A note on Monte-Carlo simulation, multiple treatment effects using IV

Author: Simen Markussen and Knut Røed

Date: 18.11.2013

Description: This is a brief and informal description of a Monte Carlo simulation of a case with two treatment effects and IV-estimation. The exercise is referred to in the paper "The effects of vocational rehabilitation". In this note, we first present the key features of the experiment/simulation. The details (e.g., regarding specific distributional assumptions) are laid out in the STATA code displayed at the end of this document.

Setting: We study the following setting:

- There are 50,000 “agents” distributed between 500 different “treatment environments”. Agents are under risk for two different treatments, p1 and p2.
- We study the effect of these treatments on some outcome y = x + alfa + b1*p1 + b2*p2 + e, where x is an individual observable characteristic, alfa is an individual unobserved characteristic. b1 and b2 are the parameters of interest. e is a normal distributed error term
- Treatment is allocated in the following manner:
    - First we construct a latent participation propensity for treatment 1 and 2
        - latent1 = x + alfa + eps1 + phi1_true
        - latent2 = x + alfa + eps2 + phi2_true

      where eps1 and eps2 are individual random variables (uncorrelated) and phi1_true and phi2_true are the true (unobserved) parameters characterizing the influence of the treatment environments.
    - Phi1_true and Phi2_true are allowed to be correlated, and in our experiment they are negatively correlated such that treatment environments pushing hard for treatment 1 typically push less for treatment 2.
    - An agent participates in treatment *s* if this treatment is the one with the highest latent propensity and this propensity is higher than a cut-off level.
- Heterogeneous effects: the effect of a treatment is heterogeneous across agents, and these effects (b1 and b2) are negatively correlated, with unconditional means equal to 1 and 2, respectively.
- Non-compliers (never-takers): The quartile with the lowest effect of treatment *s* will never participate in treatment *s* (regardless of the latent propensities).
- Since the two effects are negatively correlated the typical situation is that there is very little overlap between the two groups of never-takers. In other words, those benefiting the most from treatment 1 will often be never-takers of treatment 2, and vice versa.

- Since the unobserved variable alfa enters both in the outcome and participation equations, estimation using OLS[1] would surely give biased results. We thus need an instrumental variable for p1 and p2. The variables we want to use are the treatment environments' contribution to the two treatment propensities. However, these variables (phi1_true and phi2_true) are unobserved.

- In line with the paper, we thus construct the estimates for phi1_true and phi2_true using a leave-out mean of the treatment participation of the other agents within the same treatment environment.

    - First we run a regression to control for x

        - $p1 = c_1x + e1$

        - $p2 = c_2x + e2$

    - Then we use the estimated error terms e1 and e2 in the following way: Within each treatment environment we compute the mean of e1 and e2 and then remove agent *i* from the mean of "his" office. This is our instrumental variables phi1 and phi2.

- Finally we estimate the model using 2SLS. The model has three equations:

    - $p1 = f_1x + f_{11}phi1 + f_{12}phi2+residual$

    - $p2 = g_1x + g_{11}phi1 + g_{12}phi2+residual$

    - y = x +b1 p1_hat +b2 p2_hat+residual

    where p1_hat and p2_hat are the predictions from the two first equations.

- We then compare the estimated coefficients for b1 and b2 to the mean of the true b1 and b2 in the population of compliers for each respective treatment, and find that the model is able to identify the mean of the true effect in the complier groups.

RESULTS

We implement 1000 repetitions using the specifications as below. There are 50 000 agents and 500 treatment environments in each trial. The true average effects of treatment 1 and treatment 2 in the two complier groups are denoted mcb1 and mcb2, respectively. The estimated effects are denoted beta1 and beta2. As one can see, on average they are as good as equal. Note also that the standard errors seem fairly correct by comparing the standard deviation of beta1 and beta2 to the man of the standard error se1 and se1.

[1] The equation would then be: y = x + b1p1 + b2p2

Variable	Obs	Mean	Std. Dev.	Min	Max
i	50000	25000.5	14433.9	1	50000
office	50000	249.51	144.3388	0	500
j	50000	50.49802	28.86636	1	100
n	50000	99.99604	.4449587	1	100
beta1	1000	1.104893	.0739729	.8557182	1.336544
beta2	1000	2.105409	.0751176	1.885114	2.364618
se1	1000	.073406	.0051149	.0594072	.0895748
se2	1000	.0732459	.0049674	.0594209	.0948801
mcb1	1000	1.105917	.001154	1.101229	1.109739
mcb2	1000	2.105903	.0011671	2.102338	2.109526

STATA CODE

```
*Monte-Carlo simulation for treatment-effects using IV
clear
set obs 50000 /* Number of individuals in simulations*/

* Just some infrastructure
g i = 1
replace i = sum(i)
gen office = floor(i/100) /* Social security offices for whom the instrumental variable is to
be constructed*/
gen j = 1
bys office: replace j = sum(j)
bys office: egen n = max(j) /* Number of participants per office*/

gen beta1 = .
gen beta2 = .
gen se1 = .
gen se2 = .
gen mcb1 = .
gen mcb2 = .


* The loop, choose number of rounds
forvalues k = 1(1)100{
* This step creates the "true" practice style of each office. They are drawn from a bivariate
normal distribution
* and they can be correlated. In this baseline set-up the true office styles are negatively
correlated such that
* if an office push hard for treatment 1 they tend to push less for treatment 2
matrix G = (1, -0.5\ -0.5,1)
matrix g = (0,0)
matrix s = (1, 1)
drawnorm v1 v2 , means(g) sds(s) corr(G)
replace v1 = . if j != 1
replace v2 = . if j != 1

/*
gen v1= rnormal(0,1) if j == 1
gen v2= rnormal(0,1) if j == 1
*/

bys office: egen phi1_true = max(v1)
bys office: egen phi2_true = max(v2)


* Individual participation
gen alfa = rnormal(0,1) /* person heterogeneity, unobserved to econometrician*/
gen eps1 = rnormal(0,1) /* random component treatment 1*/
```

```
gen eps2 = rnormal(0,1) /* random component treatment 2*/
gen x = rnormal(0,1) /* observed covariate*/


* True treatment effects
* Again we use a bivariat normal distribution with negatively correlated effects
matrix C = (1, -0.5 \ -0.5,1)
matrix m = (1,2)
matrix sd = (0.25, 0.25)
drawnorm b1 b2, means(m) sds(sd) corr(C)

* We define the compliers to the percentiles with the highest true effect of the treatment
quietly summ b1, det
gen complier1 = 1
replace complier1=0 if b1 < r(p25)
quietly summ b2, det
gen complier2 = 1
replace complier2=0 if b2 < r(p25)

* This is the latent probability of participation in treatment 1 and 2
gen latent1 = x+ alfa + eps1 + phi1_true
gen latent2 = x+ alfa + eps2 + phi2_true

* A person participates in treatment 1 if: the latent1 index exceeds 0.4 and latent1 > latent2
gen p1 = 0
gen p2 = 0
replace p1 = 1 if latent1 > latent2 & latent1 > 0.4
replace p2 = 1 if latent1 < latent2 & latent2 > 0.4

* The non-compliers do not participate anyhow
replace p1 = 0 if complier1 == 0
replace p2 = 0 if complier2 == 0

* Here we find the mean true effect for the compliers
quietly summ b1 if complier1 == 1
replace mcb1 = r(mean) if `k' == i
quietly summ b2 if complier2 == 1
replace mcb2 = r(mean) if `k' == i

tab p1 p2 /* A table of participants*/
drop eps1 eps2

* The outcome equation
gen y = x + alfa + b1*p1 + b2*p2 + rnormal(0,1)

* Constructing the Instrumental variables
quietly reg p1 x
predict fit1
gen eps1 = p1-fit1
bys office: egen meps1 = mean(eps1)
gen phi1 = (n*meps1-eps1)/(n-1) /* A leave-out mean*/

quietly reg p2 x
predict fit2
gen eps2 = p2-fit2
bys office: egen meps2 = mean(eps2)
gen phi2 = (n*meps2-eps2)/(n-1) /* A leave-out mean*/


* Simultaneous model
quietly ivregress 2sls y x (p1 p2 = phi1 phi2)
replace beta1 = _b[p1] if `k' == i
replace beta2 = _b[p2] if `k' == i
replace se1 = _se[p1] if `k' == i
replace se2 = _se[p2] if `k' == i

keep beta* se* j office n i mcb1 mcb2
}
summ
```