ARE 213 **Applied Econometrics**

UC Berkeley Department of Agricultural and Resource Economics

Selection On Unobservables Designs:
Part 7, Control Function Approaches

1 The Forbidden Regression

The "Forbidden Regression" was forbidden by Jerry Hausman back in the 1970s. If you take away one thing from this set of notes, it should be: Do not run the Forbidden Regression. The Forbidden Regression combines a non-linear first-stage in a 2SLS-style procedure. For example, you might estimate a probit first stage, take the fitted values from that ($\hat{d}_i$) first stage, and insert $\hat{d}_i$ as the regressor of interest in the second stage regression. The Forbidden Regression is tempting because it's simple and it may yield more precise estimates than 2SLS, but this precision is deceptive — effectively it is pulling identifying variation from the covariates rather than the instrument, which is not what you want to do. This is (potentially) a *much* more significant problem than, for example, the "incidental parameters problem" (i.e. including fixed effects in nonlinear panel data models) that we discussed earlier in the course. Bottom line: Do not mix non-linear models and 2SLS.[1]

2 Control Function Approach

An alternative method for leveraging instruments is the control function (CF) approach. The CF approach is sufficiently flexible to accommodate non-linear models in many cases. In most cases it comes down to throwing in first-stage residuals, possibly with interactions, as controls in your structural equation. But in some cases you might exploit additional functional-form

[1] It is, of course, okay to put in higher order terms of the treatment, or interactions between the treatment and covariates, but you then need to run a separate first stage for each of these transformations of the treatment, which means you need to generate additional instruments by creating higher order terms of the instrument or interacting the instrument with covariates. This invites issues with weak instruments.

Probit: $\hat{d} = \Phi(\hat{\gamma}_1 z + X\hat{\gamma}_2) \longrightarrow y = \beta\hat{d} + X\delta + \varepsilon$ don't do this!

Linear: $\hat{d} = \hat{\gamma}_1 z + \hat{\gamma}_2 x \longrightarrow y = \beta\hat{d} + X\delta + \varepsilon$

$= \beta(\hat{\gamma}_1 z + \hat{\gamma}_2 X) + X\delta + \varepsilon$

$= \beta\hat{\gamma}_1 z + (\beta\hat{\gamma}_2 + \delta)X + \varepsilon$

Control Function: $d = \gamma_1 z + X\gamma_2 + u$

$y = \beta_1 d + X\beta_2 + \beta_3\hat{u} + \varepsilon$

$\hat{u}$ = endogenous portion of d

$\hat{\beta}_3$ tells us if we needed to instr. in the first place
→ tells us if OLS vs CF β_1 are different

Probit CF: $y = \beta_1 d + X\beta_2 + \beta_3\hat{r} + \varepsilon$

$\hat{r}_i = d_i \, IMR(\hat{\gamma}_1 z_i + X_i\hat{\gamma}_2) + (1-d)\, IMR(-\hat{\gamma}_1 z_i - X_i\hat{\gamma}_2)$

$IMR(\cdot) = \frac{\phi(\cdot)}{\Phi(\cdot)}$

Endogenous FS:

$y = \beta_1 d + \beta_2(d-\bar{d})X + \beta_2 X + \varepsilon$

CF: $y = \beta_1 d + \beta_2(d-\bar{d})X + \beta_3 X + \beta_4\hat{u} + \varepsilon$

Random Coefficients: $y_i = \overset{ATE}{\bar{\beta}_1} d_i + \beta_{1i} d_i + X_i\beta_2 + \varepsilon_i$ can't est. β_{1i}

CF: $y_i = \beta_1 d_i + X_i\beta_2 + \beta_3\hat{u}_i + \beta_4\hat{u}_i d_i + \varepsilon$

assumptions (e.g., in the Heckman selection model we saw with the LaLonde paper near the beginning of the course).

The basic intuition is as follows. The IV or 2SLS estimators throw away the potentially endogenous variation in D (the treatment) and use only the variation induced by the instrument(s) to estimate causal effects. But, alternatively, you could run a regression of Y on D and include the endogenous part of D as a control, to get rid of the "bad" variation. The simplest example of the CF approach is to apply it to the conventional linear model that we would typically estimate via 2SLS. In this case, 2SLS and the CF approach generate identical estimates. You can apply the CF approach via the following two-stage procedure:

1. $d_i = \gamma_0 + \gamma_1 z_i + x_i\gamma_2 + u_i$
2. $y_i = \beta_0 + \beta_1 d_i + x_i\beta_2 + \beta_3\hat{u}_i + \varepsilon_i$

The estimate of β_1 is identical to what you'd get by running 2SLS and using z_i as an instrument for d_i. But there is one additional benefit: the test of $\beta_3 = 0$ is a test for the endogeneity of D, and it can easily accommodate heteroskedastic or clustered data (unlike a conventional Hausman test).

The real benefit of the CF approach, however, is that it opens up other possibilities with more complicated models. We discuss the following scenarios, drawn primarily from Wooldridge (2015), in lecture. As a brief reference:

1. Model with a probit first stage (e.g. Heckman/Lalonde)

   First stage: $E[D_i|X_i, Z_i] = \Phi(\gamma_0 + \gamma_1 d_i + x_i\gamma_2 + u_i)$

   Structural equation: $y_i = \beta_0 + \beta_1 d_i + x_i\beta_2 + \varepsilon_i$

   CF regression: $y_i = \beta_0 + \beta_1 d_i + x_i\beta_2 + \beta_3\hat{r}_i + \varepsilon_i$

   Where $\hat{r}_i = d_i \cdot IMR(\hat{\gamma}_0 + \hat{\gamma}_1 z_i + x_i\hat{\gamma}_2) + (1 - d_i) \cdot IMR(-\hat{\gamma}_0 - \hat{\gamma}_1 z_i - x_i\hat{\gamma}_2)$, and IMR is the inverse Mills Ratio, $\phi(.)/\Phi(.)$.

2. Models with a structural equation that incorporates interactions (or higher order terms) of the endogenous treatment

   First stage: $d_i = \gamma_0 + \gamma_1 z_i + \gamma_2 x_i + u_i$

   Structural equation: $y_i = \beta_0 + \beta_1 d_i + \beta_2 (d_i - \bar{d}) x_i + \beta_3 x_i + \varepsilon_i$

   CF regression: $y_i = \beta_0 + \beta_1 d_i + \beta_2 (d_i - \bar{d}) x + \beta_3 x_i + \beta_4 \hat{u}_i + \varepsilon_i$

   Note: If d_i is binary, in this case you may replace the continuous first stage with a probit first stage and substitute $\hat{r}_i$ for $\hat{u}_i$.

3. Models with a structural equation that incorporates a random coefficients model (heterogeneous treatment effect)

   First stage: $d_i = \gamma_0 + \gamma_1 z_i + \gamma_2 x_i + u_i$

   Structural equation: $y_i = \beta_0 + \bar{\beta}_1 d_i + \beta_{1i} d_i + x_i \beta_2 + \varepsilon_i$

   CF regression: $y_i = \beta_0 + \beta_1 d_i + x_i \beta_2 + \beta_3 \hat{u}_i + \beta_4 \hat{u}_i d_i + \varepsilon_i$

   A positive (negative) value of β_4 implies that those with an unobserved "taste" for treatment have higher (lower) treatment effects.

   Note: If d_i is binary, in this case you may replace the continuous first stage with a probit first stage and substitute $\hat{r}_i$ for $\hat{u}_i$.

4. Models with a structural equation that incorporates interactions (or higher order terms) of the endogenous treatment and a random coefficients model

   First stage: $d_i = \gamma_0 + \gamma_1 z_i + \gamma_2 x_i + u_i$

   Structural equation: $y_i = \beta_0 + \bar{\beta}_1 d_i + \beta_{1i} d_i + \bar{\beta}_2 x_i d_i + \beta_{2i} x_i d_i + \beta_3 x_i + \varepsilon_i$

   CF regression: $y_i = \beta_0 + \beta_1 d_i + \beta_2 x_i + \beta_3 d_i x_i + \beta_4 \hat{u}_i + \beta_5 \hat{u}_i d_i + \beta_6 \hat{u}_i d_i x_i + \varepsilon_i$

   Note: If d_i is binary, in this case you may replace the continuous first stage with a probit first stage and substitute $\hat{r}_i$ for $\hat{u}_i$.

5. Models with a parametric nonlinear structural equation

   First stage: $d_i = \gamma_0 + \gamma_1 z_i + x_i \gamma_2 + u_i$

Structural equation ("Rivers-Vuong" example): $y_i = \mathbf{1}(\beta_0 + \beta_1 d_i + x_i\beta_2 + \varepsilon_i > 0)$, with $\varepsilon_i \sim N(0, \sigma^2)$

CF regression: $E[Y_i|D_i, X_i, U_i] = \Phi(\beta_0 + \beta_1 d_i + x_i\beta_2 + \beta_3\hat{u}_i)$

The coefficient estimates are not correct for the original structural equation (they're rescaled), but the average partial effects (APE) are. Thus you can compute the fitted values, differentiate these with respect to d_i, and take the sample average of them.

6. Models with a nonparametric structural equation ("Blundell-Powell")

   First stage: $d_i = f(z_i, x_i) + u_i$, where $u_i \perp (z_i, x_i)$ (this rules out a probit and other limited dependent variable models)

   Structural equation: $y_i = g(d_i, x_i, \varepsilon_i)$

   Defining $h(x_i, d_i, u_i) = E[Y_i|X_i, D_i, U_i]$, then

   CF regression: Estimate $\hat{h}(x_i, d_i, \hat{u}_i)$, then compute APE

Finally, note that for most of the cases above you may need to bootstrap the standard errors.