Measuring Bivariate Average Treatment Effect

Patrick Franco Alves[1]
Brasília University - UnB, Brasília, Brazil
Correspondence author: patrickfrancoalves@yahoo.com.br

Gustavo T. L. da Costa
Brazilian Institute of Geography and Statistics - IBGE, Rio de Janeiro, Brazil
gustavo.costa@ibge.gov.br

Abstract
This paper presents some issues concerning the measurement of two average treatment effects under a bivariate selection mechanism influence. There are many situations in which a bivariate self-selection mechanism operates over a dependent variable of interest. This paper presents the formulation of the bivariate average treatment effect from the multivariate Heckman model found in the multivariate sample-selection model literature. The self-selection processes are explained by a bivariate probabilistic model, more specifically the bi-probit model. The use of a bivariate normal distribution is necessary in order to derive the bivariate inverse mill's ratio, which appears quite different from the univariate type. Under this approach there are seven different combinations for the average treatment effect, each one measuring a specific feature of the treatments. An application case is done using a cross-section data, the Brazilian Innovation Survey (PINTEC, 2008). Since there is no direct solution to calculate the standard errors of the parameter estimates we used a bootstrap method. It is shown that this methodology can be used in any bivariate self-selection model since there is not an intricate computational solution for the problem.

Keywords: Bivariate treatment effect; Bivariate probit; Multivariate Heckman model.

1. Introduction

The average treatment effect (ATE) has an increasing importance in public policy evaluations. The availability of large observational surveys and the interest of measuring the impact of a public policy is making ATE a popular methodology. These database, which contains lots of individual characteristics, allow the implementation of more structured class of econometric models. This highlights the importance of micro-econometric developments in order to cover some issues about the selection process acting over an interest variable.

The work of Heckman (1978) has backgrounded the use of ATE for measuring the impact of a public policy program, known as a dummy endogenous variable. The quantitative framework of univariate ATE is by now widely developed and applied. On the other hand, the literature covering more than one selection mechanisms is still incipient, appearing only in De Luca and Peracchi (2006). Although none of these deals specifically with the bivariate ATE measurement problem.

The main objective concerning the existence of two self-selection mechanisms is the correlation that both processes exhibit one another and with the dependent variable of impact. This leads to a trivariate normal density in the case of a probit link-function. Consequently, the inverse of the Mill's ratio exhibits a quite different formulation from the univariate self-selection mechanism case. We begin by overviewing the multivariate Heckman model which constitutes a generalization of the bivariate approach.

---

[1] Patrick Alves thanks the financial support from FINATEC/UnB.

2. The multivariate case

The multivariate sample-selection model is given by the set of equations represented in (1) and (2) (Tauchmann, 2006), denoting the latent variable underlying the relationship between the observable variables z_{ij} and y_{ij}:

$$y^*_{ij} = x_{ij}\boldsymbol{\beta}_j + \varepsilon_{ij} \tag{1}$$

$$z_{ij} = \begin{cases} 1 & if \quad z^*_{i1} \geq 0 \\ 0 & if \quad z^*_{i1} < 0 \end{cases}; \qquad where \quad z^*_{ij} = w'_{ij}\gamma_j + \xi_{ij} \tag{2}$$

$$y_{ij} = z_{ij} y^*_{ij} \tag{3}$$

where j denotes the selection equations ($j=1,\ldots,m$) and i denotes individual observation ($i=1,\ldots,n$). The m-selection equations acts simultaneously over all the observable responses variables (q_{ij}) and interacts between themselves through the multivariate error correlation structure[2]. The linking structure of the errors can be represented by the following covariance matrix:

$$VAR(\varepsilon,\xi) = \begin{bmatrix} \sum(\varepsilon,\varepsilon)\sum{}'(\varepsilon,\xi) \\ \sum(\varepsilon,\xi)\sum(\xi,\xi) \end{bmatrix} \tag{4}$$

As noted by Tauchmann, because the computational effort is considerable non trivial it is advisable to first estimate a multivariate probit model (2) which gives the estimation for $\boldsymbol{\gamma}_j$. In a second run, using the multivariate inverse Mill's expression, it is possible to get consistent, but inefficient estimation of $\boldsymbol{\beta}_j$ (1). A consistent generalized two-step estimation of the equations (1) to (4) is provided by the following equations:

$$y_{ij} = z_{ij}\,\mathbf{x}'_{ij}\,\boldsymbol{\beta} + z_{ij}\sum_{h=1}^{H}\beta_{\lambda,ih}\,\psi_{ih}\;\phi\left(\mathbf{w}'_{hj}\,\boldsymbol{\beta}_h\right)\frac{\Phi_{h-1}\left(\tilde{A}_{hj},\,\tilde{R}_{hj}\right)}{\Phi_h(\,\cdot\,)} + z_{ij}\,\varepsilon_{ij} \tag{5}$$

$$\psi_{ih} = 2z_{ih} - 1 \tag{6}$$

$$\tilde{A}_{hj} = \frac{\psi_{lj}(\mathbf{w}'_{lj}\,\boldsymbol{\gamma}_l - \rho_{lh}^{\mu\mu}\,\mathbf{w}'_{hj}\,\boldsymbol{\gamma}_l)}{\sqrt{1-(\rho_{lh}^{\mu\mu})^2}}, \quad h = 1,\ldots,H; \quad l \neq h \tag{7}$$

where $\psi_{ih}\,(i \neq h)$ are used as the elements for a diagonal matrix Ψ and $\tilde{R}_{hj}$ define a partial correlation matrix $R_{hj} = \underset{\sim}{\rho}[\mu_j \mid \mu_{hj}]$, with elements $\tilde{R}_{hj} = \Psi_{hj} R_{hj} \Psi_{hj}$. In (5) the normal multivariate distributions with dimension h and $h-1$ are denoted by respectively, Φ_h and Φ_{h-1}. The parameters estimation of (5) provided by OLS yield to unbiased estimators, but the standard errors (SEs) are still inconsistent. In the absence of the analytic expression for the covariance parameter estimator matrix a bootstrap approach is done for the SES[3].

[2] In Tauchmann's, because of the weightings, the variables (y_{ij}) are not observable for $y_{ij}=0$.

[3] In section 4 we present the methodology of Murphy &Topel (1985) for asymptotic covariance from a two step procedure.

3. BATE: Bivariate Analysis Treatment Effect

3.1 The bivariate selection mechanism

Let the bivariate self-selection mechanism be generated by a bivariate normal density where ρ is the correlation parameter.

$$z_{i1} = \begin{cases} 1 & if \quad z^*_{i1} \geq 0 \\ 0 & if \quad z^*_{i1} < 0 \end{cases}; \qquad where \quad z^*_{i1} = w'_{i1}\gamma_1 + \xi_{i1} \tag{8}$$

$$z_{i2} = \begin{cases} 1 & if \quad z^*_{i2} \geq 0 \\ 0 & if \quad z^*_{i2} < 0 \end{cases}; \qquad where \quad z^*_{i2} = w'_{i2}\gamma_2 + \xi_{i2} \tag{9}$$

$$y_i = \delta_1 z_{i1} + \delta_2 z_{i2} + \mathbf{x}'_i\boldsymbol{\beta} + \beta_{\lambda1}\lambda_{i1} + \beta_{\lambda2}\lambda_{i2} + \varepsilon_i \tag{10}$$

where z^*_{i1} and z^*_{i2} denote the latent correlated selection process which generates the dichotomous variables in (8) and (9). Using the example of the product (z^*_{i1}) and process innovation (z^*_{i2}), this can be seen as an unobserved *desire* from the firms' management to implement the two kinds of innovation. In (10) the dependent variable simultaneously affected by y^*_{i1} and y^*_{i2} is denoted by q_i

$$\phi_2\left(z^*_{i1}, z^*_{i2}, \rho\right) = \frac{1}{2\pi\sqrt{1-\rho[\xi_1;\xi_2]^2}} \exp - \left[\frac{z^{*\,2}_{i1} + z^{*\,2}_{i2} - 2\rho[\xi_1;\xi_2]\, z^*_{i1}\, z^*_{i2}}{2\{1-\rho[\xi_1;\xi_2]^2\}}\right] \tag{11}$$

The bivariate inverse mills expression are developed using the following ancillary functions, which follows a notation close to Greene (2007, 787):

$$q_{i1} = 2z_{ij} - 1; \quad q_{i2} = 2z_{ij} - 1; \quad \rho_{*j} = q_{i1}\, q_{i2}\, \rho[\xi_1;\xi_2] \tag{12}$$

$$P(Z_1 = z_1, Z_2 = z_2 \,|\, \mathbf{w}_1;\mathbf{w}_2) = \Phi_2(q_{i1} s_{i1},\, q_{i2} s_{i2},\, \rho_{*j}) \tag{13}$$

$$k_{ij} = s_{ij}\, q_{ij}, \qquad where \quad s_{ij} = w_{ij}\,\gamma_{ij} \tag{14}$$

$$g_{ij} = \phi(k_{ij})\Phi([k_{ik\neq j} - \rho_{*j}k_{ij}] / \sqrt{1-\rho^2_{*j}}) \tag{15}$$

Dividing (15) by the bivariate cumulative normal density leads to the inverse Mills Ratio. In the univariate selection process there were only two expressions for the inverse Mills Ratio. In the present case there are seven cases corresponding to all combinations of z_{i1} and z_{i2}:

$$\lambda_{i1} = \phi(k_{i1}) \frac{\Phi([k_{i2} - k_{i1}] / \sqrt{1-\rho^2_{*j}})}{\Phi_2(k_{i1}, k_{i2}, \rho_{*j})} \tag{16}$$

$$\lambda_{i2} = \phi(k_{i2}) \frac{\Phi([k_{i1} - k_{i2}] / \sqrt{1-\rho^2_{*j}})}{\Phi_2(k_{i1}, k_{i2}, \rho_{*j})} \tag{17}$$

The equations below give the conditional mathematical expectancies. The final expressions for BATE are described on table 3.

$$E\left[\, y_i \,|\, z_{i1} = 0, z_{i2} = 0\right] = \mathbf{x}'_i\boldsymbol{\beta} + \beta_{\lambda1}\lambda_{i1} + \beta_{\lambda2}\lambda_{i2} \tag{18}$$

$$E\left[\, y_i \,|\, z_{i1} = 1, z_{i2} = 0\right] = \delta_1 + \mathbf{x}'_i\boldsymbol{\beta} + \beta_{\lambda1}\lambda_{i1} + \beta_{\lambda2}\lambda_{i2} \tag{19}$$

$$E\left[\, y_i \,|\, z_{i1} = 0, z_{i2} = 1\right] = \delta_2 + \mathbf{x}'_i\boldsymbol{\beta} + \beta_{\lambda1}\lambda_{i1} + \beta_{\lambda2}\lambda_{i2} \tag{20}$$

$$E\left[\, y_i \,|\, z_{i1} = 1, z_{i2} = 1\right] = \delta_1 + \delta_2 + \mathbf{x}'_i\boldsymbol{\beta} + \beta_{\lambda1}\lambda_{i1} + \beta_{\lambda2}\lambda_{i2} \tag{21}$$

3.2. Robust covariance matrix and Murphy and Topel theorem

The robust covariance matrix estimative consider: $y = h(\mathbf{x}, \boldsymbol{\beta}, \mathbf{w}, \boldsymbol{\gamma}) + \varepsilon$, and the covariance matrix $\mathbf{V}_b$, where $\boldsymbol{\gamma}$ are estimated in a first step having $\mathbf{x}$ as the explanatory variables (from Murphy & Topel, 1985). The second step is $y = h(\mathbf{x}, \boldsymbol{\beta}, \mathbf{w}, \boldsymbol{\gamma}) + \varepsilon$. Assuming asymptotic normality and that the covariance estimates for $\boldsymbol{\gamma}$ are unbiased[4]. The step below indicates the construction of the second stage covariance matrix estimate ($\hat{\mathbf{V}}_b$). Let $\hat{\boldsymbol{\beta}}$ be an unbiased estimator of $\boldsymbol{\beta}$ and $\hat{\mathbf{V}}_b$ the covariance matrix estimator. If $s^2\,\hat{\mathbf{V}}_b$ is an appropriate estimation[5] of $\sigma^2\,\mathbf{V}_b = \sigma^2(\mathbf{X}^{0\prime}\mathbf{X}^0)^{-1}$, then, under the necessary conditions the covariance matrix given by:

$$\hat{\mathbf{V}}_b = \frac{1}{n}[\sigma^2\mathbf{V}_b + \mathbf{V}_b(\mathbf{C}\,\mathbf{V}_c\,\mathbf{C}' - \mathbf{C}\;\mathbf{V}_c\,\mathbf{R}' - \mathbf{R}\,\mathbf{V}_c\,\mathbf{C}')\mathbf{V}_b] \qquad \textbf{(22)}$$

$$\mathbf{C} = p\lim\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^0\,\hat{\varepsilon}_i^2\left[\frac{\partial h(\mathbf{x}_i, \boldsymbol{\beta}, \mathbf{w}_i, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}'}\right] \qquad \textbf{(23)}$$

$$\mathbf{R} = p\lim\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^0\,\hat{\varepsilon}_i^2\left[\frac{\partial g(\mathbf{w}_i, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}'}\right] \qquad \textbf{(24)}$$

and $\partial g(.)/\partial\gamma$ is the gradient of the ith term in the likelihood for $\boldsymbol{\gamma}$ in the first stage. The first derivatives of the expressions (16) and (17) are necessary in order to get the **C** matrix, as the **R** matrix depends on the score function of the bivariate probit likelihood. Given the trouble calculation for the appropriate variance construction, in the present paper the robust variance estimates are done via bootstrap method.

3.3. Bootstrap Estimation

One of the main concerns of our work is in the estimation of Bate's models variance estimators. To outline the problem it is not trivial to construct the algebra of the cited problem and we will present it in a future work. One way to surpass this problem was the use of simulation strategies to investigate the properties from the estimators. In the case of regression models the main concern is about the consistency of the estimators originated from the method used to generate the replicas from the study. We used a nonparametric bootstrap analysis where the main concern was to generate robust SES and confidence intervals for the estimates. These methods are weak consistent for the parameters estimates what is a sufficient characteristic for most statistical problems (Shao and Tu, 1995).

To avoid problems in estimation of the SES and confidence limits (CLs) in cases where the population could be generated from a highly skewed distribution we used 5.000 replicas[6]. There are a few bootstrap confidence sets in the literature, bootstrap-t, bootstrap percentile, bootstrap bias-corrected percentile (BC), bootstrap accelerated bias-corrected percentile (BCa) and the hybrid bootstrap. Even though, BCa and the bootstrap-t methods are more accurate they are not an easy task to implement. We chose the hybrid method (Shao and Tu, 1995) which has the same accuracy as the traditional normal approximation when a considerable size of resample is used.

4. Application using PINTEC data and other database

Three data base were gathered, on firm level for the year 2008: the Brazilian Annual Survey of Industry (PIA/IBGE-A), the Brazilian Innovation Survey (PINTEC/IBGE-B) and the Annual Relation of Social Information (RAIS/MTE-C). We describe below the variables. The list of the variables is:

(1)*Product (Process) Innovation* (*Product(Process)*)(B): Dummy variable concatenation of two variables: "Between 2006-2008, the firm introduced a new product (*process*) or significantly improved for the firm, but existing in the domestic

---

[4] $\hat{\boldsymbol{\gamma}}$ is the first round estimate of $\boldsymbol{\gamma}$, having variance-covariance matrix $\mathbf{V}_c$.

[5] $\mathbf{X}^{\text{o}}$ is the pseudo-regressions evaluated at the true parameter values: $\mathbf{x}_i^0 = \partial h(\mathbf{x}, \boldsymbol{\beta}, \mathbf{w}, \hat{\boldsymbol{\gamma}})/\partial\beta$.

[6] For the SE and CL of the estimates, two SAS macros, %BOOT and %BOOTIC, respectively.

market?" and "Between 2006-2008, the firm introduced a new product (*process*) or significantly improved a new product for the domestic market?"; **(2)***National*: Dummy variable indicating if the firm has no foreign controller capital; **(3)***Firm Size* (*l*)(A): Total numbers of employees; **(4)***Labor Productivity(y/l)*(A): Total revenue over total number of employees*(l)*; **(5)***Market Share* (*Share*)(A): Ratio total # of employees (Firm Level/Sector Level); **(6)***R&D*(B): Total spending on R&D activities. Includes intra and extramural R&D; **(7)***Total Revenue*: Total revenue; **(8)***R&D Effort* (B): Ratio between R&D and total sales; **(9)***Schooling* (*Skill*)(C): Weighted average of employees schooling; **(10)***Gross fix capital stock*(*k*)(A/C): Capital stock measured by perpetual inventory method, according to methodology proposed by Alves and Silva (2007); **(11)***Turnover rate* (rot)(C): Percentage of employees that leaves the firm on the next year; **(12)***Geographic Localization* (*locus*)(B): Geographic five major brazilian regions; **(13)***Economic Class.* (*ocde*[7])(B): (i)Extractive; (ii)High-Tech;(iii)Medium-HT;(iv) Medium-LT;(v) LT;(vi) Services; **(14)***Cooperation for innovation (Coop)*(B)– Cooperation of the firm. Seven levels; **(15)***Age in years of the firm (age)*(C)

4.1 Results for the first stage bi-variate probit model

We report the results in which the objective was to investigate the impact of Product and Process Innovation over Labor Productivity. This impact is what we call a bivariate average treatment. In the microeconomic literature the process and product innovation appears to be correlated in the sense that entrepreneurs decide to do both kinds of innovations in order to successfully internalize the economics results. This is the reason why many authors criticized the ATE, that separately explains the product and process innovation and seek its impacts on economic performance. Considering the correlation between the two of innovative activities we present all models in (34).

$$[prod_{imkl};\, proc_{imkl}] = \Phi[\gamma_0 + \gamma_1 RDeffort_i + \gamma_2 Skill_i + \gamma_3 Share_i + \\ + \gamma_4 National_i + \gamma_m\, coop_{im} + \eta_k\, locus_{ik} + \mu_l\, ocde_{il}] \quad \textbf{(25)}$$

Table 1 - First Step Estimation: Bivariate Probit Model (2008)

		Product				*Process*			
Parameter	Levels	Estimate	Std Error	**t** Value	P-Value	Estimate	Std Error	t value	P-Value
Intercept		-1,951	0,198	-9,87	<0,001	-1,810	0,184	-9,83	<0,001
R&D Effort		0,002	0,001	3,56	0,001	0,001	0,001	0,15	0,879
Skill		0,705	0,073	9,69	<0,001	0,665	0,067	9,95	<0,001
Share		5,019	1,072	4,68	<0,001	5,468	1,177	4,64	<0,001
National		-0,233	0,060	-3,86	0,001	-0,051	0,060	-0,86	0,391
Cooperation	*1*	1,212	0,095	12,81	<0,001	1,051	0,095	11,01	<0,001
	2	0,955	0,086	11,05	<0,001	1,466	0,101	14,53	<0,001
	3	1,412	0,300	4,70	<0,001	2,125	0,439	4,84	<0,001
	4	0,974	0,197	4,93	<0,001	0,643	0,190	3,38	0,001
	5	1,157	0,217	5,33	<0,001	1,238	0,232	5,34	<0,001
	6	0,838	0,204	4,11	<0,001	1,456	0,236	6,18	< 0,001
	7	-	-	-	-	-	-	-	-
Correlation		0,716	0,010	72,65	< 0,001				

Source: Elaborated by the authors from 2008, Brazilian Innovation Survey, PIA and RAIS.
Note: The algorithm converged with the *qlim procedure* (SAS), using the method NEWRAP.

The process and product correlation estimated was 71.6%, so we do have a bivariate process between these two variables and the average schooling of the labor

[7] Classification from Organization for Economic Cooperation and Development (OECD).

force had a positive and significant high impact over both the product and process innovation. The same is true for the market concentration. With a closer look at the results of BATE 2, 4, 6 and 7 it is clear that innovating only in product gives much more impact against innovating only in process or not innovating at all.

Table 2 – Second Step Estimation: Linear regression conditional to first step

Variable	Estim.	Boot. SE	t value	p-value	Variable	Estim.	Boot. SE	t-value	p-value
Int.	1.646	0.071	8,76	<0.001	*ln(cap)*	0.064	0.002	34.1	<0.001
Prod	0.292	0.043	6.79	<0.001	*ln(age)*	0.167	0.020	8.55	<0.001
Proc	-0.029	0.035	-0.84	0.401	Rot	-0.090	0.040	-2.24	0.025
ln(posgrad)	0.173	0.055	3.16	0.002	Mills(Prod)	-0.045	0.022	-2.01	0.044
ln(employee)	0.060	0.013	4.49	<0.001	Mills(Proc)	0.194	0.044	4.43	<0.001
R^2	0.194								

Source: Elaborated by the authors from 2008, Brazilian Innovation Survey, PIA and RAIS.
Note: Results from the *reg procedure* (SAS).

Table 3 – Bivariate Average Treatments

Effects	Est.	Boot Lower C.L.	Boot Upper C.L.
$BATE1 = E[y_i \mid z_{i1}=1; z_{i2}=1] - E[y_i \mid z_{i1}=0; z_{i2}=0]$	0.832	0.801	0.866
$BATE2 = E[y_i \mid z_{i1}=1; z_{i2}=0] - E[y_i \mid z_{i1}=0; z_{i2}=0]$	3.876	3.711	4.069
$BATE3 = E[y_i \mid z_{i1}=0; z_{i2}=1] - E[y_i \mid z_{i1}=0; z_{i2}=0]$	1.223	1.179	1.276
$BATE4 = E[y_i \mid z_{i1}=1; z_{i2}=1] - E[y_i \mid z_{i1}=1; z_{i2}=0]$	-3.043	-3.206	-2.909
$BATE5 = E[y_i \mid z_{i1}=1; z_{i2}=1] - E[y_i \mid z_{i1}=0; z_{i2}=1]$	-0.389	-0.426	-0.364
$BATE6 = E[y_i \mid z_{i1}=1; z_{i2}=0] - E[y_i \mid z_{i1}=0; z_{i2}=1]$	2.654	2.519	2.811
$BATE7 = E[y_i \mid z_{i1}=1; z_{i2}=0] - E[y_i \mid z_{i1}=1; z_{i2}=0]$	-2.009	-2.146	-1.873

Source: Elaborated by the authors from 2008, Brazilian Innovation Survey, PIA and RAIS.

5. Conclusions

The bivariate analysis of treatment (BATE) can provide a great analytic potential compared with the univariate analysis of treatment. This is because the BATE considers the join process of self-selection acting over an impact variable of interest. For example, in BATE methodology we can compute seven different possibilities for the effects. In a future work we can construct the analytical asymptotic expression for the covariance matrix in the second stage.

6. References

Alves, P. F. and Silva, A. M, (2008), “Estimativa do Estoque de Capital das Empresas Industriais Brasileiras”. Discussion Text 1325, IPEA, Brasília.

De Luca, G. and Peracchi, F.(2007), “A Sample Selection Model for Unit and Item Nonresponse in Cross-Sectional Surveys”. CEIS Research Paper 95, Tor Vergata University, CEIS.

Greene, W. (2007) *Econometric Analysis*, Prentice Hall, 6th ed.

Heckman, J. J. (1976), “Dummy Endogenous Variables in a Simultaneous Equation System”, Econometrica, V. 46, 6; pg. 931.

Murphy, K. M. and Topel, R. H. (1985), “Estimation and Inference in Two-Step Econometric Models” Journal of Business & Economic Statistics, ASA, vol. 3(4), p. 370-79.

Shao, J and Tu, D. (1995). *The Jackknife and Bootstrap,* Springer Series in Statistic.

Tauchmann, H.*A* (2006) *“*Note on Consistency of Heckman-type Two-step Estimators for the Multivariate Sample-Selection Model*”*, RWI: Discussion Paper Nº 40.