Treatment evaluation

DEIP, Lima, Peru

March 1, 2017

References

- Greene, 7th ed., chap. 19, section 19.6
- Angrist, J. and J.-S. Pischke (2015), Mastering 'Metrics. Princeton University Press, Princeton.
- Angrist, J. and J.-S. Pischke (2009), Mostly Harmless Econometrics. Princeton University Press, Princeton.
- Blundel, R. and M. Costas Dias (2008), "Alternative approaches to evaluation in empirical microeconomics", Cemmep working paper CWP26/08.

Example

- Are R&D subsidies effective in stimulating firms to do more R&D?
- Collect data on firms with and without subsidies
- Construct treatment dummy variable
- Regress R&D on treatment dummy
- Problem(s)?

Omitted variable

- Other variables may influence amount of R&D or even choice of doing R&D
  - Scale
  - Technological opportunity
  - Cost of doing R&D
  - Uncertainty,...
- Try to include the main relevant control variables

Endogeneity of treatment

- Firms with and without R&D might be different from the outset.
- Government may channel funds to the most innovating firms

Missing counterfactual

- A firm cannot at the same time be in the treatment group and in the control group
- Firms in the two groups may be different for other reasons
- The same firm may be in the two situations but at different moments in time, but then other things may be different (new tax policy, recession,…)

Selection bias

- $Avg_n[Y_{1i}|D_i=1] – Avg_n[Y_{0i}|D_i=0]$
  $= \{Avg_n[Y_{1i}|D_i=1] – Avg_n[Y_{0i}|D_i=1]\} +$
  $\{Avg_n[Y_{0i}|D_i=1] – Avg_n[Y_{0i}|D_i=0]\}$

- Difference in group means
  = average causal effect + selection bias

- We do not observe $Avg_n[Y_{i0}|D_i=1]$!

How to handle the endogeneity?

- Matching estimator (treatment ignorable, matching on observables)
- Diff-in-diff: control for unobserved time-invariant heterogeneity
- Controlled diff-in-diff
- Instrumental variables (2SLS, LATE)
- Regression discontinuity design (sharp, fuzzy)
- Control function approach (also unobservables)
- Randomized controlled trials (endogeneity no issue)

Randomized trials

- Construct two randomly chosen groups, the treated and the untreated, i.e. they should otherwise have the same composition
- By the law of large numbers, the sample averages are consistent estimators of the population averages.
- The only difference between the two means is due to the treatment.

Randomized controlled trials (RCT)

- $E[Y_{0i}|D_i=1] = Avg_n[Y_{0i}|D_i=0]$
- Hence $Avg_n[Y_{1i}|D_i=1] – Avg_n[Y_{0i}|D_i=0]$

  $= \{Avg_n[Y_{1i}|D_i=1] – Avg_n[Y_{0i}|D_i=1]\} +$

  $\{Avg_n[Y_{0i}|D_i=1] – Avg_n[Y_{0i}|D_i=0]\}$

  $= \{Avg_n[Y_{1i}|D_i=1] – Avg_n[Y_{0i}|D_i=1]\}$

- Interesting to compare different RCT →external validity
- Check the balancing condition

Instrumental variables

- Example: does private school attendance increase student performance? Problem: attendance may be endogenous (family background). But there is also a lottery ticket for a subset of the seats offered. If some of those winning the lottery may because of that decide to go to the private school and otherwise would not have gone, then we can use those to compute a LATE (local average treatment effect).

LATE

- Y=outcome variable (e.g. math scores)
- D=treatment variable (school attendance)
- Z=instrumental variable (random school offer)

- LATE= $\frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]}$
- Causal story:

  effect of offers on scores

  = (effect of offers on attendance)x(effect of attendance on scores)
- LATE = effects of attendance on scores= effect of offers on scores/effect of offers on attendance

Interpretation of LATE

- LATE= $E[Y_{1i}-Y_{0i}|C_i=1]$ assuming no defiers

		Lottery losers Z=0	
		Not attending D=0	Attending D=1
Lottery winners Z=1	Not attending D=0	Never-takers	defiers
	Attending D=1	Compliers (C)	Always-takers

- We can only have a causal interpretation of the instrument for the compliers (hence <u>local</u> average treatment effect)

- External validity: other LATES for same or similar treatment

2SLS

- Generalization to multiple instruments with potentially many control variables

- $Y = X\beta + \varepsilon$

  Z

- 2 steps: Project X on Z and then replace X by the projection of X on Z

Example: omitted variables

- example: regress earnings on schooling knowing that ability plays also a role, but ability cannot be measured
- The true model: $Y_i = \alpha + \rho S_i + \beta A_i + \varepsilon_i$
- The regression of Y on S (omitting A) would yield $\rho + \beta\gamma$ as the coefficient of S where γ is the regression coefficient from a regression of A on S.
- Absence of a bias if β=0 or if γ=0.

Good instruments

- “Good instrument is correlated with the endogenous regressor for reasons the researcher can verify and explain, but uncorrelated with the outcome variable for reasons beyond its effect on the endogenous regressor.” (Angrist and Krueger, p. 73)

- If Z are weak instruments, i.e. not strongly correlated with X, or if the instruments are correlated with the outcome variable (i.e. with ε), IV can lead to a bias maybe even larger than OLS

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'(X\beta+\varepsilon) = \beta + (Z'X)^{-1}Z'\varepsilon \approx \beta + \mathrm{cov}(Z,\varepsilon)/\mathrm{cov}(Z,X)$$

Choice of instruments

- “In our view, good instruments often come from detailed knowledge of the economic mechanism and institutions determining the regressor of interest.” (Angrist and Krueger , p.73)

- Natural experiments (environment similar to randomized experiment)

Example: Angrist and Krueger, QJE, 1991

- Take the earnings-schooling example. Schooling is endogenous because omitted variable “ability” is correlated with both schooling and earnings.
- Natural experiment: differences in length of schooling because of different quarter of birth: those born in the quarter before Dec 31 enter at 5 ¾ years at school, those born in the quarter after Dec 31 enter with 6 ¼ years, and they all have to stay in school until they reach 16.

Intuition behind instrumental variables

- Hence those born before Dec 31, benefit from one more year of schooling.
- Use year of birth as instrument for schooling
- "Instrumental variables solve the omitted variable problem by using only part of the variability in schooling – specifically, a part that is uncorrelated with the omitted variables – to estimate the relationship between schooling and earnings." (Angrist and Krueger, p. 39)
- Of course, this instrument only works for students leaving school just after reaching 16. After that, the length of schooling becomes again endogenous.

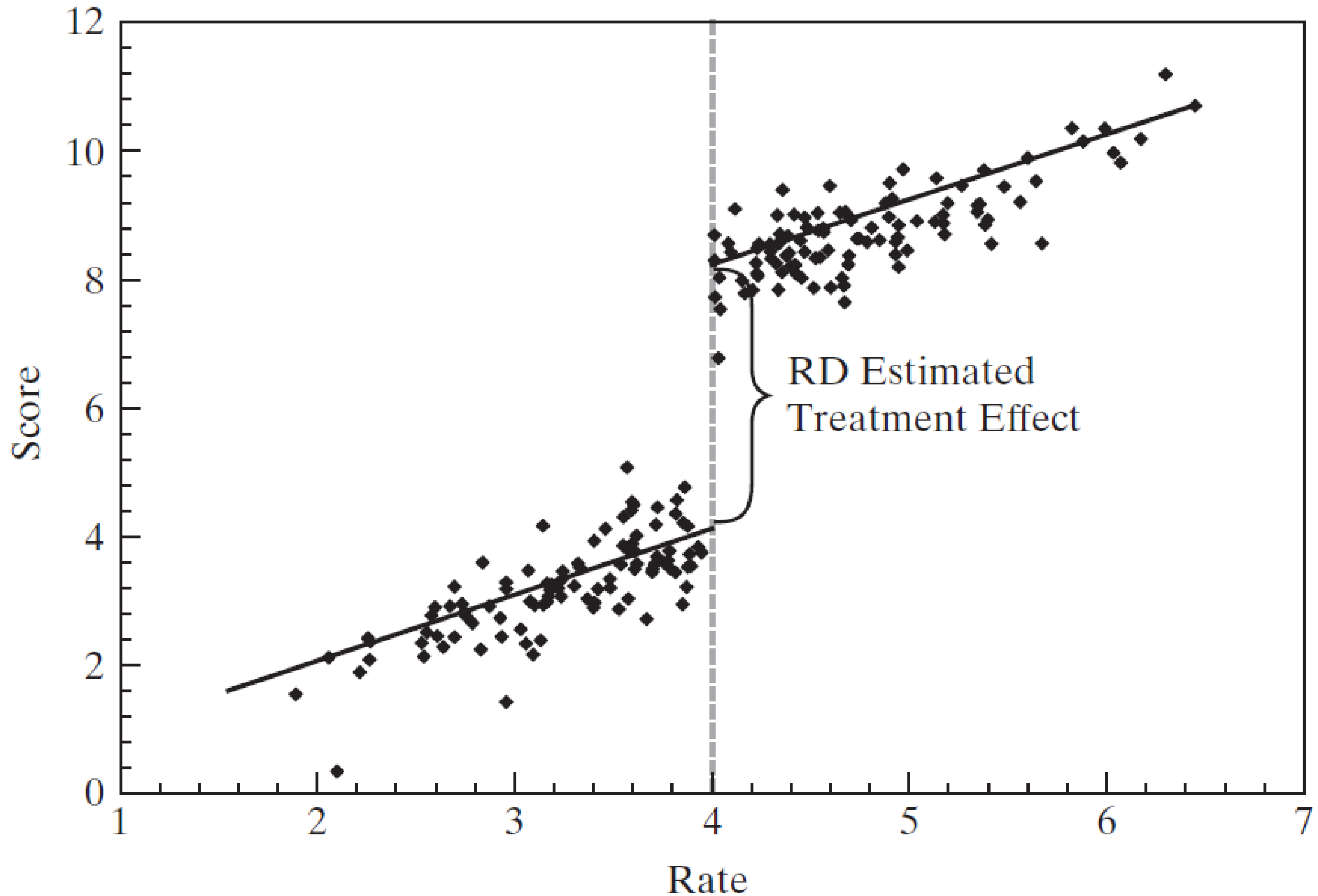
In practice

- Instruments must be valid (uncorrelated with error term, i.e. potentially omitted variables) and may not be weak (i.e. poorly correlated with the troublesome explanatory variables
- A valid instrumental variable may not affect the dependent variable through any other way than via the endogenous variable
- Underidentification is when there are fewer exclusion restrictions than troublesome variables
- Use of lagged endogenous variables is problematic if error term is autocorrelated

Regression discontinuity design (RD)

- Outcome variable is continuous function of a running variable. Treatment switches on or off as the running variable passes a cutoff.
- E.g. alcohol prohibition until age of 21. Mortality depends on age.

RD



12
10
8
6
4
2
0
Score
1
2
3
4
5
6
7
Rate
RD Estimated
Treatment Effect

RD

- Idea is to compare points close to the cut-off point to the right and to the left of it. Nearly randomized sample.
- Compare with different distances to the cut-off point.
- Estimate regressions like

$$Y = \alpha + \rho D a + \gamma a + e$$

with ρ capturing the RD effect.

RD

- The equation could also have a non-linear or curvilinear relation with the running variable, with no actual jump but an apparent jump when using linear specifications.
- Sharp and fuzzy discontinuities

Fuzzy RD

- With fuzzy discontinuity the intensity of treatment varies with the distance to the threshold
- Example: you may enter private high school if your entrance exam score is above a certain level. The average quality of peer students depends on test scores with a jump at entrance qualification level. The level 8 math scores depend on average quality of the peers.
- Application of IV

Differences-in differences

(1)

- Compare outcome before and after an exogenous shock for the treated and the controlled

$$Y_{it} = \alpha_0 + \gamma G + \tau T + \alpha G.T + \sum_j \beta^j X_{it}^j + \varepsilon_{it}$$

- Corrects for individual effects
- Allows to control for other co-determinants

Matching estimator

- Compare treated and untreated that are otherwise similar on the basis of observables
- For every treated observation find a matched untreated
  - Nearest neighbor
  - Caliper
  - Kernel weighted average of all untreated
  - With or without replacement

Propensity score matching

- When the number of observables one wants to control for is too big, matching cells become thinly populated
- Propensity score matching: first estimate a probit or logit and then do the matching on the propensity scores

Underlying conditions for matching

- Conditional mean independence

$$E[y_{i0} \mid \mathbf{x}_i, C_i = 1] = E[y_{i0} \mid \mathbf{x}_i, C_i = 0] = E[y_{i0} \mid \mathbf{x}].$$

- Overlap assumption

For any value of $\mathbf{x}$, $0 < \mathrm{Prob}(C_i = 1 \mid \mathbf{x}) < 1$

Pros and Cons of matching estimators :

👍 Well suited for cross-sectional data

👍 No assumption on functional form or distribution

👎 Only controlling for observed heterogeneity among treated and non treated firms

Example of propensity score matching

- Berube, Charles and Pierre Mohnen, “Are firms that received R&D subsidies more innovative?”, *Canadian Journal of Economics*, 42(1), 206-225, 2009

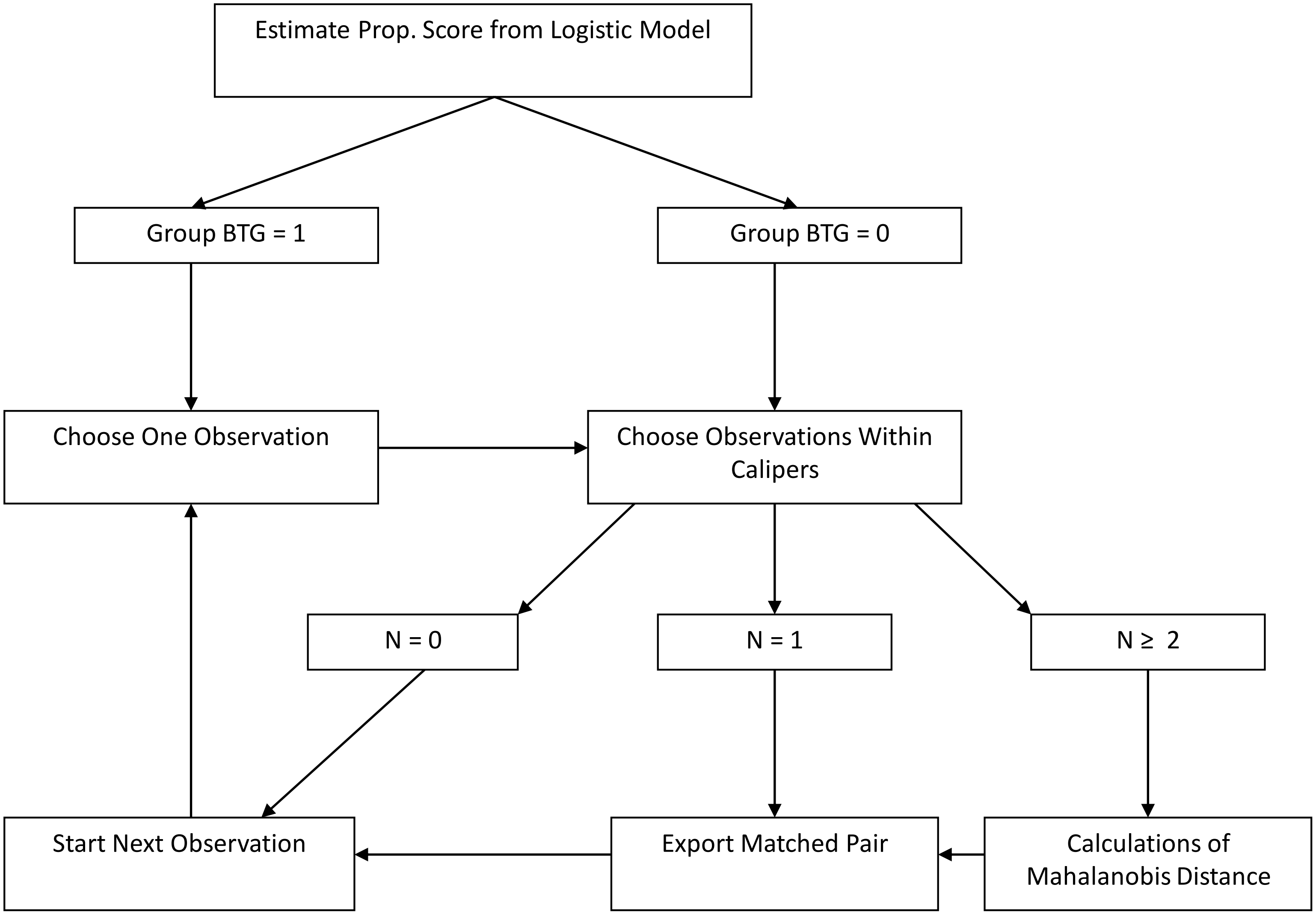
Impact of grants on innovation

- We compare firms that received tax credits only (untreated) with those that received tax credits and grants (treated)
- Outcome: different degrees of innovation:
    - Province first, Canadian first, North-American first and World first innovations
    - Already-on-the-market and first-to-market product
    - Number of new products

DATA: SURVEY OF INNOVATION 2005

- Confidential micro-data of 6,143 completed questionnaires
- Sample: 2,785 manufacturing plants*
- Of which:
  - 2,200 used tax credits only
  - 585 used tax credits and grants program
- Rejected plants: 3,292 did not use any programs and 66 used grants only

MATCHING PROTOCOL (1)



MATCHING PROTOCOL (2)

- To allow matching with replacement introduces a bias in the ordinary t-statistic for testing mean differences.

- The results presented in this paper are from a matching process that does not allow different treated firms to be matched to the same non-treated firm.

- Mahalanobis distance was used to get a unique match. Remaining treated firms that were matched to the same non-treated firm after a full cycle have to be rematched until all treated firms are uniquely match.

Table 1*: Mean and proportions of relevant characteristics before matching

Characteristics	Tax Credits only (N=2200)	Tax credits + R&D grants (N=585)	P-Value
Lnemp	4.3346	4.2499	<0.0001
Mean of predicted probabilities	0.1786	0.2317	<0.0001
Atlantic Region	3.63 %	5.70 %	0.0298
Quebec Region	36.37 %	42.11 %	0.0145
Ontario Region	43.89 %	38.68 %	0.0038
Western Region	16.10 %	18.51 %	0.1806
Resources Ind.	24.50 %	23.78 %	0.7315
Labour Ind.	24.15 %	27.75 %	0.0851
Scale Ind.	24.45 %	19.44 %	0.0148
Specialized Ind.	18.69 %	15.04 %	0.0502
Science Ind.	7.52 %	13.13 %	<0.0001
Niche	37.06 %	44.89 %	0.0009
New ind. Standards	10.63 %	17.10 %	<0.0001
Environment	34.10 %	37.22 %	0.1761
Applied for patents	21.39 %	31.93 %	<0.0001
Outsourcing R&D	20.65 %	31.26 %	<0.0001
External funding	48.45 %	61.46 %	<0.0001

*Statistics Canada, Survey of Innovation 2005, authors' calculations.

Table 2*: Proportions of relevant outcome measures before matching

Outcome variables	Tax Credits only (N=2200)	Tax credits + R&D grants (N=585)	P-Value
Province First	52.89%	64.80%	<0.0001
Canadian First	41.41%	53.01%	<0.0001
North A. First	27.24%	38.24%	<0.0001
World First	13.24%	25.26%	<0.0001
New innovation > 0	70.11%	80.49%	<0.0001
New innovation > 2	52.95%	64.75%	<0.0001
% Rev. First-to-market > 0	48.86%	60.84%	<0.0001
% Rev. Already-on-market > 0	40.71%	43.96%	0.1727

*Statistics Canada, Survey of Innovation 2005, authors' calculations.

Table 3*: Logit Model on the R&D tax credits and grants dummy

Variables	Estimate	P-value
Intercept	-2.3283	<0.0001
Ln Employment	-0.097	0.0792
Atlantic Region	0.7841	0.0009
Quebec Region	0.4266	0.0003
Ontario Region	*Reference	*Ref.
Western Region	0.4981	0.0007
Resources Ind.	0.2645	0.0851
Labour Ind.	0.3345	0.0221
Scale Ind.	*Reference	*Ref.
Specialized Ind.	0.0398	0.8137
Science Ind.	0.6962	0.0002
Niche	0.2447	0.0168
New ind. Standards	0.5257	0.0003
Enviro.	0.2161	0.0499
Applied for patents	0.5313	<0.0001
Outsourcing R&D	0.4336	0.0001
External funding	0.4778	<0.0001

*Statistics Canada, Survey of Innovation 2005, authors' calculations.

Table 4*: Mean and proportions of relevant characteristics after matching

Characteristics	Tax Credits only (N=584)	Tax credits + R&D grants (N=584)	P-Value
Lnemp	4.2691	4.2503	0.7370
Mean of predicted probabilities	0.2288	0.2311	0.7000
Atlantic Region	5.48 %	5.71 %	0.8730
Quebec Region	44.86 %	42.16 %	0.3849
Ontario Region	33.70 %	33.72 %	0.9928
Western Region	15.97 %	18.41 %	0.3019
Resources Ind.	23.61 %	23.81 %	0.9401
Labour Ind.	25.41 %	28.63 %	0.2468
Scale Ind.	20.98 %	19.47 %	0.5465
Specialized Ind.	16.56 %	15.06 %	0.5104
Science Ind.	13.43 %	13.02 %	0.8492
Niche	45.28 %	44.82 %	0.8828
New Ind. Standards	16.80 %	16.99 %	0.9348
Environment	38.10 %	37.15 %	0.7525
Applied for patents	32.45 %	31.84 %	0.8343
Outsourcing R&D	30.19 %	31.18 %	0.7314
External funding	61.18 %	61.42 %	0.9371

*Statistics Canada, Survey of Innovation 2005, authors' calculations.

Table 5*: Proportions of relevant outcome measures after matching

Outcome variables	Tax Credits only (N=584)	Tax credits + R&D grants (N=584)	P-Value
Province first	58.84%	64.75%	0.0819
Canadian first	47.84%	52.96%	0.1068
North A. First	31.4%	38.17%	0.0598
World first	17.24%	25.29%	0.0046
New innovation > 0	71.80%	80.47%	0.0011
New innovation > 2	50.86%	64.70%	<0.0001
% Rev. First-to-market > 0	52.49%	60.79%	0.0074
% Rev. Already-on-market > 0	40.13%	44.02%	0.2086

*Statistics Canada, Survey of Innovation 2005, authors' calculations.

Control function approach

- Endogenous selection or control on unobservables

$$\text{earnings}_i = \mathbf{x}_i'\boldsymbol{\beta} + \delta C_i + \varepsilon_i$$

$$C_i^* = \mathbf{w}_i'\boldsymbol{\gamma} + u_i,$$

$$C_i = 1 \quad \text{if } C_i^* > 0, 0 \text{ otherwise}$$

$$\begin{aligned} E[y_i \mid C_i = 1, \mathbf{x}_i, \mathbf{w}_i] &= \mathbf{x}_i'\boldsymbol{\beta} + \delta + E[\varepsilon_i \mid C_i = 1, \mathbf{x}_i, \mathbf{w}_i] \\ &= \mathbf{x}_i'\boldsymbol{\beta} + \delta + \rho\sigma_\varepsilon \lambda(-\mathbf{w}_i'\boldsymbol{\gamma}) \end{aligned}$$

$$E[y_i \mid C_i = 0, \mathbf{x}_i, \mathbf{w}_i] = \mathbf{x}_i'\boldsymbol{\beta} + \rho\sigma_\varepsilon \left[\frac{-\phi(\mathbf{w}_i'\boldsymbol{\gamma})}{1 - \boldsymbol{\Phi}(\mathbf{w}_i'\boldsymbol{\gamma})}\right]$$

$$E[y_i \mid C_i = 1, \mathbf{x}_i, \mathbf{w}_i] - E[y_i \mid C_i = 0, \mathbf{x}_i, \mathbf{w}_i] = \delta + \rho\sigma_\varepsilon \left[\frac{\phi_i}{\boldsymbol{\Phi}_i(1 - \boldsymbol{\Phi}_i)}\right]$$

If different outcome equations

$$C_i^* = \mathbf{w}_i'\boldsymbol{\gamma} + u_i,\, C_i = \mathbf{1} \text{ if } C_i^* > 0 \text{ and } 0 \text{ otherwise,}$$

$$y_{i0} = \mathbf{x}_i'\boldsymbol{\beta}_0 + \varepsilon_{i0},$$

$$y_{i1} = \mathbf{x}_i'\boldsymbol{\beta}_1 + \varepsilon_{i1}.$$

$$\begin{pmatrix} u_i \\ \varepsilon_{i0} \\ \varepsilon_{i1} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_0\theta_0 & \rho_1\theta_1 \\ \rho_0\theta_0 & \theta_0^2 & \theta_{01} \\ \rho_1\theta_1 & \theta_{01} & \theta_1^2 \end{pmatrix}\right].$$

$$\begin{aligned} \text{ATET} &= E[y_{i1} \mid C_i = 1,\, \mathbf{x}_i,\, \mathbf{w}_i] - E[y_{i0} \mid C_i = 1,\, \mathbf{x}_i,\, \mathbf{w}_i] \\ &= \mathbf{x}_i'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0) + (\rho_1\theta_1 - \rho_0\theta_0)\frac{\phi\left(\mathbf{w}_i'\boldsymbol{\gamma}\right)}{\Phi\left(\mathbf{w}_i'\boldsymbol{\gamma}\right)}. \end{aligned}$$