Technical note

The treatment effect: Comparing the ESR and PSM methods with an artificial example

By: Araar, A.: April 2015:

In this brief note, we propose to use an artificial example –data- in order to compare the results of PSM and those of the ESR model. We review also the theoretical framework to assess accurately the average treatment effects with the ESR model. Based on this, we show the subtle error in the theoretical framework of the paper of Sajaia and Luskin (2004)- and then in their mspredict post command. In addition to the corrected *mspredict* post command, a new *movestay* post command is produced (*msat*). This new post command can be used to estimate the ATT and ATE (see also Fuglie, K. O and D. J. Bosch (1995) for the theoretical framework). For more details, see the part B of this note.

- Fuglie, K. O and D. J. Bosch (1995). Implications of soil nitrogen testing: a switching regression analysis. *American Journal of Agricultural Economics* Vol.77: 891–900.

PART A: *The artificial example*

We assume that the number of observations is 1000:

```
set more off
clear all
set seed 1234
set obs 1000
```

Also, we assume that we have three regions:

```
gen    region    = 1 in   1/300
replace region   = 2 in 301/600
replace region   = 3 in 601/1000
```

It is assumed that the first region has more working age population:

```
gen      age  = min(int(runiform()*65+15), 65)
replace age  = age+5 if region==1
gen      educ = min(int(runiform()*5+1), 6)
```

It is assumed that the program is not randomly attributed and the population in region_1 have more probability to be selected. Also, it is assumed that the selection depends partially on the age:

```
set seed 7421
gen treatment=3*runiform()*(region==1)+0.5*runiform()*(region==2)+0.5*runiform()*(region==3)+(0.2+0.8*runiform())*(age> 30)
replace treatment    =  treatment > 1

local a =      0.6
local b =      0.1
gen e= `a'*runiform()

sum e if treatment ==1
qui replace e = e - r(mean) if  treatment ==1
sum e if treatment ==0
qui replace e = e - r(mean) if  treatment ==0
```

The outcome (income) depends on education, age, and the treatment. The parameter **a** enables to control for the predictive power of the two outcome models with the ESR method. The higher is **a,** the lower is the predictive power of the model. The parameter **b** enables to control for the contribution of the variable endogeneity (age). The higher is **b**, the higher is the endogeneity. In this artificial example, we know the exact value of the effect of the program, which is equal to **2**:

```
local at = 2
```

```
gen income = 60+0.5*educ+`b'*age+`at'*treatment + e
```

It is assumed that the variable age is not observed, but it affects jointly the program selection and the outcome. This raises the endogeneity problem, and we will need to use or to construct an instrumental variable (inst). The latter is assumed to be not explained by the outcome:

```
gen income0 =income -`at'*treatment
gen ins = (0.5+uniform())*age
regress ins income0
predict  inst, res
```

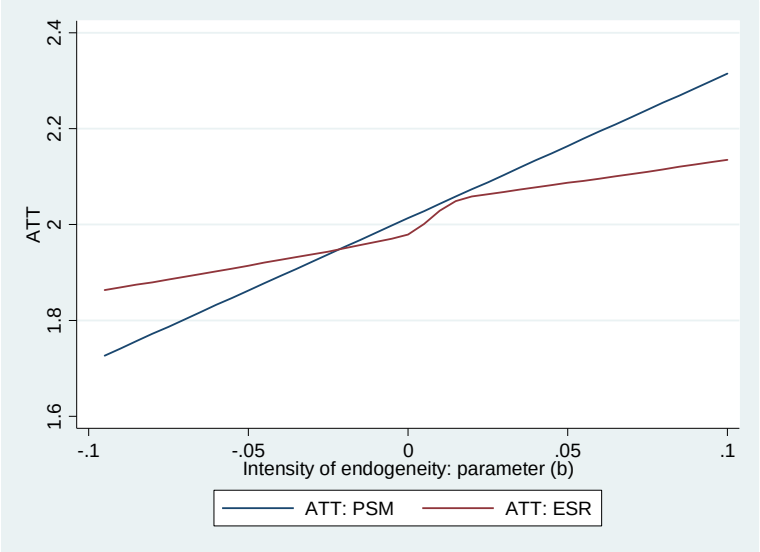
At this stage, we can estimate the effects with the PSM and ESR methods:

```
gen pw=1
xi: psmatch2 treatment i.region ins , outcome(income) cal(0.1) pw(pw) ate
local att_psm = r(att)
local atu_psm = r(atu)
gen lincome = log(income)
set seed 5241
xi: movestay lincome educ   , select(treatment  i.region ins )
msat treatment, expand(yes)
```

	PSM	ESR
ATT	2.212559	2.1918372
ATU	2.5462603	2.4961996

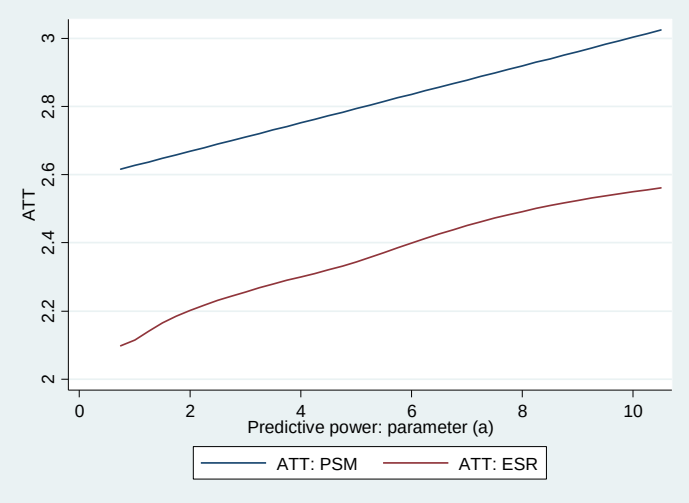
Based on the results above, the first conclusion is that the two models succeed to well capture the effect when the predictive power of the outcome models is high (lower level of a). Note that with the low endogeneity, the ESR turns to be an exogenous switching model, but the structure of the model continues to capture accurately the effect of the program.

Now, we would like to do more tests and check how the results are affected by the level of endogeneity (parameter b) where the latter varies between -0.1 and 0.1. For this end, we select a moderate level of the parameter a (a=0.6).

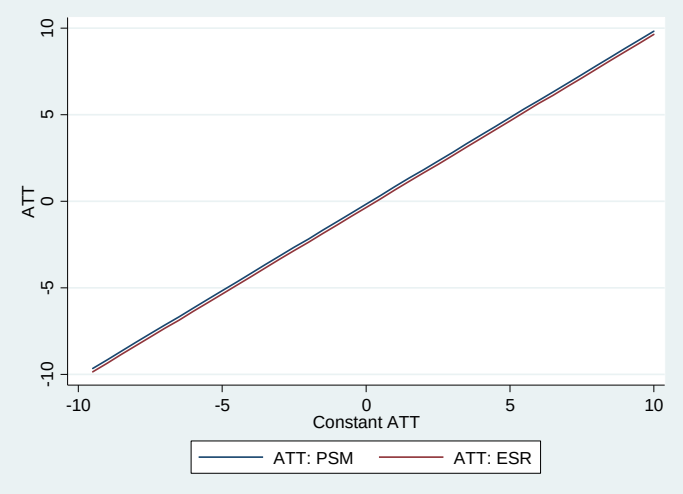


Better than the PSM, the ESR model seems to be helpful in presence of endogeneity and where the CIA PSM condition becomes less checked.

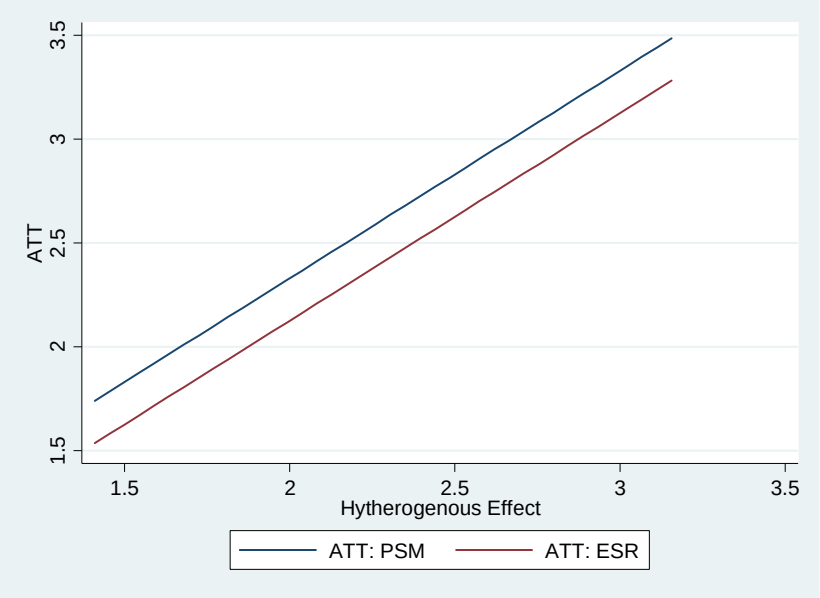
To checked sensitivity of results with predictive power of the ESR outcome models (inversely linked with the parameter a), we show the results according to a when b is fixed to 0.1. The two models succeed in estimating the affect, but the ESR shows a better performance.



Now, we control the parameters a and b, (a=0.6 and b=0.1) and we vary the predefined ATT (see the command to generate the income)



In the previous examples, we assume a homogenous treatment effect and even a reduced models can be used to estimate the impact. Now we assume that the treatment effect depends on the observed covariates (and no the observed covariate age: at=10+v*educt-0.01*age) and varies between 0.1 and 0.3 (also we have that: (a=0.6 and b=0.1)).



The PSM which is not conceived to treat the endogeneity problem is more biased in the case of heterogeneity through the unobservable part.

PART B: ***The ESR model and the estimation of average treatment effects***

Mainly, we assume a switching equation sorts individuals over two different states. With the *Endogenous Switching Regression model,* the ESR we assume that the observable outcome continue. Precisely, we have a model in which Consider the behavior of an agent with two binary outcome equations (participate to the program or not) and a criterion function T_i that determines which regime the agent faces (with migrant / without migrant). T_i can be interpreted as a treatment.

$$T_i=1 \quad \text{if} \quad \gamma Z_i + u_i > 0$$
$$T_i=0 \quad \text{if} \quad \gamma Z_i + u_i \leq 0 \qquad (1)$$
$$\text{Regime1}: y_{1i} = X_{1i}\beta_1 + : \epsilon_{1i} \quad \text{and} \quad y_{1i} = \mathrm{I}[y_{1i} \geq 0] \qquad (2)$$
$$\text{Regime0}: y_{0i} = X_{0i}\beta_0 + : \epsilon_{0i} \quad \text{and} \quad y_{0i} = \mathrm{I}[y_{0i} \geq 0] \qquad (3)$$

Where Are y_{1i}^* and y_{2i}^* are the two latent variables. We assume that the three residual: u_i , ϵ_{1i} et ϵ_{0i} are jointly normally distributed, with a mean-zero vector and correlation matrix

$$\Omega = \begin{bmatrix} \sigma_u^2 & \sigma_{1,u} & \sigma_{2,u} \\ \sigma_{1,u} & \sigma_1^2 & \\ \sigma_{2,u} & & \sigma_2^2 \end{bmatrix}$$

Where $\rho_l = Cov(u, \epsilon_l)$ *and* $l \in \{0,1\}$. Since y_{1i} and y_{0i} are not observed simultaneously, the joint distribution of (ϵ_1, ϵ_0) cannot be identified. In this estimation, we assume that $\rho_{0,1} = 1$. The estimation is done by the Full specification of Maximum Likelihood model. This model enables also to estimate the treatment effect on treated and untreated. The log likelihood function is defined as follows:

$$\ln L = \sum_i \big(I_i w_i \big[\ln\{F(\eta_{1i})\} + \ln\{f(\epsilon_{1i}/\sigma_1)/\sigma_1\} \big] + (1 - I_i) w_i \big[\ln\{F(\eta_{2i})\} + \ln\{f(\epsilon_{2i}/\sigma_2)/\sigma_2\} \big] \big)$$

Where $F(.)$ and $f(.)$ are respectively the density and cumulative density functions. w_i is an optional. Also, we have that:

$$\eta_{ji} = \frac{\gamma Z_i + \rho_j \epsilon_{ji}/\sigma_j}{\sqrt{1 - \rho_j^2}} \quad j = 1,2$$

Where ρ_j is the coefficient of correlation between ϵ_j and u .

The results of ESR can also be used to generate conditional expectations which will provide a concise measure of any efficiency differences among firms based on the credit market outcome. The following expressions are considered:

$$E(y_{1k}|I_k = 1, X_k) = X_{1k}\beta_1 + \rho_1\sigma_1 f(Z_k\gamma)/F(\gamma Z_k) \quad (10)$$

$$E(y_{1k}|I_k = 0, X_k) = X_{2k}\beta_1 - \rho_1\sigma_1 f(Z_k\gamma)/\{1 - F(Z_k\gamma)\} \quad (11)$$

$$E(y_{2k}|I_k = 1, X_k) = X_{1k}\beta_2 + \rho_2\sigma_2 f(Z_k\gamma)/F(Z_k\gamma) \quad (12)$$

$$E(y_{2k}|I_k = 0, X_k) = X_{2k}\beta_2 - \rho_2\sigma_2 f(Z_k\gamma)/\{1 - F(Z_k\gamma)\} \quad (13)$$

As we can observe, the sign that precedes ρ_1 and ρ_2 is corrected compared to what was reported in Sajaia and Luskin (2004)-movestay command- as well as their related Stata paper. The subtle error comes from omitting the negative sign for the de definition of the Mills ratio for the non-participants group (i.e. : $Mills(2) = +\rho_2 * (-1) * f(Z_k\gamma)/\{1 - F(Z_k\gamma)\}$). Thus, even if the results of *movestay* Stata command are accurate, some of those of the *mspredict* are wrong. At this stage, I have corrected temporarily this post command (mspredict_ar.ado) and I will contact the *movestay* authors. I addition, I have programmed for the PEP teams the post command *msat* to estimate the treatment effects.

syntax : **msat** *varlist(min=1 max1)*, [hhsize(varname) expand(string)]

The post-command **msat** enables to estimate quickly the average treatment affects after the movestay command. The varname that follows this command is the dummy variable of the treatment. The estimation takes automatically into account the sampling weight.

Options:

hsize Household size. For example, to compute poverty at the individual level, one will want to weight household-level observations by household size (in addition to sampling weights, best set in survey design).

expand If we use the log of the outcome variable with the movestay command and we like to estimate the treatment effect on the outcome (not on the log of outcome), the user can add the option: expand(yes).

Example:

Estimated treatment effects based on the endogenous switching regression model

Index	Estimate	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ATT	2.192045	.0704571	31.1118	0.0000	2.053784	2.330306
ATU	2.496426	.0627154	39.8056	0.0000	2.373357	2.619495
ATE	2.374978	.0472344	50.2807	0.0000	2.282288	2.467668

Note: The estimated standard errors ommit the part of samplind errors related to the estimation of the Beta's coefficients with the ESR model.

where *lincome* is the log of the income (the outcome in this example).

How to add the new post commands?

Copy simply the *mspredict_ar.ado* and *msat.ado* files in c:/ado/plus/m/