What’s New in Econometrics?
Lecture 6
Control Functions and Related Methods

Jeff Wooldridge
NBER Summer Institute, 2007

1. Linear-in-Parameters Models: IV versus Control Functions
2. Correlated Random Coefficient Models
3. Some Common Nonlinear Models and Limitations of the CF Approach
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 $z$ the $1 \times L$ vector of exogenous variables (with $z_1 = 1$):

$$y_1 = z_1 \delta_1 + \alpha_1 y_2 + u_1,$$

where $z_1$ is a $1 \times L_1$ strict subvector of $z$. First consider the exogeneity assumption
\[ E(z'u_1) = 0. \] \hspace{1cm} (2)

Reduced form for \( y_2 \):
\[ y_2 = z\pi_2 + v_2, \hspace{0.5cm} E(z'v_2) = 0 \] \hspace{1cm} (3)

where \( \pi_2 \) is \( L \times 1 \). Write the linear projection of \( u_1 \) on \( v_2 \), in error form, as
\[ u_1 = \rho_1 v_2 + e_1, \] \hspace{1cm} (4)

where \( \rho_1 = E(v_2u_1)/E(v_2^2) \) is the population regression coefficient. By construction, \( E(v_2e_1) = 0 \) and \( E(z'e_1) = 0 \).

Plug (4) into (1):
\[ y_1 = z_1\delta_1 + \alpha_1y_2 + \rho_1v_2 + e_1, \] \hspace{1cm} (5)

where we now view \( v_2 \) as an explanatory variable in the equation. By controlling for \( v_2 \), the error \( e_1 \) is uncorrelated with \( y_2 \) as well as with \( v_2 \) and \( z \).

- Two-step procedure: (i) Regress \( y_2 \) on \( z \) and
obtain the reduced form residuals, $\hat{v}_2$; (ii) Regress
\[ y_1 \text{ on } z_1, y_2, \text{ and } \hat{v}_2. \tag{6} \]

The implicit error in (6) is
\[ e_{i1} + \rho_1 z_i (\hat{\pi}_2 - \pi_2), \]
which depends on the sampling error in $\hat{\pi}_2$ unless $\rho_1 = 0$. OLS estimators from (6) will be consistent for $\delta_1, \alpha_1,$ and $\rho_1$. Simple test for null of exogeneity is (heteroskedasticity-robust) $t$ statistic on $\hat{v}_2$.

- The OLS estimates from (6) are control function estimates.
- The OLS estimates of $\delta_1$ and $\alpha_1$ from (6) are identical to the 2SLS estimates starting from (1).
- Now extend the model:
\[ y_1 = z_1 \delta_1 + \alpha_1 y_2 + \gamma_1 y_2^2 + u_1 \tag{7} \]
\[ E(u_1|z) = 0. \tag{8} \]
Let $z_2$ be a scaler not also in $z_1$. 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 $(z_1, z_2, z_2^2)$ for $(z_1, y_2, y_2^2)$.

- What does CF approach entail? We require an assumption about $E(u_1|z, y_2)$, say

$$E(u_1|z, y_2) = E(u_1|v_2) = \rho_1 v_2, \quad (9)$$

where the first equality would hold if $(u_1, v_2)$ is independent of $z$ – a nontrivial restriction on the reduced form error in (3), not to mention the structural error $u_1$. Linearity of $E(u_1|v_2)$ is a substantive restriction. Now,

$$E(y_1|z, y_2) = z_1 \delta_1 + \alpha_1 y_2 + \gamma_1 y_2^2 + \rho_1 v_2, \quad (10)$$

and a CF approach is immediate: replace $v_2$ with $\hat{v}_2$.
and use OLS on (10).

- These CF estimates are not the same as the 2SLS estimates using any choice of instruments for \((y_2, y_2^2)\). CF approach likely more efficient, but less robust. For example, (8) implies \(E(y_2|z) = z\pi_2\).

- CF approaches can impose extra assumptions even in the simple model (1). For example, if \(y_2\) is a binary response, the CF approach based on \(E(y_1|z, y_2)\) involves estimating

\[
E(y_1|z, y_2) = z_1\delta_1 + \alpha_1 y_2 + E(u_1|z, y_2). \tag{11}
\]

If \(y_2 = 1[z\delta_2 + e_2 \geq 0]\), \((u_1, e_2)\) is independent of \(z\), \(E(u_1|e_2) = \rho_1 e_2\), and \(e_2 \sim \text{Normal}(0, 1)\), then

\[
E(u_1|z, y_2) = \rho_1[y_2\lambda(z\delta_2) - (1 - y_2)\lambda(-z\delta_2)], \tag{12}
\]

where \(\lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)\) is the inverse Mills ratio (IMR). This leads to the Heckman two-step
estimate (for endogeneity, not sample selection). Obtain the probit estimate $\hat{\delta}_2$ and add the “generalized residual,”

$$\hat{gr}_{i2} = y_{i2} \lambda(z_i \hat{\delta}_2) - (1 - y_{i2}) \lambda(-z_i \hat{\delta}_2)$$
as a regressor: $y_{i1}$ on $z_{i1}, y_{i2}, \hat{gr}_{i2}, i = 1, \ldots, N$.

- Consistency of the CF estimators hinges on the model for $D(y_2|z)$ being correctly specified, along with linearity in $E(u_1|v_2)$. If we just apply 2SLS directly to (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, $\Phi(z_i \hat{\delta}_2)$, from the first stage probit, and then use these as IVs for $y_{i2}$. This is fully robust to misspecification of the probit model and the usual standard errors from IV are asymptotically valid. It
is the efficient IV estimator if
\[ P(y_2 = 1|z) = \Phi(z\delta_2) \] and \[ Var(u_1|z) = \sigma_1^2. \]

2. Correlated Random Coefficient Models

Modify (1) as
\[ y_1 = \eta_1 + z_1\delta_1 + a_1y_2 + u_1, \tag{13} \]

where \( a_1 \), the “random coefficient” on \( y_2 \). Think of \( a_1 \) as an omitted variable that interacts with \( y_2 \). Following Heckman and Vytlacil (1998), we refer to (13) as a correlated random coefficient (CRC) model.

- Write \( a_1 = \alpha_1 + \nu_1 \) where \( \alpha_1 = E(a_1) \) is the object of interest. We can rewrite the equation as
\[ y_1 = \eta_1 + z_1\delta_1 + \alpha_1y_2 + \nu_1y_2 + u_1 \tag{14} \equiv \eta_1 + z_1\delta_1 + \alpha_1y_2 + e_1, \tag{15} \]

- The potential problem with applying instrumental variables to (15) is that the error term \( \nu_1y_2 + u_1 \) is
not necessarily uncorrelated with the instruments $z$, even under

$$E(u_1|z) = E(v_1|z) = 0. \quad (16)$$

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

$$\text{Cov}(v_1, y_2|z) = \text{Cov}(v_1, y_2), \quad (17)$$

and this is sufficient for IV to consistently estimate $(\alpha_1, \delta_1)$.

- 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 $\hat{y}_2 = z\hat{\pi}_2$. See notes.
• Condition (17) cannot really hold for discrete $y_2$. 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 $y_2 = 1[z \delta_2 + \nu_2 \geq 0]$ and $\nu_2|z$ is Normal$(0, 1)$, then

$$E(y_1|z, y_2) = \eta_1 + z_1 \delta_1 + a_1 y_2$$

$$+ \rho_1 h_2(y_2, z \delta_2) + \xi_1 h_2(y_2, z \delta_2) y_2,$$

where

$$h_2(y_2, z \delta_2) = y_2 \lambda(z \delta_2) - (1 - y_2) \lambda(-z \delta_2)$$

is the generalized residual function. The two-step estimation method is the one due to Heckman (1976).

• Can also interact the exogenous variables with $h_2(y_{i2}, z_i \hat{\delta}_2)$. Or, allow $E(\nu_1|\nu_2)$ to be more
flexible, as in Heckman and MaCurdy (1986).

3. Some Common Nonlinear Models and Limitations of the CF Approach

• CF approaches are more difficult to apply to nonlinear models, even relatively simple ones. Methods are available when the endogenous explanatory variables are continuous, but few if any results apply to cases with discrete $y_2$.

Binary and Fractional Responses

Probit model:

$$ y_1 = 1[z_1 \delta_1 + \alpha_1 y_2 + u_1 \geq 0], $$

where $u_1|z \sim \text{Normal}(0, 1)$. Analysis goes through if we replace $(z_1, y_2)$ with any known function $x_1 \equiv g_1(z_1, y_2)$.

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

$$y_2 = \mathbf{z} \pi_2 + \upsilon_2, \quad \upsilon_2 | \mathbf{z} \sim \text{Normal}(0, \tau_2^2). \quad (19)$$

A key point is that the RV approach essentially requires

$$(u_1, \upsilon_2) \text{ independent of } \mathbf{z}. \quad (20)$$

If we also assume

$$(u_1, \upsilon_2) \sim \text{Bivariate Normal} \quad (21)$$

with $\rho_1 = \text{Corr}(u_1, \upsilon_2)$, then we can proceed with MLE based on $f(y_1, y_2 | \mathbf{z})$. A CF approach is available, too, based on

$$P(y_1 = 1 | \mathbf{z}, y_2) = \Phi(\mathbf{z}_1 \delta_{\rho_1} + \alpha_{\rho_1} y_2 + \theta_{\rho_1} y_2) \quad (22)$$

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

The RV two-step approach is
(i) OLS of $y_2$ on $z$, to obtain the residuals, $\hat{v}_2$.
(ii) Probit of $y_1$ on $z_1, y_2, \hat{v}_2$ 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,

$$\text{ASF}(z_1, y_2) = N^{-1} \sum_{i=1}^{N} \Phi(x_1 \tilde{\beta}_{\rho_1} + \hat{\theta}_{\rho_1} \hat{v}_{i2}), \quad (23)$$

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:

$$E(y_1|z, y_2, q_1) = \Phi(x_1 \beta_1 + q_1), \quad (24)$$

where $x_1$ is a function of $(z_1, y_2)$ 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 $\text{Var}(q_1|v_2)$ 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 \( z \) (and therefore \( y_2 \)) to obtain

\[
P(y_1 = 1|z, v_2) = P(y_1 = 1|z, y_2, v_2),
\]
we can obtain \( P(y_1 = 1|z) \). To find the latter probability, we plug in the reduced form for \( y_2 \) to get

\[
y_1 = 1[z_1 \delta_1 + \alpha_1(z_2\delta_2) + \alpha_1 v_2 + u_1 > 0].
\]

Because \( \alpha_1 v_2 + u_1 \) is independent of \( z \) and normally distributed, \( P(y_1 = 1|z) = \Phi\{[z_1 \delta_1 + \alpha_1(z_2\delta_2)]/\omega_1\} \).

So first do OLS on the reduced form, and get fitted values, \( \hat{y}_{i2} = z_i \hat{\delta}_2 \). Then, probit of \( y_{i1} \) on \( z_{i1}, \hat{y}_{i2} \).

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 \( y_2 z_1 \). 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 \( z \), 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.

- Extension to random coefficients:

$$E(y_1|z,y_2,c_1) = \Phi(z_1\delta_1 + a_1y_2 + q_1), \quad (25)$$

where $a_1$ is random with mean $\alpha_1$ and $q_1$ again has mean of zero. If we want the partial effect of $y_2$, evaluated at the mean of heterogeneity, is

$$\alpha_1\phi(z_1\delta_1 + \alpha_1y_2). \quad (26)$$

The APE in this case is much messier.

- Could just implement flexible CF approaches without formally starting with a “structural” model.
For example, could just do Bernoulli QMLE of $y_{i1}$ on $z_{i1}$, $y_{i2}$, $\hat{v}_{i2}$, and $y_{i2}\hat{v}_{i2}$. Even here, APE can be different sign from $\alpha_1$.

• Lewbel (2000) has made some progress in estimating parameters up to scale in the model $y_1 = 1[z_1\delta_1 + \alpha_1 y_2 + u_1 > 0]$, where $y_2$ might be correlated with $u_1$ and $z_1$ is a $1 \times L_1$ vector of exogenous variables. Let $z$ be the vector of all exogenous variables uncorrelated with $u_1$. Then Lewbel requires a continuous element of $z_1$ with nonzero coefficient – say the last element, $z_{L_1}$ – that does not appear in $D(u_1|y_2, z)$ or $D(y_2|z)$. ($y_2$ cannot play the role) Cannot be an instrument as we usually think of it. Can be a variable randomized to be independent of $y_2$ and $z$.

• Returning to the response function
\( E(y_1|z, y_2, q_1) = \Phi(x_1\beta_1 + q_1) \), we can understand the limits of the CF approach for estimating nonlinear models with discrete EEVs. The Rivers-Vuong approach does not work. We cannot write \( D(y_2|z) = \text{Normal}(z\pi_2, \tau_2^2) \). There are no known two-step estimation methods that allow one to estimate a probit model or fractional probit model with discrete \( y_2 \), even if we make strong distributional assumptions.

- There are some poor strategies that still linger.

Suppose \( y_1 \) and \( y_2 \) are both binary and

\[
y_2 = 1[z\delta_2 + \nu_2 \geq 0]
\]  

(27)

and we maintain joint normality of \((u_1, \nu_2)\). We should not try to mimic 2SLS as follows: (i) Do probit of \( y_2 \) on \( z \) and get the fitted probabilities, \( \hat{\Phi}_2 = \Phi(z\hat{\delta}_2) \). (ii) Do probit of \( y_1 \) on \( z_1, \hat{\Phi}_2 \), that is,
just replace $y_2$ with $\hat{\Phi}_2$.

- Currently, the only strategy we have is maximum likelihood estimation based on $f(y_1|y_2,z)f(y_2|z)$.

(Perhaps this is why some, such as Angrist (2001), promote the notion of just using linear probability models estimated by 2SLS.)

- Yes, “bivariate” probit software be used to estimate the probit model with a binary endogenous variable. In fact, with any function of $z_1$ and $y_2$ as explanatory variables.

- Parallel discussions hold for ordered probit, Tobit.

**Multinomial Responses**

- Recent push, by Villas-Boas (2005) and 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

\[ y_2 = z\Pi_2 + v_2, \]  

(28)

then we jump directly to convenient models for \( P(y_1 = j|z_1, y_2, v_2) \). The average structural functions are obtained by averaging the response probabilities across \( \hat{v}_{i2} \). No convincing way to handle discrete \( y_2 \), though.

**Exponential Models**

- Both IV approaches and CF approaches are available for exponential models. With a single EEV, write

\[ E(y_1|z, y_2, r_1) = \exp(z_1\delta_1 + \alpha_1y_2 + r_1), \]  

(29)

where \( r_1 \) is the omitted variable. (Extensions to
general nonlinear functions \( x_1 = g_1(z_1, y_2) \) are immediate; we just add those functions with linear coefficients to (29). CF methods based on

\[
E(y_1|z, y_2, r_1) = \exp(z_1 \delta_1 + \alpha_1 y_2) E[\exp(r_1)|z, y_2]
\]

This has been worked through when \( D(y_2|z) \) is homoskedastic normal (Wooldridge, 1997 – see notes for a random coefficient version where \( \alpha_1 \) becomes \( a_1 \) with \( E(a_1) = \alpha_1 \) and \( D(y_2|z) \) follows a probit (Terza, 1998). In the latter case,

\[
E(y_1|z, y_2) = \exp(z_1 \delta_1 + \alpha_1 y_2) h(y_2, z \pi_2, \theta_1)
\]

\[
h(y_2, z \pi_2, \theta_1) = \exp(\theta_1^2/2) \{ y_2 \Phi(\theta_1 + z \pi_2)/\Phi(z \pi_2) \\
+ (1 - y_2)[1 - \Phi(\theta_1 + z \pi_2)]/[1 - \Phi(z \pi_2)] \} 
\]

• IV methods that work for any \( y_2 \) are also available, as developed by Mullahy (1997). If

\[
E(y_1|z, y_2, r_1) = \exp(x_1 \beta_1 + r_1)
\]  

(30)
and $r_1$ is independent of $z$ then

$$E[\exp(-x_1 \beta_1) y_1 | z] = E[\exp(r_1) | z] = 1, \quad (31)$$

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

$$E[\exp(-x_1 \beta_1) y_1 - 1 | z] = 0. \quad (32)$$

4. Semiparametric and Nonparametric Approaches

Blundell and Powell (2004) show how to relax distributional assumptions on $(u_1, v_2)$ in the model $y_1 = 1[x_1 \beta_1 + u_1 > 0]$, where $x_1$ can be any function of $(z_1, y_2)$. Their key assumption is that $y_2$ can be written as $y_2 = g_2(z) + v_2$, where $(u_1, v_2)$ is independent of $z$, which rules out discreteness in $y_2$. Then

$$P(y_1 = 1 | z, v_2) = E(y_1 | z, v_2) = H(x_1 \beta_1, v_2) \quad (33)$$
for some (generally unknown) function $H(\cdot, \cdot)$. The average structural function is just

$$\text{ASF}(z_1, y_2) = E_{v_{i2}}[H(x_1 \beta_1, v_{i2})].$$

- Two-step estimation: Estimate the function $g_2(\cdot)$ and then obtain residuals $\hat{v}_{i2} = y_{i2} - \hat{g}_2(z_i)$. BP (2004) show how to estimate $H$ and $\beta_1$ (up to scaled) and $G(\cdot)$, the distribution of $u_1$. The ASF is obtained from $G(x_1 \beta_1)$ or

$$\hat{\text{ASF}}(z_1, y_2) = N^{-1} \sum_{i=1}^{N} \hat{H}(x_1 \hat{\beta}_1, \hat{v}_{i2}); \quad (34)$$

- Blundell and Powell (2003) allow $P(y_1 = 1|z, y_2)$ to have the general form $H(z_1, y_2, v_2)$, and then the second-step estimation is also entirely nonparametric. They also allow $\hat{g}_2(\cdot)$ to be fully nonparametric. Parametric approximations in each stage might produce good estimates of the APEs.
BP (2003) consider a very general setup, which starts with $y_1 = g_1(z_1, y_2, u_1)$, and then discuss estimation of the ASF, given by

$$ASF_1(z_1, y_2) = \int g_1(z_1, y_2, u_1) dF_1(u_1),$$  \hspace{1cm} (35)

where $F_1$ is the distribution of $u_1$. The key restrictions are that $y_2$ can be written as

$$y_2 = g_2(z) + v_2,$$  \hspace{1cm} (36)

where $(u_1, v_2)$ is independent of $z$. The key is that the ASF can be obtained from $E(y_1|z_1, y_2, v_2) = h_1(z_1, y_2, v_2)$ by averaging out $v_2$, and fully nonparametric two-step estimates are available.

Provides justification for the parametric versions discussed earlier, where the step of modeling $g_1(\cdot)$ in $y_1 = g_1(z_1, y_2, u_1)$ can be skipped.
• Imbens and Newey (2006) consider the triangular system, but without additivity in the reduced form of $y_2$,

$$y_2 = g_2(z, e_2), \quad (37)$$

where $g_2(z, \cdot)$ is strictly monotonic. Rules out discrete $y_2$ but allows some interaction between the unobserved heterogeneity in $y_2$ and the exogenous variables. When $(u_1, e_2)$ is independent of $z$, a valid control function to be used in a second stage is $v_2 = F_{y_2|z}(y_2|z)$, where $F_{y_2|z}$ is the conditional distribution of $y_2$ given $z$.

5. Methods for Panel Data

• Combine methods for handling correlated random effects models with control function methods to estimate certain nonlinear panel data models with unobserved heterogeneity and EEVs.
Illustrate a parametric approach used by Papke and Wooldridge (2007), which applies to binary and fractional responses.

In this model, 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 $T$. Perhaps jackknife methods can be adapted, but currently the assumptions are very strong (serial independence, homogeneity over time, exogenous regressors).

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

$$E(y_{it1}|y_{it2}, z_i, c_{i1}, v_{it1}) = \Phi(\alpha_1 y_{it2} + z_{it1} \delta_1 + c_{i1} + v_{it1}).$$

(38)

Allow the heterogeneity, $c_{i1}$, to be correlated with
\( y_{it2} \) and \( z_i \), where \( z_i = (z_{i1}, \ldots, z_{iT}) \) is the vector of strictly exogenous variables (conditional on \( c_{i1} \)).

The time-varying omitted variable, \( v_{it1} \), is uncorrelated with \( z_i \) – 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 \( z_{it} = (z_{it1}, z_{it2}) \), so that the time-varying IVs \( z_{it2} \) are excluded from the “structural.”

- Chamberlain approach:

\[
\begin{align*}
    c_{i1} &= \psi_1 + \bar{z}_i \xi_1 + a_{i1}, a_{i1}|z_i \sim \text{Normal}(0, \sigma_{a_1}^2). \tag{39}
\end{align*}
\]

We could allow the elements of \( z_i \) to appear with separate coefficients, too. Note that only exogenous variables are included in \( \bar{z}_i \). Next step:

\[
    E(y_{it1}|y_{it2}, z_i, r_{it}) = \Phi(\alpha_1 y_{it2} + z_{it1} \delta_1 + \psi_1 + \bar{z}_i \xi_1 + r_{it1})
\]
where \( r_{it1} = a_{i1} + v_{it1} \). Next, we assume a linear reduced form for \( y_{it2} \):

\[
y_{it2} = \psi_2 + z_{it} \delta_2 + \bar{z}_i \xi_2 + v_{it2}, t = 1, \ldots, T. \tag{40}
\]

Rules out discrete \( y_{it2} \) because

\[
r_{it1} = \eta_1 v_{it2} + e_{it1}, \tag{41}
\]

\[
e_{it1}|(z_i, v_{it2}) \sim \text{Normal}(0, \sigma_{e_1}^2), t = 1, \ldots, T. \tag{42}
\]

Then

\[
E(y_{it1}|z_i, y_{it2}, v_{it2}) = \Phi(\alpha_{e1} y_{it2} + z_{it1} \delta_{e1} + \psi_{e1} + \bar{z}_i \xi_{e1} + \eta_{e1} v_{it2}) \tag{43}
\]

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 $\delta_2$ is
the fixed effects estimate. (ii) Use the pooled probit
(quasi)-MLE of $y_{it1}$ on $y_{it2}, z_{it1}, \bar{z}_i, \hat{v}_{it2}$ to estimate
$\alpha_{e1}, \delta_{e1}, \psi_{e1}, \xi_{e1}$ and $\eta_{e1}$.

- 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,

$$E_{(c_{i1}, v_{it1})}[\Phi(\alpha_1 y_{t2} + z_{t1} \delta_1 + c_{i1} + v_{it1})],$$

which is consistently estimated as
\[ N^{-1} \sum_{i=1}^{N} \Phi(\hat{\alpha}_e y_{it2} + \mathbf{z}_{it1} \hat{\delta}_e + \hat{\psi}_e + \bar{z}_i \hat{\xi}_e + \hat{\eta}_e \hat{v}_{it2}). \] (45)

These APEs, typically with further averaging out across \( t \) and the values of \( y_{it2} \) and \( \mathbf{z}_{it1} \), can be compared directly with fixed effects IV estimates.

- We can use the approaches of Altonji and Matzkin (2005), Blundell and Powell (2003), and Imbens and Newey (2006) to make the analysis less parametric. For example, we might replace (40) with \( y_{it2} = g_2(\mathbf{z}_{it}, \bar{z}_i) + \nu_{it2} \) or \( y_{it2} = g_2(\mathbf{z}_{it}, \bar{z}_i, e_{it2}) \) under monotonicity in \( e_2 \). Then a reasonable assumption is

\[ D(c_{i1} + \nu_{it1}|\mathbf{z}_i, y_{it2}, \nu_{it2}) = D(c_{i1} + \nu_{it1}|\bar{z}_i, \nu_{it2}) \] (46)

where, in the Imbens and Newey case,

\[ \nu_{it2} = F_{y_{it2}|(\mathbf{z}_i, \bar{z})}(y_{it2}|\mathbf{z}_{it}, \bar{z}_i). \] After a first stage
estimation, the ASF can be obtained by estimating 
\[ E(y_{it1}|z_{it1}, y_{it2}, \bar{z}_i, v_{it2}) \] and then averaging out across 
(\bar{z}_i, \hat{v}_{it2}).