimation.
. * Can use the panel bootstrap. Might also look for partial effects at
. * different parts of the spending distribution.
```

Count and Other Multiplicative Models

- Conditional mean with multiplicative heterogeneity:

$$E(y_{it}|\mathbf{x}_{it}, c_i) = c_i \exp(\mathbf{x}_{it}\boldsymbol{\beta}) \tag{36}$$

where $c_i \geq 0$. Under strict exogeneity in the mean,

$$E(y_{it}|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, c_i) = E(y_{it}|\mathbf{x}_{it}, c_i), \tag{37}$$

the "fixed effects" Poisson estimator is attractive: it does not restrict $D(y_{it}|\mathbf{x}_i, c_i)$, $D(c_i|\mathbf{x}_i)$, or serial dependence.

- The FE Poisson estimator is the conditional MLE derived under a Poisson and conditional independence assumptions. It is one of the rare cares where treating the c_i as parameters to estimate gives a consistent estimator of $\boldsymbol{\beta}$.
- The FE Poisson estimator is fully robust to any distributional failure and serial correlation. y_{it} does not even have to be is not a count variable! Fully robust inference is easy (`xtpqml` in Stata).

• For endogeneity there are control function and GMM approaches, with the former being more convenient but imposing more restrictions.

• CF uses same approach as before.

• Start with an omitted variables formulation:

$$E(y_{it1}|y_{it2}, \mathbf{z}_i, c_{i1}, r_{it1}) = \exp(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + c_{i1} + r_{it1}). \tag{38}$$

• The $\{\mathbf{z}_{it}\}$ – including the excluded instruments – are assumed to be strictly exogenous here.

- If y_{it2} is (roughly) continuous we might specify

$$y_{it2} = \psi_2 + \mathbf{z}_{it}\boldsymbol{\pi}_2 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_2 + v_{it2}.$$

- Also write

$$c_{i1} = \psi_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + a_{i1}$$

so that

$$E(y_{it1}|y_{it2},\mathbf{z}_i,v_{it1}) = \exp(\psi_1 + \mathbf{x}_{it1}\boldsymbol{\beta}_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + v_{it1}),$$

where $v_{it1} = a_{i1} + r_{it1}$.

• Reasonable (but not completely general) to assume (v_{it1}, v_{i2}) is independent of $\mathbf{z}_i$.

• If we specify $E[\exp(v_{it1})|v_{it2}] = \exp(\eta_1 + \rho_1 v_{it2})$ (as would be true under joint normality), we obtain the estimating equation

$$E(y_{it1}|y_{it2}, \mathbf{z}_i, v_{it2}) = \exp(\kappa_1 + \mathbf{x}_{it1}\boldsymbol{\beta}_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + \rho_1 v_{it2}). \quad (39)$$

- Now apply a simple two-step method. (1) Obtain the residuals $\hat{v}_{it2}$ from the pooled OLS estimation y_{it2} on 1, $\mathbf{z}_{it}$, $\bar{\mathbf{z}}_i$ across t and i. (2) Use a pooled QMLE (perhaps the Poisson or NegBin II) to estimate the exponential function, where $(\bar{\mathbf{z}}_i, \hat{v}_{it2})$ are explanatory variables along with $(\mathbf{x}_{it1})$. (As usual, a fully set of time period dummies is a good idea in the first and second steps).
- Note that y_{it2} is not strictly exogenous in the estimating equation. and so GLS-type methods account for serial correlation should not be used. GMM with carefully constructed moments could be.

• Estimating the ASF is straightforward:

$$\widehat{ASF}_t(y_{t2}, \mathbf{z}_{t1}) = N^{-1} \sum_{i=1}^{N} \exp(\hat{\kappa}_1 + \mathbf{x}_{t1}\hat{\boldsymbol{\beta}}_1 + \bar{\mathbf{z}}_i\hat{\boldsymbol{\xi}}_1 + \hat{\rho}_1\hat{v}_{it2});$$

that is, we average out $(\bar{\mathbf{z}}_i, \hat{v}_{it2})$.

• Test the null of contemporaneous exogeneity of y_{it2} by using a fully robust t statistic on $\hat{v}_{it2}$.

• Can allow more flexibility by iteracting $(\bar{\mathbf{z}}_i, \hat{v}_{it2})$ with $\mathbf{x}_{it1}$, or even just year dummies.

- A GMM approach – which slightly extends Windmeijer (2002) – modifies the moment conditions under a sequential exogeneity assumption on instruments and applies to models with lagged dependent variables.
- Write the model as

$$y_{it} = c_i \exp(\mathbf{x}_{it}\boldsymbol{\beta}) r_{it} \tag{40}$$

$$E(r_{it} | \mathbf{z}_{it}, \ldots, \mathbf{z}_{i1}, c_i) = 1, \tag{41}$$

which contains the case of sequentially exogenous regressors as a special case ($\mathbf{z}_{it} = \mathbf{x}_{it}$).

- Now start with the transformation

$$\frac{y_{it}}{\exp(\mathbf{x}_{it}\boldsymbol{\beta})} - \frac{y_{i,t+1}}{\exp(\mathbf{x}_{i,t+1}\boldsymbol{\beta})} = c_i(r_{it} - r_{i,t+1}). \tag{42}$$

- Can easily show that

$$E[c_i(r_{it} - r_{i,t+1})|\mathbf{z}_{it}, \ldots, \mathbf{z}_{i1}] = 0, t = 1, \ldots, T - 1.$$

• Using the moment conditions

$$E\left[\frac{y_{it}}{\exp(\mathbf{x}_{it}\boldsymbol{\beta})} - \frac{y_{i,t+1}}{\exp(\mathbf{x}_{i,t+1}\boldsymbol{\beta})}\middle|\mathbf{z}_{it},\ldots,\mathbf{z}_{i1}\right] = 0,\ t = 1,\ldots,T-1 \qquad (43)$$

generally causes computational problems. For example, if $x_{itj} \geq 0$ for some j and all i and t – for example, if x_{itj} is a time dummy – then the moment conditions can be made arbitarily close to zero by choosing β_j larger and larger.

• Windmeijer (2002, Economics Letters) suggested multiplying through by $\exp(\boldsymbol{\mu}_{\mathbf{x}}\boldsymbol{\beta})$ where $\boldsymbol{\mu}_{\mathbf{x}} = T^{-1}\sum_{r=1}^{T} E(\mathbf{x}_{ir})$.

• So, the modified moment conditions are

$$E\left[\frac{y_{it}}{\exp[(\mathbf{x}_{it}-\boldsymbol{\mu}_{\mathbf{x}})\boldsymbol{\beta}]}-\frac{y_{i,t+1}}{\exp[(\mathbf{x}_{i,t+1}-\boldsymbol{\mu}_{\mathbf{x}})\boldsymbol{\beta}]}\Big|\mathbf{z}_{it},\ldots,\mathbf{z}_{i1}\right]=0. \tag{44}$$

• As a practical matter, replace $\boldsymbol{\mu}_{\mathbf{x}}$ with the overall sample average,

$$\bar{\mathbf{x}}=(NT)^{-1}\sum_{i=1}^{N}\sum_{r=1}^{T}\mathbf{x}_{ir}. \tag{45}$$

• The deviated variables, $\mathbf{x}_{it}-\bar{\mathbf{x}}$, will always take on positive and negative values, and this seems to solve the GMM computational problem.