

Endogeneity & Accounting

part II

Doug Schroeder
Ohio State University

Purdue University workshop
October 18-19, 2007

1

self-selection in a generalized (linear) Roy model:

$$\begin{aligned} \text{DGP: } y_j &= \mu_j(X) + v_j \quad j = 0, 1 && \text{(outcome equations)} \\ &= \mu_j + X\beta_j + v_j \\ U_D &= W\theta - V_D && \text{(selection equation)} \\ D &= 1 \quad U_D > 0 \\ D &= 0 \quad \text{otherwise} \\ y &= D y_1 + (1-D) y_0 && \text{(observable response)} \\ \Sigma &= \text{Var}[V_D, v_1, v_0|X, Z] \end{aligned}$$

3

some typical treatment effects:

$$\begin{aligned} \text{ATE} &= E_X[E[y_1 - y_0|X = x]] \\ \text{ATT} &= E_X[E[y_1 - y_0|X = x, D = 1]] \\ \text{ATUT} &= E_X[E[y_1 - y_0|X = x, D = 0]] \\ \text{LATE} &= E_X[E[y_1 - y_0|X = x, D_1 - D_0 = 1]] \\ &\quad \text{(for a binary instrument)} \\ \text{MTE} &= E[y_1 - y_0|X = x, V_D = v_D] \\ &= \text{LIV} = \partial E[y|X = x, P(z) = p] / \partial p|_{p=v} \end{aligned}$$

5

Outline – Identification & classical estimation of treatment effects

- I. Generalized Roy model and treatment effects
- II. Identification conditions
 - A. Ignorable treatment
 - B. IV methods
 - 1. Homogeneous response
 - 2. Heterogeneous response
- III. Next session
 - Bayesian analysis of treatment effects

2

self-selection in a generalized (linear) Roy model:

$$\begin{aligned} \text{net benefit or utility from treatment: } U_D &= W\theta - V_D \\ &= y_1 - y_0 - W\gamma - V_c \\ &= \mu_1(X) - \mu_0(X) - W\gamma \\ &\quad + v_1 - v_0 - V_c \\ \text{gross benefits of treatment: } &\mu_1(X) - \mu_0(X) \\ &= \mu_1 + X\beta_1 - (\mu_0 + X\beta_0) \\ \text{cost associated with treatment: } &W\gamma + V_D \\ \text{observable cost of treatment: } &W\gamma \\ \text{observable net benefit of treatment: } &\mu_1(X) - \mu_0(X) - W\gamma \\ \text{unobservable net benefit of treatment: } &-V_D = v_1 - v_0 - V_c \\ \text{observables: } &W = [X \quad Z] \end{aligned}$$

4

Marginal Treatment Effects

with homogeneous potential outcome gains, MTE is a constant

for heterogeneous potential outcome gains, MTE is typically a nonlinear function of v_D

6

Ignorable treatment & nonparametric identification

Assume $E[y_1 | X, D] = E[y_1 | X]$ and $E[y_0 | X, D] = E[y_0 | X]$,

conditional mean independence; also called “selection on observables.”

For binary treatment, this implies $E[y_1 | X, D=1] = E[y_1 | X, D=0]$ and

$E[y_0 | X, D=1] = E[y_0 | X, D=0]$.

Notice this is difficult to test as a direct test requires observing the counterfactuals, $E[y_1 | X, D=0]$ and $E[y_0 | X, D=1]$.

$ATE(X) = pE[y_1 | X, D=1] + (1-p)E[y_1 | X, D=0]$

$- pE[y_0 | X, D=1] - (1-p)E[y_0 | X, D=0] = E[y_1 | X] - E[y_0 | X]$

is *nonparametrically identified*.

7

Ignorable treatment & propensity score approach

The propensity score approach is an alternative to imposing functional form on conditional means of the unobservables.

Assume $E[y_1 | X, D] = E[y_1 | X]$ and $E[y_0 | X, D] = E[y_0 | X]$, conditional mean independence,

and $0 < p(X) = Pr(D=1 | X) < 1$ for all X .

Then

$ATE = E[(D - p(X))y / \{p(X)(1-p(X))\}]$

and

$ATT = E[(D - p(X))y / (1-p(X))] / Pr(D=1)$

Procedure: Estimate $p(X)$ via some flexible model (say, nonparametric regression) and ATE and ATT are consistently estimated via sample analogs to above.

9

ATT, ignorable treatment & propensity score

Assume $E[y_1 | X, D] = E[y_1 | X]$ and $E[y_0 | X, D] = E[y_0 | X]$, conditional mean independence,

and $0 < p(X) = Pr(D=1 | X) < 1$ for all X .

$ATT = E[(D - p(X))y / (1-p(X))] / Pr(D=1)$

Derivation: $E[(D - p(X))y | X] = p(X)[1-p(X)][m_1(X) - m_0(X)]$ (from ATE)

$E[p(X)(1-p(X))(m_1(X) - m_0(X)) / (1-p(X))] = E[p(X)(m_1(X) - m_0(X))]$

$= Pr(D=1)E[D(y_1 - y_0) | D=1] + Pr(D=0)E[D(y_1 - y_0) | D=0]$

$= Pr(D=1)E[D(y_1 - y_0) | D=1] + 0$

Hence, $E[(D - p(X))y / (1-p(X))] / Pr(D=1) = E[D(y_1 - y_0) | D=1] = ATT$

11

Ignorable treatment & exogenous dummy variable regression

Assume linear response and errors are normally distributed.

ATE is identified by *exogenous dummy variable regression*

$y = \delta + \zeta D + X\beta_0 + D X(\beta_1 - \beta_0) + \varepsilon$

where $ATE = \zeta + E[X](\beta_1 - \beta_0)$ and $\zeta = \mu_1 - \mu_0$.

The equation can be rewritten as

$y = \delta + (\zeta D + D E[X](\beta_1 - \beta_0)) + X\beta_0$
 $+ (D X(\beta_1 - \beta_0) - D E[X](\beta_1 - \beta_0)) + \varepsilon$

Hence, if we estimate

$y = \delta + \alpha D + X\beta_0 + D(X - E[X])(\beta_1 - \beta_0) + \varepsilon$

$\alpha = \zeta + E[X](\beta_1 - \beta_0) = ATE$.

8

ATE, ignorable treatment & propensity score

Assume $E[y_1 | X, D] = E[y_1 | X]$ and $E[y_0 | X, D] = E[y_0 | X]$, conditional mean independence,

and $0 < p(X) = Pr(D=1 | X) < 1$ for all X .

$ATE = E[(D - p(X))y / \{p(X)(1-p(X))\}]$

Derivation: $E[(D - p(X))y | X, D] = Dm_1(X) - p(X)(1-D)m_0(X) - p(X)Dm_1(X)$

where $m_f(X) = E[y_f | X]$

$E[Dm_1(X) - p(X)(1-D)m_0(X) - p(X)Dm_1(X) | X]$

$= p(X)m_1(X) - p(X)(1-p(X))m_0(X) - p^2(X)m_1(X)$

$= p(X)[1-p(X)][m_1(X) - m_0(X)]$

$E[(D - p(X))y / \{p(X)(1-p(X))\} | X] = m_1(X) - m_0(X)$

by iterated expectations,

$E[p(X)(1-p(X))(m_1(X) - m_0(X)) / p(X)(1-p(X))] = E[m_1(X) - m_0(X)] = ATE$.

10

Ignorable treatment & propensity score approach

If we add the assumption $E[y_0 | p(X)]$ and $E[y_1 | p(X)]$ are

linear in $p(X)$ then

$E[y | X, D] = \zeta_0 + \alpha D + \zeta_1 \hat{P} + D(\hat{P} - \hat{\mu}_p) \zeta_2$,

where $\hat{\mu}_p$ is the sample average of \hat{P} ,

then α consistently estimates ATE.

12

Matching estimators based on propensity score

Rosenbaum and Rubin, 1983, suggest selecting a propensity score at random from the sample,

then matching two individuals with this propensity score

(one who chose treatment and one who did not),

and constructing their outcome difference, $y_1 - y_0 | p(X)$.

Recognizes overlaps are important for treatment effect identification.

13

Matching estimators based on propensity score

Assume D and (y_1, y_0) are independent conditional on X (this is stronger than conditional mean independence).

The expected difference $E[y_1 - y_0 | p(X)]$ is ATE conditional on $p(X)$.

By iterated expectation, averaging over $p(X)$,

$$E_{p(X)}[E[y_1 - y_0 | p(X)]] = E[y_1 - y_0] = \text{ATE}.$$

A difficulty is finding matches. Heckman, Ichimura, and Todd, 1997, discuss trimming strategies in a nonparametric context and derive asymptotic standard errors.

14

Ignorable treatment & control function approaches

Assume $E[y_1 | X, D] = E[y_1 | X]$ and $E[y_0 | X, D] = E[y_0 | X]$, conditional mean independence, and $E[u_1 | X] = E[u_0 | X]$.

$$E[y | X, D] = \mu_0 + \alpha D + g_0(X)$$

where $\alpha = \text{ATE} = \text{ATT} = \text{ATUT}$, and $g_0(x) = E[u_0 | X]$.

If we add the assumption $E[u_0 | X] = \eta_0 + h_0(X)\beta_0$ for some vector (control) function $h_0(X)$, then $E[y | X, D] = \mu_0 + \eta_0 + \alpha D + h_0(X)\beta_0$.

That is, when the predicted individual-specific gain given X , $E[u_1 - u_0 | X]$, is zero we can estimate ATE by standard regression by adding a control function.

15

Ignorable treatment & control functions

If we relax $E[u_1 | X] = E[u_0 | X]$, then

$$E[y | X, D] = \mu_0 + \alpha D + g_0(x) + D[g_1(x) - g_0(x)]$$

where $\alpha = \text{ATE}$ (not necessarily equal to ATT) and

$$g_0(x) = E[u_0 | X] \text{ and } g_1(x) = E[u_1 | X]$$

The idea is that some flexible function $g(X)$ with enough controls in X has arranged it so that unobservables in (y_0, y_1) are unrelated to D .

16

Ignorable treatment & control functions

If we add the assumption $g_0(x) = \eta_0 + h_0(X)\beta_0$ and $g_1(x) = \eta_1 + h_1(X)\beta_1$, that is, g is linear in parameters then

$$E[y | X, D] = \phi + \alpha D + X\beta_0 + D(X - E[X])\delta$$

ATE(X) can be treated as a function of X , $\alpha + (X - E[X])\delta$.

(Averaged over all X this is α)

And, by same reasoning, ATT(X) can be estimated as

$$\alpha + \left(\sum_{i=1}^n D_i \right)^{-1} \left[\sum_{i=1}^n D_i (X_i - E[X]) \delta \right]$$

17

Instrumental variables approaches

Relax conditional mean independence and replace with existence of an instrument (exclusion restriction).

The exclusion restriction aids identification of the counterfactuals as an individual's probability of receiving treatment can be manipulated without affecting potential outcomes.

$$\text{DGP: } y = Dy_1 + (1-D)y_0 = \mu_0 + (\mu_1 - \mu_0)D + u_0 + (u_1 - u_0)D$$

$$U_D = Z\theta - V_D; D = 1 \text{ if } U > 0 \text{ and } D = 0 \text{ otherwise.}$$

18

Instrumental variables approaches

Homogeneous response:

$u_1 = u_0$ negates the interaction term $(u_1 - u_0)D$

$$y = \mu_0 + (\mu_1 - \mu_0)D + u_0$$

rules out individual-specific gains

where $u_j = X\beta_j + v_j$ (the stochastic portion of outcome)

ATE = ATT = ATUT

makes identification of treatment effects relatively

simple

19

Homogeneous response & IV endogenous dummy variable model

Assume (a) $u_1 = u_0$,

(b) $L(u_0 | X, Z) = L(u_0 | X)$, and

(c) $L(D | X, Z) \neq L(D | X)$,

where L is a linear projection (including unity ι).

Split y into constant and stochastic components

$$y = \mu_0 + (\mu_1 - \mu_0)D + u_0 + (u_1 - u_0)D,$$

apply (a) and (b)

$$u_0 = X\beta_0 + v_0 \text{ and}$$

since $u_1 = u_0$, $\beta_1 = \beta_0$, and $y = \mu_0 + (\mu_1 - \mu_0)D + X\beta_0 + v_0$

21

Propensity score IV

If make stronger assumptions can find more efficient IV estimator.

Assume (a) $u_1 = u_0$,

(b) $E[u_0 | X, Z] = L(u_0 | X)$, and

(c) $Pr(D=1 | X, Z) \neq Pr(D=1 | X)$ and $Pr(D=1 | X, Z) = G(X, Z; \gamma)$

is known parametric form (usually probit or logit),

and (d) $Var[u_0 | X, Z] = \sigma_0^2$.

23

Instrumental variables approaches

Heterogeneous response:

$u_1 \neq u_0$ retains the interaction term $(u_1 - u_0)D$

$$y = \mu_0 + (\mu_1 - \mu_0)D + u_0 + (u_1 - u_0)D$$

produces individual-specific gains

In general, ATE \neq ATT \neq ATUT

makes identification of treatment effects more

challenging

20

Endogenous dummy variable IV

Equation of interest is $y = \delta + \alpha D + X\beta_0 + v_0$,

where $\alpha = \text{ATE}$ and $v_0 = u_0 - L(u_0 | X, Z)$

As D and v_0 are typically correlated, OLS is inconsistent.

However, (b) means that Z is appropriately excluded from the above equation (this cannot be directly tested; may be indirectly tested via over-identifying restrictions).

Given assumptions, 2SLS (IV) estimation with $\{u, Z, X\}$ as instruments is consistent and asymptotically normal.

22

Propensity score IV

Utilize $\{u, G(X, Z; \gamma), X\}$ as instruments;

2SLS is consistent asymptotically normal (CAN)

where the equation of interest is $y = \delta + \alpha D + X\beta + v_0$.

This is more robust than OLS involving

$$y = \delta + \alpha G + X\beta + v_0$$

as the link function doesn't have to be equal to G for IV consistency but in general does for OLS.

24

Heterogeneous response

If $u_1 \neq u_0$ (heterogeneous response), the demands on the data are greater as endogeneity is more pressing.

$u_1 \neq u_0$ retains the interaction term $(u_1 - u_0)D$

produces individual-specific gains

ATE \neq ATT \neq ATUT

In general, when $u_1 \neq u_0$, the IV estimator (using Z or G as instruments for D) does not consistently estimate ATE (or ATT).

25

Heterogeneous response & propensity score IV

Suppose assumption (c) $e_1 = e_0$ above is replaced by

$$E[D(e_1 - e_0) | X, Z] = E[D(e_1 - e_0)].$$

The 2 stage propensity score IV procedure which utilizes $\{t, G, X, G(X - E[X])\}$ as instruments is consistent (though probably not efficient).

27

Ordinate control function IV approach

Estimate the equation

$$y = \gamma + \alpha D + X\beta + D(X - E[X]) + \xi\phi + \text{error}$$

by IV using instruments $\{t, \Phi, X, \Phi(X - E[X]), \text{and } \phi\}$

where Φ = cumulative standard normal distribution

and ϕ = ordinate for standard normal

each evaluated at $Z_i\theta$ from probit.

ATE is consistently estimated by α and

ϕ is a control function (obtained via IV assumptions).

29

Heterogeneous response & propensity score IV

Assume (a) conditional mean redundancy,

$$E[u_0 | X, Z] = E[u_0 | X] \text{ and } E[u_1 | X, Z] = E[u_1 | X],$$

$$(b) g_0(X) = \eta_0 + X\beta_0 \text{ and } g_1(X) - g_0(X) = (X - E[X])\delta,$$

$$(c) e_1 = e_0, \text{ where } u_0 = g_0(X) + e_0 \text{ and } E[e_0 | X, Z] = 0$$

$$\text{and } u_1 = g_1(X) + e_1 \text{ and } E[e_1 | X, Z] = 0, \text{ and}$$

$$(d) \Pr(D=1 | X, Z) \neq \Pr(D=1 | X) \text{ and } \Pr(D=1 | X, Z) = G(X, Z; \gamma)$$

where G is known parametric form (usually probit or logit).

26

Heterogeneous response & control function IV

Ordinate control function approach

Assume (a) conditional mean redundancy,

$$E[u_0 | X, Z] = E[u_0 | X] \text{ and } E[u_1 | X, Z] = E[u_1 | X],$$

$$(b) g_0(X) = \eta_0 + X\beta_0 \text{ and } g_1(X) - g_0(X) = (X - E[X])\delta,$$

$$(c) e_1 - e_0 \text{ is independent of } (X, Z),$$

$$\text{where } E[D | X, Z, e_1 - e_0] = h(X, Z) + k(e_1 - e_0)$$

for some functions h and k ,

$$u_0 = g_0(X) + e_0 \text{ and } E[e_0 | X, Z] = 0$$

$$\text{and } u_1 = g_1(X) + e_1 \text{ and } E[e_1 | X, Z] = 0,$$

$$(d) \Pr(D=1 | X, Z, e_1 - e_0) = \Phi(\pi_0 + X\pi_1 + Z\pi_2 + \varrho(e_1 - e_0)),$$

$$\pi_2 \neq 0, \text{ and}$$

$$(e) e_1 - e_0 \sim \text{Normal}(0, \tau^2).$$

28

Heckman's control function IV approach

Assume (a) conditional mean redundancy,

$$E[u_0 | X, Z] = E[u_0 | X] \text{ and } E[u_1 | X, Z] = E[u_1 | X],$$

$$(b) g_0(X) = \eta_0 + X\beta_0 \text{ and } g_1(X) - g_0(X) = (X - E[X])\delta,$$

$$u_0 = g_0(X) + e_0 \text{ and } E[e_0 | X, Z] = 0$$

$$\text{and } u_1 = g_1(X) + e_1 \text{ and } E[e_1 | X, Z] = 0,$$

$$(c) D = I[\theta_0 + X\theta_1 + Z\theta_2 + a \geq 0], \text{ where}$$

(a, e_0, e_1) is independent of (X, Z) with joint normal distribution

especially, $a \sim \text{Normal}(0, 1)$.

30

Heckman's control function IV approach

Procedure: (a) estimate θ 's via a probit of D on $\{t, X, Z\}$.
 (b) regress y on $\{t, D, X, D(X - E[X]), D(\phi/\Phi), (1-D)[- \phi/(1-\Phi)]\}$.

The coefficient on D is a consistent estimator of ATE.

Intuition: $E[y | X, D = j] = X\beta_j + E[e_j | D = j]$

Given assumptions, $E[e_j | D = j] \neq 0$ unless $\rho_{jv} = \text{correlation}(e_j, v) = 0$

If, say, $D = 1$ and $\rho_{1v} \neq 0$, then $E[e_1 | D = 1] = \rho_{1v}\sigma_v E[v | v > -Z\theta]$

The latter term is the expected value of a truncated normal where

$$E[v | v > -Z\theta] = \phi(-Z\theta)/[1-\Phi(-Z\theta)] = \phi(Z\theta)/\Phi(Z\theta).$$

Similar reasoning yields $E[u_0 | D = 0] = -\rho_{0v}\sigma_v \phi(Z\theta)/[1-\Phi(Z\theta)]$.

31

Estimating ATT by IV when $u_1 \neq u_0$

$$\text{ATT} = E[y_1 - y_0 | X, D = 1] = (\mu_1 - \mu_0) + E[u_1 - u_0 | X, D = 1]$$

Assume (a) $E[u_0 | X, Z] = E[u_0 | X]$,

(b) $E[u_1 - u_0 | X, Z, D = 1] = E[u_1 - u_0 | X, D = 1]$, and

(c) $\text{Pr}(D=1 | X, Z) \neq \text{Pr}(D=1 | X)$ and $\text{Pr}(D=1 | X, Z) = G(X, Z; \gamma)$

is known parametric form (usually probit or logit).

x

LATE & linear IV (Imbens and Angrist, 1994)

Let z be a binary instrument and write observed treatment status as

$$D = (1 - z)D_0 + zD_1 = D_0 + z(D_1 - D_0)$$

Then, $y = y_0 + D(y_1 - y_0)$ becomes

$$y = y_0 + D_0(y_1 - y_0) + z(D_1 - D_0)(y_1 - y_0).$$

Assume (a) z is independent of (y_0, y_1) ,

(b) $D_1 \geq D_0$, and

(c) $\text{Pr}(D = 1 | z = 1) \neq \text{Pr}(D = 1 | z = 0)$.

32

Discussion of IV control function approaches

The ordinate control function approach is probably more robust than Heckman's as it doesn't require normality.

Further, the ordinate approach allows us to separate endogeneity of D and heterogeneous response (where heterogeneity is indicated by $\xi \neq 0$).

However, if e_0 is independent of (X, Z) with a normal distribution,

Heckman's control function approach is likely more efficient (may be more useful, especially if the former procedure yields estimates that are too imprecise to be useful).

If X is not in the equation and Z is binary, the ordinate control function approach fails to identify ATE while it can be identified by Heckman's procedure.

x

Estimating ATT by IV when $u_1 \neq u_0$

Write $y = \mu_0 + g_0(x) + D[\mu_1 - \mu_0] + E[u_1 - u_0 | X, D = 1]$

$$+ D\{(u_1 - u_0) - E[u_1 - u_0 | X, D = 1]\} + e_0$$

$$= \mu_0 + g_0(x) + \text{ATT}(X) + a + e_0$$

Let $r = a + e_0$ and assume $g_0(x) = \eta_0 + h(X)\beta_0$ and

$\text{ATT} = \tau + f(X)\delta$ for functions $h(X)$ and $f(X)$, can write

$$y = \gamma_0 + h_0(X)\beta_0 + \tau D + [D f(X)]\delta + r \quad E[r | X, Z] = 0$$

Above equation can be estimated by IV

using any functions of (X, Z) as instruments.

Averaging $\tau + f(X)\delta$ over observations with $D = 1$

yields a consistent estimator for ATT.

x

LATE and linear IV

$$E[y | z = 1] = E[y_0] + E[D_0(y_1 - y_0)] + E[(D_1 - D_0)(y_1 - y_0)]$$

and

$$E[y | z = 0] = E[y_0] + E[D_0(y_1 - y_0)].$$

Now, $E[y | z = 1] - E[y | z = 0] = E[(D_1 - D_0)(y_1 - y_0)]$

$$= E[y_1 - y_0 | D_1 - D_0 = 1] \text{Pr}(D_1 - D_0 = 1)$$

$$- E[y_1 - y_0 | D_1 - D_0 = -1] \text{Pr}(D_1 - D_0 = -1)$$

Assumption (b; uniformity) implies $\text{Pr}(D_1 - D_0 = -1) = 0$ and

$$E[y | z = 1] - E[y | z = 0] = E[y_1 - y_0 | D_1 - D_0 = 1] \text{Pr}(D_1 - D_0 = 1).$$

LATE is defined as $E[y_1 - y_0 | D_1 - D_0 = 1]$ and

can be consistently estimated as α by linear IV in the simple equation

$$y = \delta_0 + \alpha D + \text{error} \text{ with binary instrument } z.$$

33

Discussion of LATE and linear IV

With binary treatment and binary instrument, the usual linear IV estimator consistently estimates LATE under weak assumptions.

LATE depends on the instrument selected (if we change the instrument LATE will likely change)

Because D_1 and D_0 are not both observable, we cannot identify the subpopulation with $D_1 - D_0 = 1$ (Angrist, Imbens, and Rubin refer to this subpopulation as the “compliers”).

Treatment effects literature is asymmetric as outcome can be heterogeneous but not treatment (requires uniformity)

34

Heckman and Vytlacil's MTE and heterogeneous response

MTE (marginal treatment effect)

MTE is the treatment effect (anticipated gains from treatment) for individuals who are indifferent between treatment and no treatment (conditional on X and V_D , unobserved utility of treatment)

$$MTE = E[y_1 - y_0 | X = x, V_D = v_D]$$

Treatment effect is identified under weak conditions similar to LATE, namely, existence of instruments (exclusion restriction satisfied) and monotonicity or uniformity – when Z changes from z to z' everyone shifts toward or away from treatment ($D = 1$).

Again the treatment effects literature is asymmetric, heterogeneous potential outcomes but homogeneous choice.

35

MTE & LIV (local instrumental variables) estimation

$$\begin{aligned} MTE &= E[y_1 - y_0 | X = x, V_D = v_D] \\ &= LIV = \partial E[y | X = x, P(Z) = p] / \partial p |_{p=v_D} \end{aligned}$$

36

MTE & LIV (cont.)

$E[y | X = x, P(Z) = p] = E[y_0 + (y_1 - y_0)D | X = x, Pr(Z) = p]$
by Bayes' theorem we have

$$E[y_0 | X = x] + E[y_1 - y_0 | X = x, D=1] Pr(D=1 | Z = z)$$

let U_D be distributed uniform(0, 1) yields

$$E[y_0 | X = x] + \int_0^p E[y_1 - y_0 | X = x, V_D = v_D] dv_D$$

the partial derivative with respect to p (evaluated at v_D) is
 $\partial E[y | X = x, P(Z) = p] / \partial p |_{p=v_D} = E[y_1 - y_0 | X = x, V_D = v_D]$
 $= MTE(x, v_D)$

37

Estimation of MTE by LIV

Earlier, we claimed LIV can estimate MTE

$$\partial E[y | X = x, Pr(Z) = p] / \partial p |_{p=v_D} = E[y_1 - y_0 | X = x, V_D = v_D]$$

For the linear separable model

$y_1 = \delta + \alpha + X\beta_1 + v_1$, and $y_0 = \delta + X\beta_0 + v_0$, yields

$E[y | X = x, Pr(Z) = z] = X\beta_0 + X(\beta_0 - \beta_1)Pr(Z) + \mathcal{K}(p)$ where

$$\mathcal{K}(p) = \alpha Pr(Z) + E[v_0 | Pr(Z) = p]$$

$$+ E[v_1 - v_0 | D = 1, Pr(Z) = p] Pr(Z)$$

LIV simplifies to $X(\beta_1 - \beta_0) + \partial \mathcal{K}(p) / \partial p |_{p=v_D}$

38

Estimation of MTE by LIV (cont.)

Since MTE is based on the partial derivative with respect to p of

$$E[y | X = x, Pr(Z) = z] = X\beta_0 + X(\beta_0 - \beta_1)Pr(Z) + \mathcal{K}(p)$$

The objective is to estimate $(\beta_1 - \beta_0)$ and the derivative of $\mathcal{K}(p)$.

LIV avoids strong distributional assumptions by employing nonparametric (local linear, kernel density) regression methods.

39

Estimation of MTE by LIV (cont.)

Step 1: Estimate $P(Z)$, the propensity score via probit, nonparametric discrete choice, etc.

Step 2: Estimate β_0 and $(\beta_1 - \beta_0)$ by using a nonparametric version of FWL (double residual regression). This involves a local linear regression (LLR) of each regressor in X and $X^*P(Z)$ onto $P(Z)$.

40

Estimation of MTE by LIV (cont.)

Step 2 (cont): For each regressor in X and $X^*P(Z)$ and for the response variable y estimate the residuals from LLR. Denote the matrix of residuals from the regressors (ordered by X followed by $X^*P(Z)$) as e_X and the residuals from y , e_y .

Step 3: Estimate $(\beta_0, \beta_1 - \beta_0)$ from a no-intercept linear regression of e_y onto e_X . That is, $[e_X^T e_X]^{-1} e_X^T e_y$.

42

Estimation of MTE by LIV (cont.)

Intuition: In the textbook linear model $y = X\beta + Z\gamma + \varepsilon$, FWL produces $E[y|X, Z] = P_Z y + (I - P_Z)Xb$ where b is the OLS estimator for β and P_Z is the projection matrix $Z(Z^T Z)^{-1} Z^T$.

Rewriting we can identify the estimator for γ (g)

$$E[y|X, Z] = Xb + P_Z(y - Xb) = Xb + Zg.$$

Hence, $g = (Z^T Z)^{-1} Z^T (y - Xb)$. That is, g is estimated from a regression of the restricted response $(y - Xb)$ onto the regressor Z . LIV employs the nonparametric analog.

44

Estimation of MTE by LIV (cont.)

LLR for X_k (the k th regressor) is

$$\{\tau_0(p), \tau_1(p)\} = \underset{\{\tau_0(p), \tau_1(p)\}}{\operatorname{arg\,min}} \left\{ \sum_{j=1}^n (X_k(j) - \tau_0 - \tau_1(P(Z_j) - p))^2 K\left(\frac{P(Z_j) - p}{h}\right) \right\}$$

where $K(W)$ is a (Gaussian, biweight, or Epanechnikov) kernel evaluated at W .

The bandwidth h is estimated by leave-one out generalized cross-validation based on the nonparametric regression of $X_k(j)$ onto $(\tau_0 + \tau_1 p)$.

41

Estimation of MTE by LIV (cont.)

Step 4: We've effectively estimated

$$\beta_0 X - (\beta_1 - \beta_0) X^* P(Z) \text{ for } E[y|X = x, P(Z) = z].$$

To estimate the derivative of $\mathcal{K}(p)$ by nonparametric FWL, define the restricted response

$$Y = y - b_0 X - (b_1 - b_0) X^* P(Z)$$

where b is the estimator for β .

43

Estimation of MTE by LIV (cont.)

Step 5: Estimate $\tau_1(p) = \partial \mathcal{K}(p) / \partial p$ by LLR of

$$Y_j = y_j - b_0 X_j - (b_1 - b_0) X_j^* P(Z_j) \text{ onto } P(Z_j) \text{ for each}$$

observation j in the set of overlaps

(the region for which MTE is identified – the subset of common support for $D = 1$ and $D = 0$).

45

Step 6: The LIV estimator of $MTE(X, v_D)$ is
 $(\beta_1 - \beta_0)X + \tau_1(p)$.

That is, conditional on X , MTE depends on the propensity score p .

In the homogeneous response setting, MTE is a constant and $MTE = ATE = ATT = ATUT$.

46

Combine MTE with weight functions to identify other treatment effects

HV connect MTE to numerous other treatment effects by utilizing weight functions h .

We'll explore a few of the standard ones.

In general, $TE = \int_0^1 MTE(x, v_D) h_{TE}(v_D) dv_D$

48

Combine MTE with weight functions to identify other treatment effects (cont.)

Unlike the weight functions for the other treatment effects, h for OLS doesn't integrate to one and can be negative (not a true weight function).

$$OLS: h_{OLS}(v_D) = \begin{cases} 1 + \frac{E[v_1|V_D = v_D]h_{TT}(v_D) - E[v_0|V_D = v_D]h_{TUT}(v_D)}{MTE(v_D)} & \text{for } MTE(x, v_D) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

50

In the heterogeneous response setting, MTE is typically a downward sloping function of p .

The intuition for this result is that for individuals who are less likely to accept treatment, a larger anticipated gain from treatment is necessary to induce selection of treatment.

While for individuals who are very likely to accept treatment, a smaller anticipated gain from treatment induces selection of treatment.

47

Combine MTE with weight functions to identify other treatment effects (cont.)

ATE: $h_{ATE}(v_D) = 1$ (assumes full support; more later)

$$ATT: h_{ATT}(v_D) = \left[\int_{u_D}^1 f(p|X=x) dp \right] \frac{1}{E[P|X=x]}$$

$$ATUT: h_{ATUT}(v_D) = \left[\int_0^{u_D} f(p|X=x) dp \right] \frac{1}{E[1-P|X=x]}$$

where f is the density function for Z .

49

Combine MTE with weight functions to identify other treatment effects (cont.)

$$LATE = \frac{1}{v_D - v_D'} \left[\int_{v_D'}^{v_D} MTE(v) dv \right]$$

In the limit as the interval gets arbitrarily small, LATE converges to MTE.

Since the treatment effects are only identified over the region of overlaps, often only some version of LATE (LATT, LATUT, etc.) is identified.

51

Concluding second session remarks

1. Causal effects lie at the center of many accounting, economics, and business questions
2. Endogeneity makes discovery of causal effects challenging
3. There exist a variety of identifying condition and approaches for estimating various treatment effects
 - existing methods seem underutilized
 - further progress calls for more innovation

What's next?

1. Bayesian analysis of selection including data augmentation and predictive distributions of treatment effects
2. Some estimation results for archival accounting choice data