CONTROL FUNCTION AND RELATED METHODS

Jeff Wooldridge
Michigan State University
LABOUR Lectures, EIEF
October 18-19, 2011

1. Linear-in-Parameters Models: IV versus Control Functions
2. Correlated Random Coefficient Models
3. Nonlinear Models
4. Semiparametric and Nonparametric Approaches
5. Methods for Panel Data

1. **Linear-in-Parameters Models**: **IV versus Control Functions**

- Most models that are linear are estimated using standard IV methods: two stage least squares (2SLS) or generalized method of moments (GMM).
- An alternative, the control function (CF) approach, relies on the same kinds of identification conditions. But even in models linear in parameters it can lead to different estimators.

- Let y_1 be the response variable, y_2 the endogenous explanatory variable (EEV), and $\mathbf{z}$ the $1 \times L$ vector of exogenous variables (with $z_1 = 1$):

$$y_1 = \mathbf{z}_1 \boldsymbol{\delta}_1 + \alpha_1 y_2 + u_1, \tag{1}$$

where $\mathbf{z}_1$ is a $1 \times L_1$ strict subvector of $\mathbf{z}$.

- First consider the exogeneity assumption

$$E(\mathbf{z}'u_1) = \mathbf{0}. \tag{2}$$

Reduced form for y_2:

$$y_2 = \mathbf{z}\boldsymbol{\pi}_2 + v_2,\ E(\mathbf{z}'v_2) = \mathbf{0} \tag{3}$$

where $\boldsymbol{\pi}_2$ is $L \times 1$. Write the linear projection of u_1 on v_2, in error form, as

$$u_1 = \rho_1 v_2 + e_1, \tag{4}$$

where $\rho_1 = E(v_2 u_1)/E(v_2^2)$ is the population regression coefficient. By construction, $E(v_2 e_1) = 0$ and $E(\mathbf{z}'e_1) = \mathbf{0}$.

Plug (4) into (1):

$$y_1 = \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + \rho_1 v_2 + e_1, \tag{5}$$

where v_2 is an explanatory variable in the equation. The new error, e_1, is uncorrelated with y_2 as well as with v_2 and $\mathbf{z}$.

• Suggests a two-step estimation procedure:

(i) Regress y_2 on $\mathbf{z}$ and obtain the reduced form residuals, $\hat{v}_2$.

(ii) Regress

$$y_1 \text{ on } \mathbf{z}_1, y_2, \text{ and } \hat{v}_2. \tag{6}$$

The implicit error in (6) is $e_{i1} + \rho_1 \mathbf{z}_i(\hat{\boldsymbol{\pi}}_2 - \boldsymbol{\pi}_2)$, which depends on the sampling error in $\hat{\boldsymbol{\pi}}_2$ unless $\rho_1 = 0$. OLS estimators from (6) will be consistent for $\boldsymbol{\delta}_1, \alpha_1$, and ρ_1.

- The OLS estimates from (6) are *control function* estimates.
- The OLS estimates of $\boldsymbol{\delta}_1$ and α_1 from (6) are *identical* to the 2SLS estimates starting from (1).
- A test of $H_0 : \rho_1 = 0$ in the equation

$$y_{i1} = \mathbf{z}_{i1}\boldsymbol{\delta}_1 + \alpha_1 y_{i2} + \rho_1 \hat{v}_{i2} + error_i$$

is the regression-based Hausman test for $H_0 : Cov(y_2, u_1) = 0$. Is easily made robust to heteroskedasticity of unknown form.

• The equivalence of IV and CF methods does not always. Add a quadratic in y_2:

$$y_1 = \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + \gamma_1 y_2^2 + u_1 \tag{7}$$

$$E(u_1|\mathbf{z}) = 0. \tag{8}$$

• Cannot get very far now without the stronger assumption (8).

• Let z_2 be a (nonbinary) scalar not also in $\mathbf{z}_1$. Under assumption (8), we can use, say, z_2^2 as an instrument for y_2^2. So the IVs would be $(\mathbf{z}_1, z_2, z_2^2)$ for $(\mathbf{z}_1, y_2, y_2^2)$. We could also use interactions $z_2\mathbf{z}_1$.

- What does the CF approach entail? Because of the nonlinearity in y_2, the CF approach is based on the conditional mean, $E(y_1|\mathbf{z}, y_2)$, rather than a linear projection.
- Therefore, we now *assume*

$$E(u_1|\mathbf{z}, y_2) = E(u_1|v_2) = \rho_1 v_2 \tag{9}$$

where

$$y_2 = \mathbf{z}\boldsymbol{\pi}_2 + v_2.$$

- Independence of (u_1, v_2) and $\mathbf{z}$ is sufficient for the first equality. Even under the independence assumption, linearity of $E(u_1|v_2)$ is a substantive restriction.

• Under $E(u_1|\mathbf{z}, y_2) = \rho_1 v_2$, we have

$$E(y_1|\mathbf{z}, y_2) = \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + \gamma_1 y_2^2 + \rho_1 v_2. \tag{10}$$

A CF approach is immediate: replace v_2 with $\hat{v}_2$ and use OLS on (10). *Not* equivalent to a 2SLS estimate.

• If the assumptions hold, CF likely more efficient; it is less robust than an IV approach, which requires only $E(u_1|\mathbf{z}) = 0$.

• At a minimum the CF approach requires $E(v_2|\mathbf{z}) = 0$ or $E(y_2|\mathbf{z}) = \mathbf{z}\boldsymbol{\pi}_2$, which puts serious restrictions on y_2.

• Even in linear models with constant coefficients, CF approaches can impose extra assumptions when we base it on $E(y_1|\mathbf{z},y_2)$, particularly when y_2 is (partially) discrete. Generally, the estimating equation is

$$E(y_1|\mathbf{z},y_2) = \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + E(u_1|\mathbf{z},y_2). \tag{11}$$

• Suppose y_2 is binary. Generally, $E(u_1|\mathbf{z},y_2)$ depends on the joint distribution of (u_1,y_2) given $\mathbf{z}$. If $y_2 = 1[\mathbf{z}\boldsymbol{\delta}_2 + e_2 \geq 0]$, (u_1,e_2) is independent of $\mathbf{z}$, $E(u_1|e_2) = \rho_1 e_2$, and $e_2 \sim Normal(0,1)$, then

$$E(u_1|\mathbf{z},y_2) = \rho_1[y_2\lambda(\mathbf{z}\boldsymbol{\delta}_2) - (1-y_2)\lambda(-\mathbf{z}\boldsymbol{\delta}_2)], \tag{12}$$

where $\lambda(\cdot)$ is the inverse Mills ratio (IMR).

• The CF approach is based on

$$E(y_1|\mathbf{z},y_2) = \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + \rho_1[y_2\lambda(\mathbf{z}\boldsymbol{\delta}_2) - (1-y_2)\lambda(-\mathbf{z}\boldsymbol{\delta}_2)]$$

and the Heckman two-step approach (for endogeneity, not sample selection):

(i) Probit to get $\hat{\boldsymbol{\delta}}_2$ and compute $\widehat{gr}_{i2} \equiv y_{i2}\lambda(\mathbf{z}_i\hat{\boldsymbol{\delta}}_2) - (1-y_{i2})\lambda(-\mathbf{z}_i\hat{\boldsymbol{\delta}}_2)$ (generalized residual).

(ii) Regress y_{i1} on $\mathbf{z}_{i1},\ y_{i2}, \widehat{gr}_{i2},\ i = 1,\ldots,N$ (and adjust the standard errors).

• Consistency of the CF estimators hinges on the model for $D(y_2|\mathbf{z})$ being correctly specified, along with linearity in $E(u_1|e_2)$. If we just apply 2SLS directly to $y_1 = \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + u_1$, it makes no distinction among discrete, continuous, or some mixture for y_2.

• How might we robustly use the binary nature of y_2 in IV estimation? Obtain the fitted probabilities, $\Phi(\mathbf{z}_i\hat{\boldsymbol{\delta}}_2)$, from the first stage probit, and then use these as IVs (not regressors!) for y_{i2}. Fully robust to misspecification of the probit model, usual standard errors from IV asymptotically valid. Efficient IV estimator if $P(y_2 = 1|\mathbf{z}) = \Phi(\mathbf{z}\boldsymbol{\delta}_2)$ and $Var(u_1|\mathbf{z}) = \sigma_1^2$.

• Similar suggestions work for y_2 a count variable or a corner solution.

2. Correlated Random Coefficient Models

- Modify the original equation as

$$y_1 = \eta_1 + \mathbf{z}_1\boldsymbol{\delta}_1 + a_1 y_2 + u_1 \tag{13}$$

where a_1, the "random coefficient" on y_2. Heckman and Vytlacil (1998) call (13) a correlated random coefficient (CRC) model. For a random draw i, $y_{i1} = \eta_1 + \mathbf{z}_{i1}\boldsymbol{\delta}_1 + a_{i1}y_2 + u_{i1}$.

- Write $a_1 = \alpha_1 + v_1$ where $\alpha_1 = E(a_1)$ is the parameter of interest. We can rewrite the equation as

$$y_1 = \eta_1 + \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + v_1 y_2 + u_1 \tag{14}$$

$$\equiv \eta_1 + \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + e_1. \tag{15}$$

• The potential problem with applying instrumental variables is that the error term $v_1 y_2 + u_1$ is not necessarily uncorrelated with the instruments $\mathbf{z}$, even under

$$E(u_1|\mathbf{z}) = E(v_1|\mathbf{z}) = 0. \tag{16}$$

We want to allow y_2 and v_1 to be correlated, $Cov(v_1, y_2) \equiv \tau_1 \neq 0$. A suffcient condition that allows for any *unconditional* correlation is

$$Cov(v_1, y_2|\mathbf{z}) = Cov(v_1, y_2), \tag{17}$$

and this is sufficient for IV to consistently estimate $(\alpha_1, \boldsymbol{\delta}_1)$.

- The usual IV estimator that ignores the randomness in a_1 is more robust than Garen's (1984) CF estimator, which adds $\hat{v}_2$ and $\hat{v}_2 y_2$ to the original model, or the Heckman-Vytlacil (1998) "plug-in" estimator, which replaces y_2 with $\hat{y}_2 = \mathbf{z}\hat{\boldsymbol{\pi}}_2$.
- The condition $Cov(v_1, y_2|\mathbf{z}) = Cov(v_1, y_2)$ cannot really hold for discrete y_2. Further, Card (2001) shows how it can be violated even if y_2 is continuous. Wooldridge (2005) shows how to allow parametric heteroskedasticity.

• In the case of binary y_2, we have what is often called the "switching regression" model. If $y_2 = 1[\mathbf{z}\boldsymbol{\delta}_2 + v_2 \geq 0]$ and $v_2|\mathbf{z} \sim Normal(0,1)$, then

$$
\begin{aligned}
E(y_1|\mathbf{z},y_2) = {} & \eta_1 + \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 \\
& + \rho_1 h_2(y_2,\mathbf{z}\boldsymbol{\delta}_2) + \xi_1 h_2(y_2,\mathbf{z}\boldsymbol{\delta}_2)y_2,
\end{aligned}
$$

where

$$
h_2(y_2,\mathbf{z}\boldsymbol{\delta}_2) = y_2\lambda(\mathbf{z}\boldsymbol{\delta}_2) - (1-y_2)\lambda(-\mathbf{z}\boldsymbol{\delta}_2)
$$

is the generalized residual function.

- Reminder: The expression for $E(y_1|\mathbf{z}, y_2)$ is an *estimating* equation for α_1. We do not use $E(y_1|\mathbf{z}, y_2)$, evaluated at $y_2 = 1$ and $y_2 = 0$, to obtain the treatment effect at different values of $\mathbf{z}$. The ATE in the model is constant and equal to α_1.

- Common to add the interactions $y_{i2}(\mathbf{z}_{i1} - \bar{\mathbf{z}}_1)$ (same as estimating $y_2 = 0$, $y_2 = 1$ separately) and then α_1 remains the average treatment effect (with the sample average $\bar{\mathbf{z}}_1$ replacing $E(\mathbf{z}_1)$.
- If $\boldsymbol{\delta}_1$ is replaced with random coefficients correlated with y_2, can interact $\mathbf{z}_1$ with $h_2(y_{i2}, \mathbf{z}_i\hat{\boldsymbol{\delta}}_2)$ under joint normality of the random coefficients and v_2.

- Can allow $E(v_1|v_2)$ to be more flexible [Heckman and MaCurdy (1986), Powell, Newey, and Walker (1990)].
- Also easy to allow for y_2 to follow a "heteroskedastic probit" model: replace v_2 with $e_2 = v_2/\exp(\mathbf{z}_2\boldsymbol{\gamma}_2)$ where $\exp(\mathbf{z}_2\boldsymbol{\gamma}_2) = sd(e_2|\mathbf{z})$. Estimate $\boldsymbol{\delta}_2, \boldsymbol{\gamma}_2$ by heteroskedastic probit.

3. Nonlinear Models

- Typically three approaches to nonlinear models with EEVs.

(1) Plug in fitted values from a first step regression (in an attempt to mimic 2SLS in linear model). More generally, try to find $E(y_1|\mathbf{z})$ or $D(y_1|\mathbf{z})$ and then impose identifying restrictions.

(2) CF approach: plug in residuals in an attempt to obtain $E(y_1|y_2,\mathbf{z})$ or $D(y_1|y_2,\mathbf{z})$.

(3) Maximum Likelihood (often limited information): Use models for $D(y_1|y_2,\mathbf{z})$ and $D(y_2|\mathbf{z})$ jointly.

- All strategies are more difficult with nonlinear models when y_2 is discrete. Some poor practices have lingered.

Binary and Fractional Responses

Probit model:

$$y_1 = 1[\mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + u_1 \geq 0], \tag{18}$$

where $u_1|z$ ~Normal$(0,1)$. Analysis goes through if we replace $(\mathbf{z}_1, y_2)$ with any known function $\mathbf{x}_1 \equiv \mathbf{g}_1(\mathbf{z}_1, y_2)$.

• The Rivers-Vuong (1988) approach [Smith and Blundell (1986) for Tobit] is to make a homoskedastic-normal assumption on the reduced form for y_2,

$$y_2 = \mathbf{z}\pi_2 + v_2,\ \ v_2|\mathbf{z} \sim Normal(0, \tau_2^2). \tag{19}$$

- RV approach comes close to requiring

$$(u_1, v_2) \text{ independent of } \mathbf{z}. \tag{20}$$

If we also assume

$$(u_1, v_2) \sim \text{Bivariate Normal} \tag{21}$$

with $\rho_1 = Corr(u_1, v_2)$, then we can proceed with MLE based on $f(y_1, y_2|\mathbf{z})$. A CF approach is available, too, based on

$$P(y_1 = 1|\mathbf{z}, y_2) = \Phi(\mathbf{z}_1\boldsymbol{\delta}_{\rho 1} + \alpha_{\rho 1}y_2 + \theta_{\rho 1}v_2) \tag{22}$$

where each coefficient is multiplied by $(1 - \rho_1^2)^{-1/2}$.

The RV two-step approach is

(i) OLS of y_2 on $\mathbf{z}$, to obtain the residuals, $\hat{v}_2$.

(ii) Probit of y_1 on $\mathbf{z}_1, y_2, \hat{v}_2$ to estimate the scaled coefficients. A simple t test on $\hat{v}_2$ is valid to test $H_0 : \rho_1 = 0$.

• Can recover the original coefficients, which appear in the partial effects. Or,

$$\widehat{ASF}(\mathbf{z}_1, y_2) = N^{-1} \sum_{i=1}^{N} \Phi(\mathbf{x}_1 \hat{\boldsymbol{\beta}}_{\rho 1} + \hat{\theta}_{\rho 1} \hat{v}_{i2}), \tag{23}$$

that is, we average out the reduced form residuals, $\hat{v}_{i2}$. This formulation is useful for more complicated models.

- The two-step CF approach easily extends to fractional responses:

$$E(y_1|\mathbf{z}, y_2, q_1) = \Phi(\mathbf{x}_1\boldsymbol{\beta}_1 + q_1), \tag{24}$$

where $\mathbf{x}_1$ is a function of $(\mathbf{z}_1, y_2)$ and q_1 contains unobservables. Can use the the *same* two-step because the Bernoulli log likelihood is in the linear exponential family. Still estimate scaled coefficients. APEs must be obtained from (23). To account for first-stage estimation, the bootstrap is convenient.

- Wooldridge (2005, Rothenberg Festschrift) describes some simple ways to make the analysis starting from (24) more flexible, including allowing $Var(q_1|v_2)$ to be heteroskedastic.

Example: Effects of school spending on student performance.

```
. sum math4 lunch rexpp found if y97

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       math4 |      1763    .6058803    .1966755       .029          1
       lunch |      2270    .3614616    .2535764      .0019      .9939
       rexpp |      2329    4261.201     789.124       1895      11779
       found |      2357    5895.984    1016.795       4816      10762

. glm math4 lrexpp lunch lenrol lrexpp94 if y97, fam(bin) link(probit) robust
note: math4 has noninteger values

Generalized linear models                          No. of obs      =      1208
Optimization     : ML                              Residual df     =      1203
                                                   Scale parameter =         1

                                                   AIC             =  .9079682
Log pseudolikelihood = -543.4128012                BIC             = -8378.243

------------------------------------------------------------------------------
             |               Robust
       math4 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lrexpp |   .2726134    .113923     2.39   0.017     .0493285    .4958983
       lunch |  -1.120498   .0582561   -19.23   0.000    -1.234678   -1.006318
      lenrol |   .0108851   .0390534     0.28   0.780    -.0656581    .0874283
    lrexpp94 |   .1854712   .0829582     2.24   0.025     .0228761    .3480664
       _cons |  -3.118799   .9207049    -3.39   0.001    -4.923347    -1.31425
------------------------------------------------------------------------------
```

```
. margeff

Average partial effects after glm
      y  = Pr(math4)

------------------------------------------------------------------------------
    variable |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lrexpp |   .0998487   .0416701     2.40   0.017     .0181768    .1815205
       lunch |  -.4103988    .020199   -20.32   0.000    -.4499882   -.3708094
      lenrol |   .0039868   .0143069     0.28   0.781    -.0240542    .0320278
    lrexpp94 |   .0679316   .0303795     2.24   0.025     .0083889    .1274742
------------------------------------------------------------------------------

. * Compare with OLS:

. reg math4 lrexpp lunch lenrol lrexpp94 if y97, robust

Linear regression                                      Number of obs =    1208
                                                       F(  4,  1203) =  106.45
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3114
                                                       Root MSE      =    .168

------------------------------------------------------------------------------
             |               Robust
       math4 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lrexpp |   .0934219   .0415276     2.25   0.025     .0119474    .1748964
       lunch |  -.4205896   .0215956   -19.48   0.000    -.4629588   -.3782204
      lenrol |  -.0002997   .0146181    -0.02   0.984    -.0289795    .0283802
    lrexpp94 |   .0590509   .0302958     1.95   0.052    -.0003877    .1184894
       _cons |  -.4838692   .3261972    -1.48   0.138    -1.123848    .1561094
------------------------------------------------------------------------------
```

```
. * Estimate the reduced form treating lrexpp as endogenous and
. * lfound as its IV:

. reg lrexpp lfound lunch lenrol lrexpp94 if y97 & e(sample)

      Source |       SS       df       MS              Number of obs =    1208
-------------+------------------------------           F(  4,  1203) =  244.58
       Model |  14.1288484     4   3.5322121           Prob > F      =  0.0000
    Residual |  17.3739986  1203  .014442227           R-squared     =  0.4485
-------------+------------------------------           Adj R-squared =  0.4467
       Total |   31.502847  1207  .026100122           Root MSE      =  .12018

------------------------------------------------------------------------------
      lrexpp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lfound |   .5443139   .0302413    18.00   0.000     .4849824    .6036454
       lunch |   .1419003   .0135798    10.45   0.000     .1152575    .1685431
      lenrol |  -.0961736   .0083231   -11.55   0.000     -.112503   -.0798442
    lrexpp94 |   .1358278   .0235749     5.76   0.000     .0895753    .1820803
       _cons |   3.036333   .2151718    14.11   0.000     2.614179    3.458486
------------------------------------------------------------------------------

. predict v2h, resid
(7694 missing values generated)
```

```
. glm math4 lrexpp lunch lenrol lrexpp94 v2h if y97, fam(bin) link(probit) robust
note: math4 has noninteger values

Generalized linear models                          No. of obs      =       1208
Optimization     : ML                              Residual df     =       1202
                                                   Scale parameter =          1
Deviance         =  157.5338068                    (1/df) Deviance =   .1310597
Pearson          =  146.5468582                    (1/df) Pearson  =   .1219192

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

                                                   AIC             =   .9083169
Log pseudolikelihood = -542.6234317                BIC             =  -8372.725

------------------------------------------------------------------------------
             |               Robust
       math4 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lrexpp |   .9567996   .2012636     4.75   0.000     .5623302    1.351269
       lunch |   -1.18315   .0585657   -20.20   0.000    -1.297937   -1.068363
      lenrol |   .0616713   .0399644     1.54   0.123    -.0166574         .14
    lrexpp94 |  -.0784249   .1063432    -0.74   0.461    -.2868536    .1300039
         v2h |  -.8593559   .2374808    -3.62   0.000     -1.32481   -.3939021
       _cons |  -6.966712   1.259552    -5.53   0.000    -9.435388   -4.498036
------------------------------------------------------------------------------

. * Easily reject null that spending is exogenous.
```

```
. margeff

Average partial effects after glm
      y  = Pr(math4)

------------------------------------------------------------------------------
    variable |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lrexpp |   .3501127   .0734569     4.77   0.000     .2061399    .4940855
       lunch |   -.432939   .0202452   -21.38   0.000    -.4726189    -.393259
      lenrol |   .0225668    .014639     1.54   0.123    -.0061252    .0512588
    lrexpp94 |  -.0286973   .0389071    -0.74   0.461    -.1049537    .0475592
         v2h |   -.314455    .086817    -3.62   0.000    -.4846132   -.1442968
------------------------------------------------------------------------------

. * Standard errors need to be fixed up for two-step estimation.
```

```
. ivreg math4 lunch lenrol lrexpp94 (lrexpp = lfound) if y97, robust

Instrumental variables (2SLS) regression               Number of obs =    1208
                                                       F(  4,  1203) =  114.28
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2908
                                                       Root MSE      =   .1705

------------------------------------------------------------------------------
             |               Robust
       math4 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lrexpp |   .3082997   .0710983     4.34   0.000     .1688093    .4477901
       lunch |  -.4389034   .0219076   -20.03   0.000    -.4818848   -.3959221
      lenrol |   .0155435   .0154313     1.01   0.314    -.0147317    .0458187
    lrexpp94 |   -.026911   .0393419    -0.68   0.494    -.1040974    .0502754
       _cons |   -1.66758   .4360251    -3.82   0.000    -2.523035   -.8121261
------------------------------------------------------------------------------
Instrumented:  lrexpp
Instruments:   lunch lenrol lrexpp94 lfound
------------------------------------------------------------------------------
```

• The control function approach has some decided advantages over another two-step approach – one that appears to mimic the 2SLS estimation of the linear model.

• Consider the binary response case. Rather than conditioning on v_2 along with $\mathbf{z}$ (and therefore y_2) to obtain $P(y_1 = 1|z, v_2) = P(y_1 = 1|\mathbf{z}, y_2, v_2)$, we can obtain $P(y_1 = 1|\mathbf{z})$.

• To find $P(y_1 = 1|\mathbf{z})$, we plug in the reduced form for y_2 to get $y_1 = 1[\mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1(\mathbf{z}\boldsymbol{\delta}_2) + \alpha_1 v_2 + u_1 > 0]$. Because $\alpha_1 v_2 + u_1$ is independent of $\mathbf{z}$ and normally distributed, $P(y_1 = 1|\mathbf{z}) = \Phi\{[\mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1(\mathbf{z}\boldsymbol{\delta}_2)]/\omega_1\}$. So first do OLS on the reduced form, and get fitted values, $\hat{y}_{i2} = \mathbf{z}_i\hat{\boldsymbol{\delta}}_2$. Then, probit of y_{i1} on $\mathbf{z}_{i1}, \hat{y}_{i2}$ to estimate scaled coefficients. Harder to estimate APEs and test for endogeneity.

• Danger with plugging in fitted values for y_2 is that one might be tempted to plug $\hat{y}_2$ into nonlinear functions, say y_2^2 or $y_2\mathbf{z}_1$. This does **not** result in consistent estimation of the scaled parameters or the partial effects. If we believe y_2 has a linear RF with additive normal error independent of $\mathbf{z}$, the addition of $\hat{v}_2$ solves the endogeneity problem regardless of how y_2 appears. Plugging in fitted values for y_2 only works in the case where the model is linear in y_2. Plus, the CF approach makes it much easier to test the null that for endogeneity of y_2 as well as compute APEs.

• Can understand the limits of CF approach by returning to $E(y_1|\mathbf{z}, y_2, q_1) = \Phi(\mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + q_1)$, where y_2 is discrete. Rivers-Vuong approach does not generally work.

• Suppose y_1 and y_2 are both binary and

$$y_2 = 1[\mathbf{z}\boldsymbol{\delta}_2 + v_2 \geq 0] \tag{25}$$

and we maintain joint normality of (u_1, v_2). We should *not* try to mimic 2SLS as follows: (i) Do probit of y_2 on $\mathbf{z}$ and get the fitted probabilities, $\hat{\Phi}_2 = \Phi(\mathbf{z}\hat{\boldsymbol{\delta}}_2)$. (ii) Do probit of y_1 on $\mathbf{z}_1, \hat{\Phi}_2$, that is, just replace y_2 with $\hat{\Phi}_2$.

- In general, the only strategy we have is maximum likelihood estimation based on $f(y_1|y_2,\mathbf{z})f(y_2|\mathbf{z})$. [Perhaps this is why some, such as Angrist (2001), Angrist and Pischke (2009), promote the notion of just using linear probability models estimated by 2SLS.]
- "Bivariate probit" software can be used to estimate the probit model with a binary endogenous variable.
- Parallel discussions hold for ordered probit, Tobit.

Multinomial Responses

• Recent push by Petrin and Train (2006), among others, to use control function methods where the second step estimation is something simple – such as multinomial logit, or nested logit – rather than being derived from a structural model. So, if we have reduced forms

$$\mathbf{y}_2 = \mathbf{z}\mathbf{\Pi}_2 + \mathbf{v}_2, \tag{26}$$

then we jump directly to convenient models for $P(y_1 = j|\mathbf{z}_1, \mathbf{y}_2, \mathbf{v}_2)$. The average structural functions are obtained by averaging the response probabilities across $\hat{\mathbf{v}}_{i2}$. No convincing way to handle discrete $\mathbf{y}_2$, though.

Exponential Models

• IV and CF approaches available for exponential models. Write

$$E(y_1|\mathbf{z}, y_2, r_1) = \exp(\mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + r_1), \tag{27}$$

where r_1 is the omitted variable. As usual, CF methods based on

$$E(y_1|\mathbf{z}, y_2) = \exp(\mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2)E[\exp(r_1)|\mathbf{z}, y_2].$$

For continuous y_2, can find $E(y_1|\mathbf{z}, y_2)$ when $D(y_2|\mathbf{z})$ is homoskedastic normal (Wooldridge, 1997) and when $D(y_2|\mathbf{z})$ follows a probit (Terza, 1998). In the probit case,

$$E(y_1|\mathbf{z}, y_2) = \exp(\mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2)h(y_2, \mathbf{z}\boldsymbol{\pi}_2, \theta_1)$$

$$h(y_2, \mathbf{z}\boldsymbol{\pi}_2, \theta_1) = \exp(\theta_1^2/2)\{y_2\Phi(\theta_1 + \mathbf{z}\boldsymbol{\pi}_2)/\Phi(\mathbf{z}\boldsymbol{\pi}_2) + (1 - y_2)[1 - \Phi(\theta_1 + \mathbf{z}\boldsymbol{\pi}_2)]/[1 - \Phi(\mathbf{z}\boldsymbol{\pi}_2)]\}.$$

- IV methods that work for any $\mathbf{y}_2$ are available [Mullahy (1997)]. If

$$E(y_1|\mathbf{z}, \mathbf{y}_2, r_1) = \exp(\mathbf{x}_1\boldsymbol{\beta}_1 + r_1) \tag{28}$$

and r_1 is independent of $\mathbf{z}$ then

$$E[\exp(-\mathbf{x}_1\boldsymbol{\beta}_1)y_1|\mathbf{z}] = E[\exp(r_1)|\mathbf{z}] = 1, \tag{29}$$

where $E[\exp(r_1)] = 1$ is a normalization. The moment conditions are

$$E[\exp(-\mathbf{x}_1\boldsymbol{\beta}_1)y_1 - 1|\mathbf{z}] = 0. \tag{30}$$

4. Semiparametric and Nonparametric Approaches

- Blundell and Powell (2004) show how to relax distributional assumptions on (u_1, v_2) in the model $y_1 = 1[\mathbf{x}_1\boldsymbol{\beta}_1 + u_1 > 0]$, where $\mathbf{x}_1$ can be any function of $(\mathbf{z}_1, y_2)$. Their key assumption is that y_2 can be written as $y_2 = g_2(\mathbf{z}) + v_2,$ where (u_1, v_2) is independent of $\mathbf{z}$, which rules out discreteness in y_2. Then

$$P(y_1 = 1|\mathbf{z}, v_2) = E(y_1|\mathbf{z}, v_2) = H(\mathbf{x}_1\boldsymbol{\beta}_1, v_2) \quad (31)$$

for some (generally unknown) function $H(\bullet,\bullet)$. The average structural function is just $ASF(\mathbf{z}_1, y_2) = E_{v_{i2}}[H(\mathbf{x}_1\boldsymbol{\beta}_1, v_{i2})]$.

- Two-step estimation: Estimate the function $g_2(\cdot)$ and then obtain residuals $\hat{v}_{i2} = y_{i2} - \hat{g}_2(\mathbf{z}_i)$. BP (2004) show how to estimate H and $\boldsymbol{\beta}_1$ (up to scaled) and $G(\cdot)$, the distribution of u_1. The ASF is obtained from $G(\mathbf{x}_1\boldsymbol{\beta}_1)$ or

$$\widehat{ASF}(\mathbf{z}_1, y_2) = N^{-1}\sum_{i=1}^{N}\hat{H}(\mathbf{x}_1\hat{\boldsymbol{\beta}}_1, \hat{v}_{i2}); \tag{32}$$

- Blundell and Powell (2003) allow $P(y_1 = 1|\mathbf{z}, y_2)$ to have general form $H(\mathbf{z}_1, y_2, v_2)$, and the second-step estimation is entirely nonparametric. Further, $\hat{g}_2(\cdot)$ can be fully nonparametric. Parametric approximations might produce good estimates of the APEs.

- BP (2003) consider a very general setup: $y_1 = g_1(\mathbf{z}_1, \mathbf{y}_2, u_1)$, with

$$ASF_1(\mathbf{z}_1, \mathbf{y}_2) = \int g_1(\mathbf{z}_1, \mathbf{y}_2, u_1) dF_1(u_1), \tag{33}$$

where F_1 is the distribution of u_1. Key restrictions are that $\mathbf{y}_2$ can be written as

$$\mathbf{y}_2 = \mathbf{g}_2(\mathbf{z}) + \mathbf{v}_2, \tag{34}$$

where $(u_1, \mathbf{v}_2)$ is independent of $\mathbf{z}$.

• Key: ASF can be obtained from $E(y_1|\mathbf{z}_1, \mathbf{y}_2, \mathbf{v}_2) = h_1(\mathbf{z}_1, \mathbf{y}_2, \mathbf{v}_2)$ by averaging out $\mathbf{v}_2$, and fully nonparametric two-step estimates are available.

• The focus on the ASF is liberating. It justifies flexible parametric approaches that need not be tied to "structural" equations. In particular, we can just skip modeling $g_1(\cdot)$ and start with $E(y_1|\mathbf{z}_1, \mathbf{y}_2, \mathbf{v}_2)$.

• For example, if y_1 is binary or a fraction and y_2 is a scalar,

$$\begin{aligned} E(y_1|\mathbf{z}_1, y_2, v_2) = \Phi(&\mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + \rho_1 v_2 + \eta_1 v_2^2 \\ &+ \mathbf{z}_1 v_2 \boldsymbol{\xi}_1 + \omega_1 y_2 v_2 + \ldots) \end{aligned}$$

5. Methods for Panel Data

- Combine methods for correlated random effects models with CF methods for nonlinear panel data models with unobserved heterogeneity and EEVs.
- Illustrate a parametric approach used by Papke and Wooldridge (2008), which applies to binary and fractional responses.
- Nothing appears to be known about applying “fixed effects” probit to estimate the fixed effects while also dealing with endogeneity. Likely to be poor for small T.

• Model with time-constant unobserved heterogeneity, c_{i1}, and time-varying unobservables, v_{it1}, as

$$E(y_{it1}|y_{it2},\mathbf{z}_i,c_{i1},v_{it1}) = \Phi(\alpha_1 y_{it2} + \mathbf{z}_{it1}\boldsymbol{\delta}_1 + c_{i1} + v_{it1}). \tag{35}$$

Allow the heterogeneity, c_{i1}, to be correlated with y_{it2} and $\mathbf{z}_i$, where $\mathbf{z}_i = (\mathbf{z}_{i1},\ldots,\mathbf{z}_{iT})$ is the vector of strictly exogenous variables (conditional on c_{i1}). The time-varying omitted variable, v_{it1}, is uncorrelated with $\mathbf{z}_i$ – strict exogeneity – but may be correlated with y_{it2}. As an example, y_{it1} is a female labor force participation indicator and y_{it2} is other sources of income.

• Write $\mathbf{z}_{it} = (\mathbf{z}_{it1}, \mathbf{z}_{it2})$, so that the time-varying IVs $\mathbf{z}_{it2}$ are excluded from the "structural."

• Chamberlain approach:

$$c_{i1} = \psi_1 + \bar{\mathbf{z}}_i \boldsymbol{\xi}_1 + a_{i1}, a_{i1} | \mathbf{z}_i \sim Normal(0, \sigma_{a_1}^2). \tag{36}$$

Next step:

$$E(y_{it1} | y_{it2}, \mathbf{z}_i, r_{it1}) = \Phi(\alpha_1 y_{it2} + \mathbf{z}_{it1} \boldsymbol{\delta}_1 + \psi_1 + \bar{\mathbf{z}}_i \boldsymbol{\xi}_1 + r_{it1})$$

where $r_{it1} = a_{i1} + v_{it1}$. Next, assume a linear reduced form for y_{it2}:

$$y_{it2} = \psi_2 + \mathbf{z}_{it} \delta_2 + \bar{\mathbf{z}}_i \boldsymbol{\xi}_2 + v_{it2}, t = 1, \ldots, T. \tag{37}$$

- Rules out discrete y_{it2} because

$$r_{it1} = \eta_1 v_{it2} + e_{it1}, \tag{38}$$

$$e_{it1}|(\mathbf{z}_i, v_{it2}) \sim Normal(0, \sigma_{e_1}^2), t = 1, \ldots, T. \tag{39}$$

Then

$$\begin{aligned} E(y_{it1}|\mathbf{z}_i, y_{it2}, v_{it2}) = \Phi(&\alpha_{e1} y_{it2} + \mathbf{z}_{it1}\boldsymbol{\delta}_{e1} \\ &+ \psi_{e1} + \bar{\mathbf{z}}_i \boldsymbol{\xi}_{e1} + \eta_{e1} v_{it2}) \end{aligned} \tag{40}$$

where the "e" subscript denotes division by $(1 + \sigma_{e_1}^2)^{1/2}$. This equation is the basis for CF estimation.

• Simple two-step procedure: (i) Estimate the reduced form for y_{it2} (pooled across t, or maybe for each t separately; at a minimum, different time period intercepts should be allowed). Obtain the residuals, $\hat{v}_{it2}$ for all (i, t) pairs. The estimate of $\boldsymbol{\delta}_2$ is the fixed effects estimate. (ii) Use the pooled probit (quasi)-MLE of y_{it1} on $y_{it2}, \mathbf{z}_{it1}, \bar{\mathbf{z}}_i, \hat{v}_{it2}$ to estimate $\alpha_{e1}, \boldsymbol{\delta}_{e1}, \psi_{e1}, \boldsymbol{\xi}_{e1}$ and η_{e1}.

• Delta method or bootstrapping (resampling cross section units) for standard errors. Can ignore first-stage estimation to test $\eta_{e1} = 0$ (but test should be fully robust to variance misspecification and serial independence).

- Estimates of average partial effects are based on the average structural function,

$$E_{(c_{i1},v_{it1})}[\Phi(\alpha_1 y_{t2} + \mathbf{z}_{t1}\boldsymbol{\delta}_1 + c_{i1} + v_{it1})], \tag{41}$$

which is consistently estimated as

$$N^{-1}\sum_{i=1}^{N}\Phi(\hat{\alpha}_{e1}y_{t2} + \mathbf{z}_{t1}\hat{\boldsymbol{\delta}}_{e1} + \hat{\psi}_{e1} + \bar{\mathbf{z}}_i\hat{\boldsymbol{\xi}}_{e1} + \hat{\eta}_{e1}\hat{v}_{it2}). \tag{42}$$

These APEs, typically with further averaging out across t and perhaps over y_{t2} and $\mathbf{z}_{t1}$, can be compared directly with fixed effects IV estimates.

EXAMPLE: **Effects of Spending on Test Pass Rates**

- Reform occurs between 1993/94 and 1994/95 school year; its passage was a surprise to almost everyone.
- Since 1994/95, each district receives a foundation allowance, based on revenues in 1993/94.
- Intially, all districts were brought up to a minimum allowance – $4,200 in the first year. The goal was to eventually give each district a basic allowance ($5,000 in the first year).
- Districts divided into three groups in 1994/95 for purposes of initial foundation allowance. Subsequent grants determined by statewide School Aid Fund.

- Catch-up formula for districts receiving below the basic. Initially, more than half of the districts received less than the basic allowance. By 1998/99, it was down to about 36%. In 1999/00, all districts began receiving the basic allowance, which was then $5,700. Two-thirds of all districts now receive the basic allowance.
- From 1991/92 to 2003/04, in the 10th percentile, expenditures rose from $4,616 (2004 dollars) to $7,125, a 54 percent increase. In the 50th percentile, it was a 48 percent increase. In the 90th percentile, per pupil expenditures rose from $7,132 in 1992/93 to $9,529, a 34 percent increase.

• Response variable: *math*4, the fraction of fourth graders passing the MEAP math test at a school.

• Spending variable is log(*avgrexppp*), where the average is over the current and previous three years.

• The linear model is

$$math4_{it} = \theta_t + \beta_1 \log(avgrexp_{it}) + \beta_2 lunch_{it} + \beta_3 \log(enroll_{it}) + c_{i1} + u_{it1}$$

Estimating this model by fixed effects is identical to adding the time averages of the three explanatory variables and using pooled OLS.

• The "fractional probit" model:

$$E(math4_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}) = \Phi(\theta_{at} + \mathbf{x}_{it}\boldsymbol{\beta}_a + \bar{\mathbf{x}}_i\boldsymbol{\xi}_a).$$

• Allowing spending to be endogenous. Controlling for 1993/94 spending, foundation grant should be exogenous. Exploit nonsmoothness in the grant as a function of initial spending.

$$math4_{it} = \theta_t + \beta_1 \log(avgrexp_{it}) + \beta_2 lunch_{it} + \beta_3 \log(enroll_{it})$$

$$+ \beta_{4t} \log(rexppp_{i,1994}) + \xi_1 \overline{lunch_i} + \xi_2 \overline{\log(enroll_i)} + v_{it1}$$

• And, fractional probit version of this.

```
. use meap92_01

. xtset distid year
       panel variable:  distid (strongly balanced)
        time variable:  year, 1992 to 2001
                delta:  1 unit

. des math4 avgrexp lunch enroll found

              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
math4           double %9.0g                  fraction satisfactory, 4th
                                                grade math
avgrexp         float  %9.0g                  (rexppp + rexppp_1 + rexppp_2 +
                                                rexppp_3)/4
lunch           float  %9.0g                  fraction eligible for free lunch
enroll          float  %9.0g                  district enrollment
found           int    %9.0g                  foundation grant, $: 1995-2001

. sum math4 rexppp lunch

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       math4 |      5010    .6149834    .1912023       .059          1
      rexppp |      5010     6331.99    1168.198   3553.361   15191.49
       lunch |      5010    .2802852    .1571325      .0087   .9126999
```

```
. xtreg math4 lavgrexp lunch lenroll y96-y01, fe cluster(distid)

Fixed-effects (within) regression               Number of obs      =      3507
Group variable: distid                          Number of groups   =       501

R-sq:  within  = 0.4713                         Obs per group: min =         7
       between = 0.0219                                        avg =       7.0
       overall = 0.2049                                        max =         7

                                                F(9,500)           =    171.93
corr(u_i, Xb)  = -0.1787                        Prob > F           =    0.0000

                               (Std. Err. adjusted for 501 clusters in distid)
------------------------------------------------------------------------------
             |               Robust
       math4 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |   .3770929   .0705668     5.34   0.000     .2384489    .5157369
       lunch |  -.0419467   .0731611    -0.57   0.567    -.1856877    .1017944
     lenroll |   .0020568   .0488107     0.04   0.966    -.0938426    .0979561
         y96 |  -.0155968   .0063937    -2.44   0.015    -.0281587   -.003035
         y97 |  -.0589732   .0095232    -6.19   0.000    -.0776837   -.0402628
         y98 |   .0781686   .0112949     6.92   0.000     .0559772    .1003599
         y99 |   .0642748   .0123103     5.22   0.000     .0400884    .0884612
         y00 |   .0895688   .0133223     6.72   0.000     .0633942    .1157434
         y01 |   .0630091    .014717     4.28   0.000     .0340943    .0919239
       _cons |  -2.640402   .8161357    -3.24   0.001     -4.24388   -1.036924
```

```
-------------+----------------------------------------------------------------
     sigma_u |   .1130256
     sigma_e |  .08314135
         rho |  .64888558   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. des alavgrexp alunch alenroll

              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
alavgrexp       float  %9.0g                  time average lavgrexp, 1995-2001
alunch          float  %9.0g                  time average lunch, 1995-2001
alenroll        float  %9.0g                  time average lenroll, 1995-2001
```

```
. reg math4 lavgrexp alavgrexp lunch alunch lenroll alenroll y96-y01,
      cluster(distid)

Linear regression                                      Number of obs =    3507
                                                       F( 12,   500) =  161.09
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4218
                                                       Root MSE      =  .11542

                                (Std. Err. adjusted for 501 clusters in distid)
------------------------------------------------------------------------------
             |               Robust
       math4 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |    .377092   .0705971     5.34   0.000     .2383884    .5157956
   alavgrexp |   -.286541   .0731797    -3.92   0.000    -.4303185   -.1427635
       lunch |  -.0419466   .0731925    -0.57   0.567    -.1857494    .1018562
      alunch |  -.3770088   .0766141    -4.92   0.000    -.5275341   -.2264835
     lenroll |   .0020566   .0488317     0.04   0.966     -.093884    .0979972
    alenroll |  -.0031646   .0491534    -0.06   0.949    -.0997373    .0934082
         y96 |  -.0155968   .0063965    -2.44   0.015    -.0281641   -.0030295
         y97 |  -.0589731   .0095273    -6.19   0.000    -.0776916   -.0402546
         y98 |   .0781687   .0112998     6.92   0.000     .0559678    .1003696
         y99 |    .064275   .0123156     5.22   0.000     .0400782    .0884717
         y00 |    .089569    .013328     6.72   0.000     .0633831    .1157548
         y01 |   .0630093   .0147233     4.28   0.000     .0340821    .0919365
       _cons |  -.0006233   .2450239    -0.00   0.998    -.4820268    .4807801
------------------------------------------------------------------------------
```

```
. * Now use fractional probit.

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

Generalized linear models                          No. of obs      =      3507
Optimization     : ML                              Residual df     =      3494
                                                   Scale parameter =         1
Deviance         =   237.643665                    (1/df) Deviance =  .0680148
Pearson          =  225.1094075                    (1/df) Pearson  =  .0644274

                               (Std. Err. adjusted for 501 clusters in distid)
------------------------------------------------------------------------------
             |               Robust
       math4 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |   .8810302   .2068026     4.26   0.000     .4757045    1.286356
   alavgrexp |  -.5814474   .2229411    -2.61   0.009    -1.018404   -.1444909
       lunch |  -.2189714   .2071544    -1.06   0.290    -.6249865    .1870437
      alunch |  -.9966635   .2155739    -4.62   0.000    -1.419181   -.5741465
     lenroll |   .0887804   .1382077     0.64   0.521    -.1821017    .3596626
    alenroll |  -.0893612   .1387674    -0.64   0.520    -.3613404    .1826181
         y96 |  -.0362309   .0178481    -2.03   0.042    -.0712125   -.0012493
         y97 |  -.1467327   .0273205    -5.37   0.000      -.20028   -.0931855
         y98 |   .2520084   .0337706     7.46   0.000     .1858192    .3181975
         y99 |   .2152507   .0367226     5.86   0.000     .1432757    .2872257
         y00 |   .3049632   .0399409     7.64   0.000     .2266805    .3832459
         y01 |   .2257321   .0439608     5.13   0.000     .1395705    .3118938
       _cons |  -1.855832   .7556621    -2.46   0.014    -3.336902   -.3747616
------------------------------------------------------------------------------
```

```
. margeff

Average partial effects after glm
      y  = Pr(math4)

------------------------------------------------------------------------------
    variable |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |   .2968496   .0695326     4.27   0.000     .1605682     .433131
   alavgrexp |  -.1959097   .0750686    -2.61   0.009    -.3430414   -.0487781
       lunch |  -.0737791   .0698318    -1.06   0.291    -.2106469    .0630887
      alunch |  -.3358104   .0723725    -4.64   0.000    -.4776579   -.1939629
     lenroll |   .0299132   .0465622     0.64   0.521     -.061347    .1211734
    alenroll |  -.0301089   .0467477    -0.64   0.520    -.1217326    .0615149
         y96 |  -.0122924   .0061107    -2.01   0.044    -.0242692   -.0003156
         y97 |  -.0508008   .0097646    -5.20   0.000     -.069939   -.0316625
         y98 |   .0809879   .0100272     8.08   0.000     .0613349    .1006408
         y99 |   .0696954   .0111375     6.26   0.000     .0478662    .0915245
         y00 |   .0970224   .0115066     8.43   0.000     .0744698     .119575
         y01 |   .0729829   .0132849     5.49   0.000      .046945    .0990208
------------------------------------------------------------------------------

. * These standard errors are very close to bootstrapped standard errors.
```

```
. xtgee math4 lavgrexp alavgrexp lunch alunch lenroll alenroll y96-y01,
        fa(bin) link(probit) corr(exch) robust

GEE population-averaged model                   Number of obs      =      3507
Group variable:                     distid      Number of groups   =       501
Link:                               probit      Obs per group: min =         7
Family:                           binomial                     avg =       7.0
Correlation:                  exchangeable                     max =         7
                                                Wald chi2(12)      =   1815.43
Scale parameter:                         1      Prob > chi2        =    0.0000

                                 (Std. Err. adjusted for clustering on distid)
------------------------------------------------------------------------------
             |             Semi-robust
       math4 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |    .884564   .2060662     4.29   0.000     .4806817    1.288446
   alavgrexp |  -.5835138   .2236705    -2.61   0.009      -1.0219   -.1451277
       lunch |  -.2372942   .2091221    -1.13   0.256    -.6471659    .1725775
      alunch |  -.9754696   .2170624    -4.49   0.000    -1.400904   -.5500351
     lenroll |   .0875629   .1387427     0.63   0.528    -.1843677    .3594935
    alenroll |  -.0820307   .1393712    -0.59   0.556    -.3551933    .1911318
         y96 |  -.0364771   .0178529    -2.04   0.041    -.0714681    -.001486
         y97 |  -.1471389   .0273264    -5.38   0.000    -.2006976   -.0935801
         y98 |   .2515377   .0337018     7.46   0.000     .1854833     .317592
         y99 |   .2148552   .0366599     5.86   0.000      .143003    .2867073
         y00 |   .3046286   .0399143     7.63   0.000     .2263981    .3828591
         y01 |   .2256619   .0438877     5.14   0.000     .1396437    .3116801
       _cons |  -1.914975   .7528262    -2.54   0.011    -3.390487   -.4394628
------------------------------------------------------------------------------
```

```
. margeff

Average partial effects after xtgee
      y  = Pr(math4)

------------------------------------------------------------------------------
    variable |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lavgrexp |   .2979576   .0692519     4.30   0.000     .1622263    .4336889
   alavgrexp |  -.1965515   .0752801    -2.61   0.009    -.3440978   -.0490052
       lunch |  -.0799305   .0704803    -1.13   0.257    -.2180693    .0582082
      alunch |  -.3285784   .0728656    -4.51   0.000    -.4713924   -.1857644
     lenroll |   .0294948   .0467283     0.63   0.528    -.0620909    .1210805
    alenroll |  -.0276313   .0469381    -0.59   0.556    -.1196283    .0643656
         y96 |   -.012373   .0061106    -2.02   0.043    -.0243497   -.0003964
         y97 |  -.0509306   .0097618    -5.22   0.000    -.0700633   -.0317979
         y98 |   .0808226    .010009     8.08   0.000     .0612054    .1004399
         y99 |   .0695541   .0111192     6.26   0.000     .0477609    .0913472
         y00 |   .0968972   .0115004     8.43   0.000     .0743568    .1194376
         y01 |   .0729416   .0132624     5.50   0.000     .0469478    .0989353
------------------------------------------------------------------------------
```

```
. * Now allow spending to be endogenous. Use foundation allowance, and
. * interactions, as 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
         y97 |  -1.049922    .376541    -2.79   0.005     -1.78972   -.3101244
         y98 |  -.4548311   .4958826    -0.92   0.359    -1.429102    .5194394
         y99 |  -.4360973   .5893671    -0.74   0.460    -1.594038    .7218439
         y00 |  -.3559283   .6509999    -0.55   0.585    -1.634961     .923104
         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
     le94y97 |     .11487   .0425422     2.70   0.007     .0312865    .1984534
     le94y98 |   .0599439   .0554377     1.08   0.280    -.0489757    .1688636
     le94y99 |   .0557854   .0661784     0.84   0.400    -.0742367    .1858075
     le94y00 |    .048899   .0727172     0.67   0.502    -.0939699    .1917678
```

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

. * Estimate is substantially larger than when spending is treated as exogenous.
```

```
. * 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. Should
. * use the panel bootstrap accounting for both stages.
. * Only marginal evidence that spending is endogenous, but the negative sign
. * fits the story that districts increase spending when performance is
. * (expected to be) worse, based on unobservables (to us).
```

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)	—
Scale Factor	—	.337	