Control Function Approach

- Modern control function approach is an extension of the IV method
- It relies on instruments: variables which predict treatment but are exogenous with respect to potential outcomes
- The exogenous variation induced by excluded instrumental variables provides separate variation in the residuals obtained from a reduced form,
- And these residuals serve as the control functions
- It goes beyond two stage least squares by making more explicit assumptions about the functional relationship between instruments, treatment and outcomes
- CF takes the selection model explicitly into consideration in the estimation process: Allows to test for endogeneity
- It also goes beyond the estimation of average effects to estimate the whole distribution of marginal treatment effects

Linear-in-Parameters Models: IV and CF

Models linear in endogenous variable

- Consider outcome variable y_1, endogenous regressor y_2 and a vector of exogenous variable Z
- Consider the model: $y_1 = z_1\delta_1 + \alpha_1 y_2 + u_1$ (1)
- where z_1 is a $1 \times L_1$ subvector of z with $E(z'u_1) = 0$ (Assumption 1)
- This is the same exogeneity condition we use for consistency of the 2SLS IV estimator
- reduced form of y_2: $y_2 = z\pi_2 + \nu_2$ and $E(z'\nu_2) = 0$
- Endogeneity of y_2 arises if and only if u_1 is correlated with ν_2
- linear projection of u_1 on ν_2: $u_1 = \rho_1\nu_2 + e_1$ (2)
- where $\rho_1 = E(\nu_2 u_1)/E(\nu_2^2)$ the population regression coefficient. Why?

- By definition, $E(\nu_2 e_1) = 0$ and $E(z'e_1 = 0)$ because u_1 and ν_2 are both uncorrelated with z
- Plugging (2) into (1): $y_1 = z_1\delta_1 + \alpha_1 y_2 + \rho_1\nu_2 + e_1$
- where we now view ν_2 as an explanatory variable in the equation
- e_1 is uncorrelated with ν_2 and z. Plus, y_2 is a linear function of z and ν_2, so e_1 is also uncorrelated with y_2
- Consistent estimates of parameters: run the OLS regression of y_1 on z_1, y_2, and ν_2 using a random sample
- However, we do not observe ν_2: it is the error in the reduced form equation for y_2
- $y_2 = z\pi_2 + \nu_2$ can be estimated by OLS
- so we can replace ν_2 by $\hat{\nu_2}$, the OLS residual from the first-stage regression

- $y_1 = z_1\delta_1 + \alpha_1 y_2 + \rho_1\hat{\nu_2} + error$, with $error_i = e_{i1} + \rho_1 z_i(\hat{\pi}_2 - \pi_2)$ which depends on sampling error in $\hat{\pi}_2$ unless $\rho_1 = 0$
- OLS estimates of previous equation gives consistent parameters
- This is a CF estimation: The inclusion of the residuals $\hat{\nu}_2$ "controls" for the endogeneity of y_2 in the original equation
- However, it does so with sampling error because $\hat{\pi}_2 \neq \pi_2$: SE must be adjusted for generated regressor bias
- CF estimates are identical in this case to 2SLS estimates of Equation (1) using z as IV
- We can test for endogeneity (Hausman test): $H_0 : \rho_1 = 0$ robust to heteroskedasticity

Non-linear Models in endogenous variable

- Let's extend the model: $y_1 = z_1\delta_1 + \alpha_1 y_2 + \gamma_1 y_2^2 + u_1$ (3)
- with $E(u_1|z) = 0$ (4)
- for simplicity, let's assume $\exists$ a scalar non-binary z_2 that is not in z_1
- Under Eq (4), we can use z_2^2 as IV for y_2^2: any function of z_2 is uncorrelated with u_1
- we can apply standard IV estimator with explanatory variables (z_1, y_2, y_2^2) and instruments (z_1, z_2, z_2^2)
- what would the CF approach entail in this case?

- To implement CF on Eq (3), we need $E(y_1|z, y_2)$
- A linear projection argument no longer works because of non-linearity in y_2
- We need an assumption on $E(u_1|z, y_2)$
- A standard one is: $E(u_1|z, y_2) = E(u_1|z, \nu_2) = E(u_1|\nu_2) = \rho_1 \nu_2 \qquad (5)$
- 1st equality follows because y_2 and ν_2 are 1:1 functions of each other given z
- second equality holds if (u_1, ν_2) is independent of z
- Final assumption is a linearity assumption which is more restrictive than simply defining a linear projection
- Under (5):

$$\begin{aligned} E(y_1|z, y_2) &= z_1\delta_1 + \alpha_1 y_2 + \gamma_1 y_2^2 + \rho_1(y_2 - z\pi_2) \\ &= z_1\delta_1 + \alpha_1 y_2 + \gamma_1 y_2^2 + \rho_1 \nu_2 \end{aligned}$$

- To implement CF approach, we run OLS of y_1 on $z_1, y_2, y_2^2, \hat{\nu}_2$, where $\hat{\nu}_2$ still represents the reduced form residuals
- CF estimates are not the same as the IV estimates for any choice of instruments for (y_2, y_2^2)
- The CF is likely to be more efficient than a direct IV, but less robust
- For instance, (4) and (5) $\implies E(y_2|z) = z\pi_2$
- This is a substantive restriction on the conditional distribution of y_2
- CF estimator will be inconsistent in cases where 2SLS estimator will be consistent
- However, because CF estimator solves the endogeneity of y_2 and y_2^2 by adding $\hat{\nu}_2$ to the regression, it will generally be more precise than IV estimator

Binary EEV: Formal Assumptions

Control Function Approach

- Standard IV treats all EEVs the same
- CF allows us to recognize the binary nature of some EEV
- Let's consider the heteregeneous treatment effect model:
  $y_i = d_i y_i^1 + (1 - d_i) y_i^0 = \beta + \alpha_i d_i + u_i$
- Assignment to treatment is given by the reduced form binary response:
  $P(d_i = 1) = Pr\big[g(Z_i, v_i) > 0\big] = P\big[Z_i \gamma + v_i > 0\big]$
- the CF approach is based on 2 assumptions:
  - **Assumption 1:** $(u, \alpha) \perp (d, Z) | v$
    Conditional on v, u and α are independent of d and Z
  - **Assumption 2:** $P\big[d = 1 | Z_{z-}, z\big] \neq P\big[d = 1 | Z_{z-}\big]$
    Conditional on the remaining regressors in Z (denoted by Z_{z-}), the treatment decision rule is a non-trivial (non-constant) function of z

Control Function Approach

- Under Assumption 1:

$$E[u|d, Z, v] = h_u(v) \tag{1}$$

$$E[\alpha|d, Z, v] = h_\alpha(v) \tag{2}$$

- If we knew these functions or could estimate them we could fully correct for selection on observables and recover the distribution of treatment effects
- Assumption 1 is often relaxed to **Assumption 1a**: $(u) \perp (d, Z)|v$
- which will be sufficient to recover the ATT
- Many applications of the control function approach typically make a parametric assumption on the joint distribution of the error terms, u and v

CF and treatment effect

- Consider

$$
\begin{aligned}
y_i &= \mu_0 + (\mu_1 - \mu_0)T_i + u_i \\
T_i &= g(\gamma_z Z_i + v_i)
\end{aligned}
$$

- 2 main assumptions: additivity of the error terms (most often made) and linearity in parameters in both the selection and the outcome equation
- Suppose z_i is a valid instrument: determines T_i but has no direct effect on outcome
- Assume that conditional on v_i, treatment is exogenous with respect to potential outcome: $(Y_i(1), Y_i(0)) \perp T|Z, v_i$
- one can consistently estimate average treatment effects provided one can control for $E(u_i|z_i, v_i)$: This is the control function
- It represent the effect of unobservables which both intervene in the selection process and determine potential outcomes: source of selection bias

- 2 step CF estimator
- First stage: estimate γ_z and predict v_i
- If treatment is continuous, and if g is invertible, then the v_i can be identified non parametrically
- In binary treatment case: need parametric assumptions
- $g(Z_i, v_i) = \mathbb{1}_{\{\gamma_z Z_i + v_i \geq 0\}}$ and we need to specify the distribution of v_i (often logistic or normal)
- 2nd stage: regresses outcomes on treatment and v
- The correct way to do it depends on assumptions regarding the selection process

Example 1: Heckman two step estimator (Heckit)

- Assumes linearity in both outcome and selection equation

$$\begin{aligned} y_i &= \mu_0 + (\mu_1 - \mu_0)T_i + u_i \\ T_i &= \mathbb{1}_{\{\gamma_z Z_i + v_i \geq 0\}} \end{aligned}$$

- with the additional assumption that disturbances in the selection and outcome equations are jointly normal, with covariance ρ $(u_i, v_i) \sim \mathcal{N}(0; (\sigma_u^2, \sigma_v^2, \rho))$

- Under normality of v_i, one can estimate the parameter γ_z through probit.

- Conditioning on v_i in the outcome equation yields

$$\begin{aligned} & E(Y_i|T_i, v_i) = \mu_0 + (\mu_1 - \mu_0)T_i + E(u_i|v_i) \\ = \quad & \mu_0 + (\mu_1 - \mu_0)T_i + T_i E(u_i| - \gamma z_i \geq v_i) + (1 - T_i)E(u_i| - \gamma z_i < v_i) \end{aligned}$$

Example1: Heckman two step estimator

Joint normality implies

$$\begin{aligned} E(u_i | -\gamma z_i \geq v_i) &= \rho E(v_i | -\gamma z_i \geq v_i) = \rho \lambda_1(-\gamma_z z_i) \\ E(u_i | -\gamma z_i < v_i) &= \rho \lambda_0(-\gamma_z z_i) \end{aligned}$$

- where $\lambda_1(-\gamma_z z_i) = \frac{\phi(-\gamma_z z_i)}{1-\Phi(-\gamma_z z_i)}$ and $\lambda_0(-\gamma_z z_i) = \frac{-\phi(-\gamma_z z_i)}{\Phi(-\gamma_z z_i)}$ are the inverse Mills ratios. The treatment effect can be consistently estimated by running OLS on
$y_i = \mu_0 + (\mu_1 - \mu_0)T_i + \rho\lambda_1(-\hat{\gamma}_z z_i)T_i + \rho\lambda_0(-\hat{\gamma}_z z_i)(1 - T_i) + \epsilon_i$

Example 2: Semi parametric methods

- Many methods have been proposed which relax the strong joint normality assumption of the disturbance terms
- Semi-parametric approach: make no functional assumptions about the selection process or the outcome function but simply assume additivity of the error term
- Non parametric estimation relies on power series (i.e. sum of polynomials) or splines (i.e. piece-wise polynomials)
- With discrete treatment, semi-parametric methods are in fact parametric in the first stage

Control Function Approach

- The CF method is close to a fully structural approach: it explicitly incorporates the decision process for the assignment rule in the estimation of the impact of the treatment
- The problem is how to identify the unobservable term, v, in order to include it in the outcome equation
- If T is a continuous variable and the decision rule is invertible, then T and Z are sufficient to identify v. In such case, v is a deterministic function of (T, Z)
- However, if T is discrete, and Z is continuous then we can still recover the complete distribution of treatment effects
- In this case the probability of $T = 1$ is a continuous function of z, say $P(z)$

CF and Correlated Random Coefficient Models

- Control function methods can be used for random coefficient models: models where unobserved heterogeneity interacts with endogenous explanatory variables
- Modify the outcome equation as $y_1 = \eta_1 + z_1\delta_1 + a_1 y_2 + u_1$
- where a_1 is the random coefficient on y_2
- write $a_1 = \alpha_1 + \nu_1$, whee $\alpha_1 = E(a_1)$ is the object of interest
- We can rewrite $y_1 = \eta_1 + z_1\delta_1 + \alpha_1 y_2 + \nu_1 y_2 + u_1 = \eta_1 + z_1\delta_1 + \alpha_1 y_2 + e_1$
- shows explicitly a constant coefficient on y_2 (which we hope to estimate) but also an interaction between the observed heterogeneity, ν_1, and y_2
- For a random draw, we would write: $y_{i1} = \eta_1 + z_{i1}\delta_1 + \alpha_1 y_{i2} + \nu_{i1} y_{i2} + u_{i1}$
- Makes it clear that δ_1 and α_1 are parameters to estimate and ν_{i1} is individual specific

- Assume $E(u_1|z,\nu_2) = \rho_1\nu_2$ and $E(\nu_1|z,\nu_2) = \zeta_1\nu_2$
- Then $E(y_1|z,y_2) = \eta_1 + z_1\delta_1 + \alpha_1 y_2 + \zeta_1\nu_2 y_2 + \rho_1\nu_2$
- This equation is estimable once we estimate π_2
- Garen's (1984) control function procedure is to first regress y_2 on z and obtain the reduced form residuals, $\hat{\nu}_2$, and then to run the OLS regression y_1 on $1, z_1, y_2, \hat{\nu}_2 y_2, \hat{\nu}_2$
- Under the assumptions above, Garen's method consistently estimates δ_1 and α_1
- standard errors should be adjusted for the estimation of π_2 in the first stage
- A test that y_2 is exogenous is easily obtained from the usual F test of : $H_0 : \zeta_1 = 0, \rho_1 = 0$

Next session: IV and MTE

- Heckman and Vytlacil (Econometrica, 2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation"
- Cornelissen et al (Labor Economics, 2016): "From LATE to MTE: Alternative methods for the evaluation of policy interventions"