New Developments in Econometrics
Lecture 14: Control Function and Related Methods

Jeff Wooldridge

Cemmap Lectures, UCL, June 2009

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

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

- Most models that are linear in parameters 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.
- 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 $\mathbf{z}$.

• Two-step 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 $\rho_1 = 0$ (exogeneity test). 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).
- Now extend the model:

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

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

Let z_2 be a (nonbinary) scalar not also in $\mathbf{z}_1$. Under the (8) – which is stronger than (2), and is essential for nonlinear models – 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 $(\mathbf{z}_1, y_2, y_2^2)$.

• What does 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 independence of (u_1, v_2) and $\mathbf{z}$ is sufficient for the first equality. Even under the independence assumption, linearity 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 notably more efficient, it is notably but less robust than an IV approach.

- 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 $\mathbf{z}_{i1}, y_{i2}, \widehat{gr}_{i2}, i = 1, \ldots, N$.

• 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.

- Write $a_1 = \alpha_1 + v_1$ where $\alpha_1 = E(a_1)$ is the object 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 $v_2|\mathbf{z}$ is $Normal(0,1)$, 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 evaluate 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 and Limitations of the CF Approach

• 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,

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

- 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 $\mathrm{H}_0 : \rho_1 = 0$.

• 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). In inference, we should only assume the mean is correctly specified.method can be used in the binary and fractional cases. To account for first-stage estimation, the bootstrap is convenient.

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

• 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. 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 the latter probability, 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 $\mathbf{z}_{i1}, \hat{y}_{i2}$. 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. Can also justify flexible parametric approaches and just skip modeling $g_1(\cdot)$.

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:

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

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.

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	