MODELS WITH ENDOGENOUS EXPLANATORY VARIABLES AND HETEROGENEITY

Correlated Random Effects Panel Data Models

IZA Summer School in Labor Economics
May 13-19, 2013
Jeffrey M. Wooldridge
Michigan State University

1. FEIV as a CRE Estimator
2. The Hausman Test Comparing REIV and FEIV
3. Nonlinear Models with Heterogeneity and Endogeneity
4. Probit Response Function with an EEV

1. FEIV as a CRE Estimator

- 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 $\mathbf{x}_{it}$ $(1 \times K)$ are correlated with u_{it} (or u_{ir} for $r \neq t$).

- Let $\mathbf{z}_{it}$ be a set of $1 \times L$ instrumental variables, $L \geq K$.
- Strictly exogenous elements of $\mathbf{x}_{it}$ are in $\mathbf{z}_{it}$.

- The fixed effects IV estimator (FE2SLS) relies on strict exogeneity of the IVs:

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

- For the rank condition, we must have enough time-varying IVs.
- We can obtain the FEIV estimator using a CRE approach. Nominally we assume a Mundlak form:

$$c_i = \psi + \bar{\mathbf{z}}_i\boldsymbol{\xi} + a_i$$
$$E(a_i|\mathbf{z}_i) = 0$$

- We obtain the estimating equation

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \psi + \bar{\mathbf{z}}_i\boldsymbol{\xi} + a_i + u_{it}$$
$$E(a_i + u_{it}|\mathbf{z}_i) = 0, t = 1, \ldots, T$$

- Some elements of $\mathbf{x}_{it}$ are still endogenous. Estimate the equation by by pooled 2SLS or RE2SLS, using instruments $(1, \mathbf{z}_{it}, \bar{\mathbf{z}}_i)$.
- The estimate of $\boldsymbol{\beta}$ is the FE2SLS estmator.
- We can add time-constant variables.

- As usual, make inference robust to any kind of serial correlation and heteroskedasticity in $\{a_i + u_{it}\}$, even if REIV is used.
- Currently limitation of Stata: `xtivreg` does not allow a "cluster robust" option. Can bootstrap. `xtivreg2` does not support REIV estimation.

Testing Strict Exogeneity of the Instruments

- Use an augmented model that includes lead values of the IVs:

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

- Estimate by FEIV, using instruments $(\mathbf{z}_{it}, \mathbf{z}_{i,t+1})$, and test $H_0 : \boldsymbol{\gamma} = \mathbf{0}$.

Estimating Passenger Demand: AIRFARE.DTA

```
. * 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
------------------------------------------------------------------------------

. 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

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

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

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
Dropped collinear:    ldist ldistsq
------------------------------------------------------------------------------
```

```
. 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
------------------------------------------------------------------------------
```

Testing Exogeneity of the Explanatory Variables

• Write

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

• We want to test whether $\mathbf{y}_{it2}$ is (strictly) exogenous, maining strict exogeneity of $\{\mathbf{z}_{it}\}$ where $\mathbf{z}_{it1} \subset \mathbf{z}_{it}$.

• Regression-based test: Estimate the reduced forms,

$$\mathbf{y}_{it2} = \mathbf{z}_{it}\boldsymbol{\Pi}_2 + \mathbf{v}_{it2}$$

by FE and obtain the FE residuals, $\hat{\ddot{\mathbf{v}}}_{it2}$.

• Estimate

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

by fixed effects and test $H_0 : \boldsymbol{\rho}_1 = \mathbf{0}$.

• Estimates of $\boldsymbol{\delta}_1$ and $\boldsymbol{\alpha}_1$ are actually the FEIV estimates. The is the "control function" version of fixed effects.

• The first-stage estimation does not affect the asymptotic distribution under the null.

- The above test is a robust, variable edition version of the Hausman test comparing the FE estimator and the FE2SLS estimator.
- Apply to the AIRFARE data.

```
. * What are the usual FE estimates of the demand function?

. 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 for endogeneity of lfare:

. 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
-------------+----------------------------------------------------------------

. * p-value is .103, so not strong evidence even though estimates are
. * quite different (-1.155 versus -.302).
```

2. The Hausman Test Comparing REIV and FEIV

- The traditional Hausman test is nonrobust, and can lead to computational and degrees-of-freedom problems.
- Can use a variable addition form of the Hausman test.

- As before, estimate

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

by pooled 2SLS or REIV, using instruments $(1, \mathbf{z}_{it}, \bar{\mathbf{z}}_i)$ and test Test $H_0 : \boldsymbol{\xi} = \mathbf{0}$, preferably using a fully robust test.

- A rejection is evidence that the IVs are correlated with c_i, and should use FEIV.

- Include time-constant variables in equation as extra controls.

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

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

. estimates store b_feiv

. qui xtivreg lpassen ldist ldistsq 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    -.5078761           .2063        .1555309
         y98 |    .0257147     .0307363       -.0050216         .004651
         y99 |    .0724166     .0796548       -.0072383        .0060892
         y00 |    .1127914     .1325795       -.0197881        .0151611
------------------------------------------------------------------------------
                         b = consistent under Ho and Ha; obtained from xtivreg
          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)
                          =        1.76
                Prob>chi2 =      0.7799
```

```
. 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    -.5078761           .2063        .2064464
         y98 |    .0257147     .0307363       -.0050216        .0066745
         y99 |    .0724166     .0796548       -.0072383        .0084713
         y00 |    .1127914     .1325795       -.0197881        .0202836
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from xtivreg
           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)
                          =        1.00
                Prob>chi2 =      0.9100
```

```
. * Use pooled IV with CRE to obtain robust test:

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

. 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
------------------------------------------------------------------------------

. * Robust p-value, with one df, is .046.
```

3. Nonlinear 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.

- Model endogeneity with heterogeneity as two sources of omitted variables: heterogeneity, $\mathbf{c}_{i1}$ and time-varying omitted variables, $\mathbf{r}_{it1}$.

- “Structural” expectation is

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

- $\mathbf{y}_{it2}$ is allowed to be correlated with $\mathbf{r}_{it1}$ as well as with $\mathbf{c}_{i1}$.
- The average structural function is now

$$ASF(\mathbf{y}_{t2}, \mathbf{z}_{t1}) = E_{(\mathbf{c}_{i1}, \mathbf{r}_{it1})}[m_{t1}(\mathbf{y}_{t2}, \mathbf{z}_{t1}, \mathbf{c}_{i1}, \mathbf{r}_{it1})]$$

so we average out both sources of unobservables.

• For consistent estimation of the ASF we need outside instruments, so $\mathbf{z}_{it1} \subset \mathbf{z}_{it}$.

• Most of the time should allow $\mathbf{c}_{i1}$ to be correlated with $\{\mathbf{z}_{it} : t = 1, \ldots, T\}$.

• Can have complicated time effects (such as year dummies and interactions).

- Cases where we can eliminate $\mathbf{c}_{i1}$ to obtain an estimating equation are rare:

(i) Models with a single additive effect.

(ii) Models with a single multiplicative effect.

- To handle other cases, and to obtain average partial effects, we use the CRE approach and model $D(\mathbf{c}_{i1}|\mathbf{z}_i)$ (or impose substantive restrictions).

- In nonlinear models we also need some restriction on how $\mathbf{y}_{it2}$ is related to $\mathbf{r}_{it1}$.

Exogeneity of Instruments

• The sequence $\{\mathbf{z}_{it} : t = 1,\ldots,T\}$ is strictly exogenous conditional on $\mathbf{c}_{i1}$ in the sense that

$$E(y_{it1}|\mathbf{y}_{it2},\mathbf{z}_i,\mathbf{c}_{i1},\mathbf{r}_{it1}) = E(y_{it1}|\mathbf{y}_{it2},\mathbf{z}_{it1},\mathbf{c}_{i1},\mathbf{r}_{it1})$$
$$D(\mathbf{r}_{it1}|\mathbf{z}_i,\mathbf{c}_{i1}) = D(\mathbf{r}_{it1})$$

• First incorporates standard exclusion restrictions.

• Second imposes strict exogeneity of $\{\mathbf{z}_{it}\}$ with respect to $\{\mathbf{r}_{it}\}$. Rules out feedback, so the IVs cannot react in the future given past changes to the "shocks" $\{\mathbf{r}_{it1}\}$.

• Neither restricts $D(\mathbf{c}_{i1}|\mathbf{z}_i)$.

Reduced Forms for $\mathbf{y}_{it2}$

- CRE methods combined with control function methods are convenient for allowing correlation of $\mathbf{y}_{it2}$ with $(\mathbf{c}_{i1}, \mathbf{r}_{it1})$. So two sources of endogeneity.

- Suppose y_{it2} is a scalar and its reduced form satisfies

$$\begin{aligned} y_{it2} &= m_{t2}(\mathbf{z}_{it}, \bar{\mathbf{z}}_i, \boldsymbol{\theta}_2) + v_{it2} \\ D(v_{it2}|\mathbf{z}_i) &= D(v_{it2}) \end{aligned}$$

where $\bar{\mathbf{z}}_i = T^{-1}\sum_{t=1}^{T} \mathbf{z}_{it}$ is the time average.

- Think of v_{it2} containing heterogeneity a_{i2} independent of $\{\mathbf{z}_{it}\}$.

• Often we assume a linear model for $m_{t2}(\cdot)$. The equation

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

can be gotten from

$$y_{it2} = \mathbf{z}_{it}\boldsymbol{\delta}_2 + c_{i2} + u_{it2}$$
$$c_{i2} = \psi_2 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_2 + a_{i2}$$

- In a fully nonparametric context, can add to $\bar{\mathbf{z}}_i$ other functions of $\{\mathbf{z}_{it} : t = 1, \ldots, T\}$ – such as variances and covariances – and even non-exchangeable statistics such as initial conditions, period-specific means, and individual specific trends. The larger is T the more we can add.
- However, independence between v_{it2} and $\mathbf{z}_i$ is substantive and (generally) difficult to relax.
- Criticism: In fully structural models for (y_{it1}, y_{it2}), how can we be sure y_{it2} has a reduced form with an additive, independent error?

• If we assume

$$y_{it2} = \psi_2 + \mathbf{z}_{it}\boldsymbol{\delta}_2 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_2 + v_{it2}$$
$$D(v_{it2}|\mathbf{z}_i) = D(v_{it2})$$

we can generally identify the ASF if we assume

$$D(\mathbf{c}_{i1}, \mathbf{r}_{it1}|y_{it2}, \mathbf{z}_i, v_{it2}) = D(\mathbf{c}_{i1}, \mathbf{r}_{it1}|v_{it2}, \bar{\mathbf{z}}_i)$$

• In practice, make this assumption and assume a parametric model for

$$D(\mathbf{c}_{i1}, r_{it1}|\mathbf{z}_i, v_{it2}).$$

• With sufficient variation in $\mathbf{z}_{it2}$ across i and t, we can identify

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

• An important insight is that the ASF can be obtained by averaging out $(\bar{\mathbf{z}}_i,v_{it2})$:

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

• The original structural function $m_{t1}(\bullet)$ has disappeared! We only need to be able to estimate

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

• Together, $(\bar{\mathbf{z}}_i,v_{it2})$ act as suitable proxies for $(\mathbf{c}_{i1},\mathbf{r}_{it1})$.

- When y_{it2} is discrete (such as binary) a different approach is needed. We cannot write $y_{it2} = m_{t2}(\mathbf{z}_{it}, \bar{\mathbf{z}}_i, \boldsymbol{\delta}_2) + v_{it2}$ where v_{it2} is independent of $\mathbf{z}_i$.

- But we can often use pooled MLEs along with the CRE approach by specifying

$$
\begin{aligned}
y_{it2} &= 1[\psi_2 + \mathbf{z}_{it}\boldsymbol{\delta}_2 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_2 + e_{it2} > 0] \\
D(e_{it}|\mathbf{z}_i) &= Normal(0,1)
\end{aligned}
$$

4. Probit Response Function with an EEV

• Represent endogeneity as an omitted, time-varying variable, in addition to unobserved heterogeneity. For $0 \leq y_{it1} \leq 1$,

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

where $\mathbf{x}_{it1}$ can be any function of $(y_{it2}, \mathbf{z}_{it1})$.

• y_{it1} could be a fractional reponse. If binary then

$$E(y_{it1}|y_{it2}, \mathbf{z}_i, c_{i1}, r_{it1}) = P(y_{it1} = 1|y_{it2}, \mathbf{z}_i, c_{i1}, r_{it1}).$$

• Elements of $\mathbf{z}_{it}$ are 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 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 $E(c_i|\mathbf{z}_i)$ is linear in $\bar{\mathbf{z}}_i$ and $Var(c_i|\mathbf{z}_i)$ is constant.

• Using nonparametrics we can get by with less, such as $D(c_{i1}|\mathbf{z}_i) = D(c_{i1}|\bar{\mathbf{z}}_i)$. Or add other functions of $\{\mathbf{z}_{it} : t = 1,\ldots,T\}$ such as individual-specific trends.

- To obtain an estimating equation we allow different time intercepts:

$$
\begin{aligned}
E(y_{it1}|y_{it2}, \mathbf{z}_i, a_{i1}, r_{it1}) &= \Phi(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + \psi_{t1} + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + a_{i1} + r_{it1}) \\
&\equiv \Phi(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + \psi_{t1} + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + v_{it1}).
\end{aligned}
$$

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

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

• So far we have

$$E(y_{it1}|y_{it2},\mathbf{z}_i,v_{it1}) = \Phi(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + \psi_{t1} + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + v_{it1})$$
$$y_{it2} = \psi_2 + \mathbf{z}_{it}\boldsymbol{\delta}_2 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_2 + v_{it2}$$

• With (v_{it1}, v_{it2}) independent of $\mathbf{z}_i$ we might impose a parametric assumption:

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

Easy to allow η_1 to change over time: interact v_{it2} with time dummies. With just a little more work, κ_1^2 could change over time.

• 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.

- The estimating function is

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

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

- Identification comes off of the exclusion of the time-varying exogenous variables $\mathbf{z}_{it2}$. Could replace $\bar{\mathbf{z}}_i$ with $\mathbf{z}_i$.

• We can only estimate scaled coefficients? It turns out these are exactly the coefficients that appear in an estimable formula for the APEs.

• If we start with

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

we can justify replacing $\Phi(\cdot)$ with any other continuous cdf, such as the logistic.

• Two step procedure:

1. Estimate the reduced form for y_{it2},

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

by OLS (pooled or for each t separately). Obtain the residuals, $\hat{v}_{it2}$.

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

• The scaled estimates index the APEs. For given y_{t2} and $\mathbf{z}_{t1}$, average out $\bar{\mathbf{z}}_i$ and $\hat{v}_{it2}$. The APE for y_{t2} with coefficient $\hat{\alpha}_{\kappa 1}$:

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

• Can use panel bootstrap for valid inference.

- No restriction on the dynamics of any time-varying variables, including r_{it1}.
- As shown in Wooldridge (2013, forthcoming, Journal of Econometrics), a one-step alternative is to apply the pooled "IV probit" QMLE with covariates $(\mathbf{x}_{it1}, \bar{\mathbf{z}}_i)$ in the first equation and $(\mathbf{z}_{it}, \bar{\mathbf{z}}_i)$ in the second equations. Then we estimate the unscaled coefficients. This works even for fractional y_{it1}.

• Simple endogeneity test: Clustered t statistic for $\hat{\eta}_{\kappa 1}$ in pooled probit. Or interact time dummies with $\hat{v}_{it2}$ for a T degrees-of-freedom test.

• The CRE/CF approach is much preferred to "plug-in" approaches. If $\mathbf{x}_{it1}$ includes y_{it2}^2 or $y_{it2}\mathbf{z}_{it1}$ and we replace y_{it2} with $\hat{y}_{it2}$ we get a mess:

$$y_{it2}\mathbf{z}_{it1} = (y_{it2}^* + v_{it2})\mathbf{z}_{it1} = y_{it2}^*\mathbf{z}_{it1} + v_{it2}\mathbf{z}_{it1}$$

and the term $v_{it2}\mathbf{z}_{it1}$ introduces heteroskedasticity inside a nonlinear function.

• Literature on identifying APEs is liberating. Could use a response function

$$\Phi(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + \psi_1 + \eta_1\hat{v}_{it2} + \tau_1\hat{v}_{it2}^2 + \hat{v}_{it2}\mathbf{x}_{it1}\boldsymbol{\omega}_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + \hat{v}_{it2}\bar{\mathbf{z}}_i\boldsymbol{\rho}_1). \tag{34}$$

The ASF is estimated by averaging out $(\bar{\mathbf{z}}_i, \hat{v}_{it2})$:

$$\widehat{ASF}_t(y_{t2}, \mathbf{z}_{t1}) = N^{-1}\sum_{t=1}^{T}\Phi(\mathbf{x}_{t1}\hat{\boldsymbol{\beta}}_1 + \hat{\psi}_1 + \eta_1\hat{v}_{it2} + \tau_1\hat{v}_{it2}^2 + \hat{v}_{it2}\mathbf{x}_{t1}\hat{\boldsymbol{\omega}}_1 + \bar{\mathbf{z}}_i\hat{\boldsymbol{\xi}}_1 + \hat{v}_{it2}\bar{\mathbf{z}}_i\hat{\boldsymbol{\rho}}_1) \tag{35}$$

- Could even use a "hetprobit" model in the second stage where $(\bar{\mathbf{z}}_i, \hat{v}_{it2})$ along with time dummies can appear in the variance.
- Because we impose parametric assumptions, $\bar{\mathbf{z}}_i$ can be replaced with $\mathbf{z}_i$ everywhere (in first and second stages).

Empirical Example: Papke and Wooldridge (2008)

y_{it1} = fraction of students passing 4th grade math test

y_{it2} = log(average per student spending)

N = 501 districts

T = 7 years

Instrument for spending: “foundation grant”

Model:	**Linear**	**Fractional Probit**	
Estimation Method:	**Instrumental Variables**	**Pooled QMLE**	
	Coefficient	Coefficient	APE
log(*arexppp*)	.555	1.731	.583
	(.221)	(.759)	(.255)
lunch	–.062	–.298	–.100
	(.074)	(.202)	(.068)
log(*enroll*)	.046	.286	.096
	(.070)	(.209)	(.070)
$\hat{v}_2$	–.424	–1.378	—
	(.232)	(.811)	—

Estimated Effects by Spending Percentiles

	2001	
	Spending Exogenous	Spending Endogenous
Percentile	APE (Standard Error)	APE (Standard Error)
5th	.292 (.072)	.599 (.273)
25th	.283 (.069)	.569 (.254)
50th	.276 (.065)	.543 (.234)
75th	.264 (.060)	.498 (.198)
95th	.229 (.043)	.365 (.101)

```
. use meap92_01

. reg lavgrexp lfound lfndy96-lfndy01 lunch alunch lenroll alenroll y96-y01
           lexppp94 le94y96-le94y01 if year > 1994, robust cluster(distid)

Regression with robust standard errors                 Number of obs =    3507
                                                       F( 24,   500) = 1174.57
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9327
Number of clusters (distid) = 501                      Root MSE      =  .03987

------------------------------------------------------------------------------
             |               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
       lunch |   .0707136   .0219002     3.23   0.001     .0276858    .1137414
      alunch |   .0950213   .0246069     3.86   0.000     .0466757    .1433669
     lenroll |  -.2160362   .0175133   -12.34   0.000    -.2504449   -.1816274
    alenroll |   .2224296    .017509    12.70   0.000     .1880293    .2568299
         y96 |   .3696531   .0656164     5.63   0.000     .2407352     .498571
         y97 |   .8609391   .1331651     6.47   0.000     .5993071    1.122571
         y98 |   1.324608   .1985554     6.67   0.000     .9345027    1.714714
         y99 |    1.43907   .1980365     7.27   0.000     1.049984    1.828156
         y00 |   1.369825   .2236949     6.12   0.000     .9303276    1.809323
         y01 |   1.125462   .2686096     4.19   0.000     .5977199    1.653205
    lexppp94 |   .7426653   .0386746    19.20   0.000     .6666806      .81865
     le94y96 |  -.0439181   .0229688    -1.91   0.056    -.0890454    .0012092
     le94y97 |  -.0853345   .0330271    -2.58   0.010    -.1502236   -.0204455
     le94y98 |  -.1475106   .0385361    -3.83   0.000    -.2232233    -.071798
     le94y99 |  -.2488234   .0375783    -6.62   0.000    -.3226542   -.1749927
     le94y00 |  -.2860549   .0374151    -7.65   0.000    -.3595652   -.2125447
```

```
     le94y01 |  -.3604275   .0402212    -8.96   0.000    -.4394508   -.2814042
       _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 if year > 1994, fa(bin) link(probit) cluster(distid)
note: math4 has non-integer values

Generalized linear models                          No. of obs      =      3507
Optimization     : ML: Newton-Raphson              Residual df     =      3487
                                                   Scale parameter =         1
Deviance         =  2640.161906                    (1/df) Deviance =  .7571442
Pearson          =  223.3709371                    (1/df) Pearson  =  .0640582

Variance function: V(u) = u*(1-u)                  [Bernoulli]
Link function    : g(u) = invnorm(u)               [Probit]
Standard errors  : Modified Sandwich

Log pseudo-likelihood = -1433.305294               AIC             =  .8288026
BIC                   =-25822.53226

                           (standard errors adjusted for clustering on 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
         y96 |  -3.208359   .7862358    -4.08   0.000    -4.749353   -1.667366
         y97 |  -3.072289   1.085228    -2.83   0.005    -5.199297   -.9452811
         y98 |  -2.245485   1.469355    -1.53   0.126    -5.125369    .6343984
```

y99	-2.114372	1.751352	-1.21	0.227	-5.546958	1.318215
y00	-1.915712	1.98039	-0.97	0.333	-5.797205	1.965782
y01	-2.9817	2.198171	-1.36	0.175	-7.290035	1.326636
lexppp94	-1.386843	.6393514	-2.17	0.030	-2.639949	-.1337374
le94y96	.3696914	.0913095	4.05	0.000	.1907281	.5486547
le94y97	.3350894	.122254	2.74	0.006	.095476	.5747028
le94y98	.2806904	.1642631	1.71	0.087	-.0412594	.6026401
le94y99	.259349	.1968296	1.32	0.188	-.12643	.6451279
le94y00	.2445331	.2216178	1.10	0.270	-.1898297	.6788959
le94y01	.3593392	.2458225	1.46	0.144	-.122464	.8411423
_cons	-2.455592	.7329693	-3.35	0.001	-3.892185	-1.018998

```
. margins, dydx(lavgrexp)

Average marginal effects                          Number of obs   =       3507
Model VCE    : Robust

Expression   : Predicted mean math4, predict()
dy/dx w.r.t. : lavgrexp

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |   .5830163   .2203345     2.65   0.008     .1511686    1.014864
------------------------------------------------------------------------------
```

```
. capture program drop math4_boot

.
. program math4_boot, rclass
  1.
. * First Stage
. reg lavgrexp lfound lfndy96-lfndy01 lunch alunch lenroll alenroll y96-y01
         lexppp94 le94y96-le94y01 if year>1994, robust cluster(distid)
  2. predict v2hat, resid
  3. * Second Stage
. glm math4 lavgrexp v2hat lunch alunch lenroll alenroll y96-y01 lexppp94
      le94y96-le94y01 if year>1994, fa(bin) link(probit) cluster(distid)
  4.
. return scalar blavgrexp = _b[lavgrexp]
  5. return scalar blunch = _b[lunch]
  6. return scalar blenroll = _b[lenroll]
  7. return scalar bv2hat = _b[v2hat]
  8.
. predict x1b1hat, xb
  9. gen scale=normalden(x1b1hat)
 10. gen pe1=scale*_b[lavgrexp]
 11. summarize pe1
 12. return scalar ape1=r(mean)
 13. gen pe2=scale*_b[lunch]
 14. summarize pe2
 15. return scalar ape2=r(mean)
 16. gen pe3=scale*_b[lenroll]
 17. summarize pe3
 18. return scalar ape3=r(mean)
 19.
. drop v2hat x1b1hat scale pe1 pe2 pe3
 20. end

.
.
. *Bootstrapped SE within districts
. bootstrap r(blavgrexp) r(blunch) r(blenroll) r(bv2hat) r(ape1) r(ape2) r(ape3),
          reps(500) seed(123) cluster(distid) idcluster(newid): math4_boot
```

```
(running math4_boot on estimation sample)

Bootstrap replications (500)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

..................................................   500

Bootstrap results                               Number of obs      =      3507
                                                Replications       =       500

      command:  math4_boot
        _bs_1:  r(blavgrexp)
        _bs_2:  r(blunch)
        _bs_3:  r(blenroll)
        _bs_4:  r(bv2hat)
        _bs_5:  r(ape1)
        _bs_6:  r(ape2)
        _bs_7:  r(ape3)

                                (Replications based on 501 clusters in distid)
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _bs_1 |   1.731039   .7589086     2.28   0.023     .2436054    3.218473
       _bs_2 |  -.2980214   .2023112    -1.47   0.141    -.6945441    .0985013
       _bs_3 |   .2856761   .2090573     1.37   0.172    -.1240686    .6954208
       _bs_4 |  -1.378126   .8113834    -1.70   0.089    -2.968408    .2121567
       _bs_5 |   .5830163   .2552307     2.28   0.022     .0827733    1.083259
       _bs_6 |   -.100374   .0681039    -1.47   0.141    -.2338551    .0331071
       _bs_7 |   .0962161   .0703321     1.37   0.171    -.0416323    .2340644
------------------------------------------------------------------------------
```

• Semykina and Wooldridge (2012, unpublished) cover the case when y_{it2} is binary. Now write

$$y_{it2} = 1[\psi_2 + \mathbf{z}_{it}\delta_2 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_2 + u_{it2} > 0]$$
$$u_{it2}|\mathbf{z}_i \sim Normal(0,1)$$

• If y_{it1} is also binary:

$$y_{it1} = 1[\mathbf{x}_{it1}\boldsymbol{\beta}_{\kappa 1} + \psi_{\kappa 1} + \bar{\mathbf{z}}_i\boldsymbol{\xi}_{\kappa 1} + v_{it1} > 0]$$

where $\mathbf{x}_{it1}$ can include y_{it2} and interactions $y_{it2}\mathbf{z}_{it1}$.

- If (v_{it1}, v_{it2}) is jointly normal, have the "bivariate probit" model for each t. Can use standard software by pooling and including time effects. Cluster standard errors.
- Turns out the same method words if we only assume

$$E(y_{it1}|y_{it2}, \mathbf{z}_i, a_{i1}, r_{it1}) = \Phi(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + \psi_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + v_{it1});$$

the usual `biprobit` approach can be used in a pooled QMLE framework when y_{it1} is fractional and y_{it2} is binary.

• A variable addition test (VAT) version of the score test is available in two steps.

1. Estimate a pooled probit of

$$y_{it2} \text{ on } 1, d2_t, \ldots, dT_t, \mathbf{z}_{it}, \bar{\mathbf{z}}_i$$

(or estimate for each t separately). Obtain the "generalized residuals"

$$\widehat{gr}_{it2} \equiv y_{it2} \frac{\phi(\mathbf{w}_{it}\hat{\boldsymbol{\theta}}_2)}{\Phi(\mathbf{w}_{it}\hat{\boldsymbol{\theta}}_2)} - (1 - y_{it2}) \frac{\phi(\mathbf{w}_{it}\hat{\boldsymbol{\theta}}_2)}{[1 - \Phi(\mathbf{w}_{it}\hat{\boldsymbol{\theta}}_2)]}$$

where $\mathbf{w}_{it} = (1, d2_t, \ldots, dT_t, \mathbf{z}_{it}, \bar{\mathbf{z}}_i)$.

2. Estimate a pooled probit of

$$y_{it1} \text{ on } 1, d2_t, \ldots, dT_t, \mathbf{x}_{it1}, \bar{\mathbf{z}}_i, \widehat{gr}_{it2}$$

and use a robust t test on $\widehat{gr}_{it2}$.

• Get a test (two df) for the full switching model by adding $y_{it2}\widehat{gr}_{it2}$.

• Exactly the same procedure works with y_{it1} fractional, including the pooled "bivariate probit" estimation.

• Radical idea (Wooldridge, 2013, forthcoming): Might controlling for the generalized residual in a flexible way approximately solve the endogeneity problem? In other words, estimate a probit response function of the form

$$\Phi(\mathbf{x}_{it1}\boldsymbol{\beta}_1 + \psi_1 + \eta_1 \widehat{gr}_{it2} + \tau_1 \widehat{gr}_{it2}^2 + \widehat{gr}_{it2}\mathbf{x}_{it1}\boldsymbol{\omega}_1 + \bar{\mathbf{z}}_i\boldsymbol{\xi}_1 + \widehat{gr}_{it2}\bar{\mathbf{z}}_i\boldsymbol{\rho}_1),$$

and, for fixed $\mathbf{x}_{t1}$, average out $(\widehat{gr}_{it2}, \bar{\mathbf{z}}_i)$ to estimate the ASF:

$$\widehat{ASF}_t(y_{t2}, \mathbf{z}_{t1}) = N^{-1}\sum_{t=1}^{T} \Phi(\mathbf{x}_{t1}\hat{\boldsymbol{\beta}}_1 + \hat{\psi}_1 + \hat{\eta}_1 \widehat{gr}_{it2} + \hat{\tau}_1 \widehat{gr}_{it2}^2 + \widehat{gr}_{it2}\mathbf{x}_{it1}\hat{\boldsymbol{\omega}}_1 + \bar{\mathbf{z}}_i\hat{\boldsymbol{\xi}}_1 + \widehat{gr}_{it2}\bar{\mathbf{z}}_i\hat{\boldsymbol{\rho}}_1)$$