Counterfactual approach 3

Prof. Roberto Gabriele

Methods for Empirical Economics

Department of Economics and Management, University of Trento,
Ph.D. programme in Economics and Finance

Outline

1 IV

Hidden bias

1. When **nonobservable factors** significantly drive the nonrandom assignment to treatment consistent estimations of ATEs relying only on observables (basically, the vector of covariates x) is no longer possible
2. Regardless of what kind of unobservableness the analyst has to deal with, the problem becomes: **finding suitable econometric procedures in order to produce consistent estimation of ATEs**
3. The literature has provided three methods to cope with selection on unobservables:
4. Instrumental-variables (IV), Selection-models (SM), and Difference-in-differences (DID)

More information

1. Their implementation requires, however, either **additional information or further assumptions**, which are not always available or viable
2. working under the potential presence of a hidden bias is generally recognized as much more tricky than working under overt bias

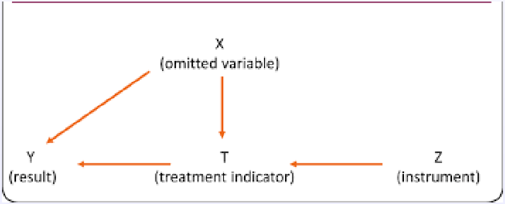
IV

1. When selection into a program is driven not only by observables but also by unobservable-to-the-analyst factors, then the **conditional mean independence (CMI) hypothesis no longer holds**
2. In the regression approach, the **treatment binary variable D becomes endogenous**, that is, correlated with the error term, thus preventing ordinary least squares (OLS) from producing consistent estimates of regression parameters, including ATE, ATET, and ATENT.
3. In the case of Matching (and propensity-score based Reweighting, for instance), **the bias depends on excluding relevant covariates from the variables generating the actual propensity-score** and/or from the matching procedure applied on units

Definition of IV

1. The need for **at least one exogenous variable z, the instrumental-variable**, which is assumed to have the following two fundamental properties:
   - z is (directly) correlated with treatment D
   - z is (directly) uncorrelated with outcome Y
2. These two requirements imply that **the selection into program should possibly depend on the same factors affecting the outcome plus z**, the instrument, assumed to not directly affect the outcome

IV representation



Finding instruments

1. **Sources of instruments** come from a combination of institutional knowledge and ideas about the processes determining the variable of interest
2. **Institutional constraints** may play a key role in generating suitable instruments, thus providing grounds for creating quasi-randomized settings approximating natural experiments

Finding instruments II

A classical example

1. For instance, Angrist and Krueger (1991) looking for the causal relation between years of schooling and personal earning
2. the authors use "quarter-of-birth" in order to instrument years of education, assumed to be endogenous
3. Why should this be a good instrument?
4. The authors argue that as it is compulsory to attend school until the age of 16 in many US states (and only after this threshold can a student freely drop out of school), and as individuals born in the first quarters of the year start school before the age of 6, while ones later born are more than 6 years old at that time, this induces a situation in which earlier born children have a longer education time than those born later.

Finding instruments III

1. Empirically, the authors find a positive relation between years of education and quarter-of-birth,
2. thus showing that this variable can serve as a good instrument for years of education.
3. In fact, the date of birth seems unrelated to the (unobservable) variables which may influence earnings such as family background, personal motivation, and genetic attitude;
4. as such, quarter-of-birth can be reliably assumed as randomly determined and, as such, purely exogenous

A recall: properties of covariance

$$\mathrm{Cov}[X,Y] \stackrel{\text{def}}{=} \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$$

Properties of Covariance:

1. $\mathrm{Cov}[X,c] = 0$ for any constant c.
2. $\mathrm{Cov}[aX,Y] = a \cdot \mathrm{Cov}[X,Y]$
   $\mathrm{Cov}[X,aY] = a \cdot \mathrm{Cov}[X,Y]$
3. $\mathrm{Cov}[X,Y] = \mathrm{Cov}[Y,X]$
4. $\mathrm{Cov}[X,X] = \mathrm{Var}[X]$
5. Bilinearity (a.k.a. *distributive property*):

$$\mathrm{Cov}[X+Y,Z] = \mathrm{Cov}[X,Z] + \mathrm{Cov}[Y,Z]$$
$$\mathrm{Cov}[X,Y+Z] = \mathrm{Cov}[X,Y] + \mathrm{Cov}[X,Z]$$

UNIVERSITÀ DI TRENTO

A formal introduction

Solving hidden bias

1. Lets start from the linear univariate model:
2. $Y = \mu + \alpha D + u$; so that: $\alpha = E[Y|D=1] - E[Y|D=0] = DIM$
3. $\alpha = cov(Y, D)/Var(D)$ (OLS)
4. if selection-into-treatment was driven by a factor x, that is unobservable-to-the-analyst:
5. $Y = \mu + \alpha D + \beta x + u = \mu + \alpha D + u*$
6. $\alpha = cov(Y, D)/Var(D) = cov(\mu + \alpha D + \beta x + u, D)/var(D) =$
7. $= \alpha Var(D)/Var(D) + \beta cov(X, D)/Var(D)$
8. OLS: $\alpha_{OLS} = \alpha + \beta cov(X, D)/Var(D)$

in the case of unobservable selection, a standard OLS is a biased estimator given that $Cov(D; u*) \neq 0$

A formal introduction II

Solving hidden bias

1. We get then:
2. $\alpha_{OLS} = \alpha + \beta E[Y|D = 1] - E[Y|D = 0] = DIM$
3. It proves that in the case of unobservable selection, a standard OLS is a biased estimator.

The solution provided by IV

1. IV approach can restore consistency, provided that an **instrumental-variable z, correlated with D but uncorrelated with u***, is available
2. If we assume that u is a pure random component, thus uncorrelated by definition with z:
3. $Cov(z, u*) = Cov(z, \beta x + u) = Cov(z, \beta x) + Cov(z, u) = \beta Cov(z, x) = 0$
4. Hence for IV estimator we have: $Cov(z, u*) = Cov(z, y - \mu - \alpha D) = Cov(z, y) - \alpha Cov(z, D) = 0$
5. $\alpha_{IV} = Cov(z, y)/Cov(z, D)$
6. which is consistent: $\alpha_{IV} = Cov(z, \alpha D + \mu + \beta x + u)/Cov(z, D) = \beta Cov(z, x) + \alpha Cov(z, D)/Cov(z, D) = \alpha$

UNIVERSITÀ DI TRENTO

A more general approach to estimate ATEs

1. Starting with POM model:
2. $Y = \mu_0 + D(\mu_1 - \mu_0) + \nu_0 + D(\nu_1 - \nu_0)$*
3. This equation, assuming that CMI does not hold, yields:
4. $E[\nu_1|D, x] \neq E[\nu_1|x]$
5. $E[\nu_0|D, x] \neq E[\nu_0|x]$

two cases: homoegenous and heterogeneous case

*From $Y = Y_1 D - Y_0(1 - D)$

Homogeneous case

$\nu_1 = \nu_0$

We get:

1. $Y = \mu_0 + D(\mu_1 - \mu_0) + \nu_0$
2. and also: $ATET = ATE = ATENT = \mu_1 - \mu_0$
3. Suppose, however, one has an instrumental-variable z. Formally:
4. $$\mathrm{E}(v_0|\mathbf{x},z) = \mathrm{E}(v_0|\mathbf{x}) \quad \Leftrightarrow \quad z \text{ is uncorrelated with } v_0$$
$$\mathrm{E}(D|\mathbf{x},z) \neq \mathrm{E}(D|\mathbf{x}) \quad \Leftrightarrow \quad z \text{ is correlated with } D$$
5. assume: $E[\nu_0|z,x] = E[\nu_0|x] = g(x) = x\beta$
6. so that: $Y = \mu_0 + D \cdot ATE + x\beta + u$

The full model

The previous assumptions and relationships can be more compactly summarized in the following two-equation structural system:

$$\begin{cases} \text{(a)} & Y_i = \mu_0 + D_i \text{ATE} + \mathbf{x}_i \boldsymbol{\beta} + u_i \\ \text{(b)} & D_i^* = \eta + \mathbf{q}_i \boldsymbol{\delta} + \varepsilon_i \\ \text{(c)} & D_i = \begin{cases} 1 & \text{if} \quad D_i^* \geq 0 \\ 0 & \text{if} \quad D_i^* < 0 \end{cases} \\ \text{(d)} & \mathbf{q}_i = (\mathbf{x}_i, z_i) \end{cases}$$

The full model II

Eq. (b) represents the **latent selection function** derived from:

1. An objective function of a supporting external agency choosing whether a unit is, or is not, suitable for treatment;
2. Self-selection into the program operated by units themselves, according to some cost/benefit contrast within a proper unit pay-offs function
3. Generally, it is assumed that D_i*, a rescaled scalar score associated with each eligible unit, is **unknown to the evaluator** as he only knows the (final) binary decision indicator D_i (selected vs. not selected for the program), along with some other observable unit characteristics (covariates) affecting this choice.

The full model III

In a system like this, **endogeneity arises when one assumes that the unobservable factors affecting the selection into program** (i.e., ε_i) are correlated with the unobservable factors affecting the realization of units' outcome (i.e., u_i).

In the case of zero correlation between these two terms, OLS of (a) produces consistent estimation of ATE.

Direct Two-Stage Least Squares

It is based on **two sequential OLS regressions** in order to calculate the predictions of the endogenous variable D in the first step, and on using these predictions as a regressor in the outcome equation in place of the actual D in the second step

1. Estimate the selection equation by running an OLS regression of D on x and z of the type: $D = \eta + x\delta_x + z\delta_z + error$ to obtain the predicted values of D_i, denoted by D_{fv}
2. Estimate the outcome equation by running a second OLS of Y on x and D_{fv}. the coefficient is a **consistent estimation of ATE**

Note: in step 1, what is fitted is a linear probability model, while in step 2, a standard OLS regression is estimated. The second step also provides the analytical estimation of ATE and of its standard error to perform usual significance tests.

Probit-2SLS

It returns consistent estimations even when the first-step probit is incorrectly specified.

1. Estimate a probit of D on x and z, getting p_{1Di}, i.e., the "predicted probability of D."
2. Run an OLS of D on (1, x, p_{1Di}), thus getting the fitted values $D_{2fv,i}$.
3. Finally, estimate a second OLS of Y on (1, x, $D_{2fv,i}$).

Heterogeneous case

$\nu_1 \neq \nu_0$

We assume:

$$E(v_0 | \mathbf{x}, z) = E(v_0 | \mathbf{x}) = g_0(\mathbf{x})$$
$$E(v_1 | \mathbf{x}, z) = E(v_1 | \mathbf{x}) = g_1(\mathbf{x})$$

that is:

$$v_0 = g_0(\mathbf{x}) + e_0 \quad \text{with} \quad E(e_0 | \mathbf{x}, z) = 0$$
$$v_1 = g_1(\mathbf{x}) + e_1 \quad \text{with} \quad E(e_1 | \mathbf{x}, z) = 0$$

From POM for Y, we obtain that:

$$Y = \mu_0 + \alpha D + g_0(\mathbf{x}) + D[g_1(\mathbf{x}) - g_0(\mathbf{x})] + e_0 + D(e_1 - e_0)$$

$$Y = \mu_0 + \text{ATE} \cdot D + \mathbf{x}\boldsymbol{\beta}_0 + D(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})\boldsymbol{\beta} + \varepsilon$$

Heterogeneous case II

Assuming also:

1. $g_1(x) = x\beta_1$, $g_0(x) = x\beta_0$
2. $\varepsilon = e_0 + D(e_1 - e_0)$

we get:

$$Y = \mu_0 + \mathrm{ATE} \cdot D + \mathbf{x}\boldsymbol{\beta}_0 + D(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})\boldsymbol{\beta} + \varepsilon$$

Heterogeneous case III

Hp: $e_1 = e_0$, only observable heterogeneity

1. This subcase assumes that unobservable heterogeneity is not at work and thus only observable heterogeneity matters.
2. This is a quite strong assumption, but one that holds in many applications, especially when the analyst has access to a large set of observable variables and is sure that diversity in units' outcome response is driven by these (available) observable factors

Selection models

Consider the model:

1. $$\begin{aligned} Y &= \mu_Y + \beta_Y x + \alpha D + u && \text{Outcome equation} \\ D &= \mu_D + \beta_D x + \varepsilon && \text{Selection equation} \\ u &= \gamma_u Q + e_u \\ \varepsilon &= \gamma_\varepsilon Q + e_\varepsilon \end{aligned}$$
2. Q a common unobservable component
3. Since Q is unobservable, it is part of both error terms u and ε
4. it can be shown that the bias of the OLS estimator takes the following form:

$$\alpha_{\text{OLS}} = \alpha + \frac{\text{Cov}(\varepsilon; u)}{\text{Var}(D)}$$

5. $$\alpha_{\text{OLS}} = \alpha + \gamma_\varepsilon \gamma_u \frac{\text{Var}(Q)}{\text{Var}(D)}$$

UNIVERSITÀ DI TRENTO

Selection models

Then we have:

1

- $\gamma_u \gamma_\varepsilon > 0$, then OLS has an upward bias.
- $\gamma_u \gamma_\varepsilon < 0$, then OLS has a downward bias.
- $\gamma_u \gamma_\varepsilon = 0$, then OLS is unbiased (consistent).

2

A *two-step* procedure can be used to estimate this equation:

1. Run a Probit of D_i on $(1, \mathbf{x}_i, z_i)$ and get: $\left(\widehat{\phi}_i, \widehat{\Phi}_i\right)$
2. Run an OLS of Y_i on: $\left[1, D_i, \mathbf{x}_i, D_i(\mathbf{x}_i - \boldsymbol{\mu}_\mathbf{x})_i, D_i\frac{\widehat{\phi}_i}{\widehat{\Phi}_i},\ (1 - D_i)\frac{\widehat{\phi}_i}{1-\widehat{\Phi}_i}\right]$

R

$$\hat{\alpha}_{\text{DDM}} = \frac{1}{n_{1t}} \sum_{i \in I_{1t} \cap S_P} \left\{ (Y_{1ti} - \sum_{j \in I_{0t} \cap S_P} W(i,j) Y_{0tj} \right\} - \frac{1}{n_{1t'}} \sum_{i \in I_{1t'} \cap S_P} \left\{ (Y_{0t'i} - \sum_{j \in I_{0t'} \cap S_p} W(i,j) Y_{0t'j} \right\}$$

To fix ideas

Bibliography

- Smith, J. A., and Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators?. Journal of econometrics, 125(1-2), 305-353.
- Cerulli, G. (2015). Econometric evaluation of socio-economic programs. Advanced Studies in Theoretical and Applied Econometrics Series, 49, Springer.