Advances and Applications in Statistics
Published Online: January 2015
Available online at http://www.pphmj.com/journals/adas.htm
Volume 43, Number 2, 2014, Pages 91-105

MEASURING BIVARIATE AVERAGE TREATMENT EFFECT

Patrick Franco Alves and Gustavo T. L. da Costa

Department of Economics
Brasília University-UnB
Brasília, Brazil
e-mail: patrickfrancoalves@yahoo.com.br

Brazilian Institute of Geography and Statistics-IBGE
Rio de Janeiro, Brazil
e-mail: gustavo.costa@ibge.gov.br

Abstract

We present the methodology for the measurement of the average treatment effects under a bivariate selection mechanism. The formulation of the bivariate average treatment effect comes from the multivariate sample-selection model, where the bivariate normal distribution is necessary in order to derive the bivariate inverse Mills ratio. Under this approach there are seven different average treatment effects. An application case is done using a cross-section data for Brazilian industrial firms. It is shown that this methodology can be easily used in any bivariate self-selection mechanism case since there is not an intricate computational solution for the problem.

Received: April 8, 2014; Revised: September 23, 2014; Accepted: September 27, 2014

2010 Mathematics Subject Classification: 62J05, 62P25, 62-07.

Keywords and phrases: bivariate treatment effect, bivariate probit, multivariate Heckman model, sample selection.

1. Introduction

The average treatment effect (ATE) estimation has an increasing importance in public policy evaluations. The availability of large observational surveys and the interest of measuring the impact of public policies is making ATE a necessary and popular methodology. The need of more structured class of econometric models 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 [4] has backgrounded the use of ATE for measuring the impact of a public policy, known as a dummy endogenous variable. The quantitative framework of the ATE is widely developed and applied. On the other hand, the literature covering more than one selection mechanism is still incipient, appearing only in [2], and do not deal specifically with the bivariate ATE measurement problem. The main concern in the existence of two self-selection mechanisms is the correlation that both processes exhibit one another and with the dependent variable of interest and, in the case of a probit link-function, this leads to a trivariate normal distribution. Consequently, the inverse of the Mills ratio exhibits a different formulation from the univariate case.

2. The Multivariate Case

We begin by overviewing the multivariate Heckman model which constitutes a generalization of the bivariate approach. The multivariate sample-selection model is given by the set of equations represented in (1) and (2) (Tauchmann [7]), denoting the latent variable underlying the relationship between the observable variables z_{ij} and y_{ij}:

$$y_{ij}^* = \mathbf{x}_{ij}\boldsymbol{\beta}'_j + \varepsilon_{ij}, \tag{1}$$

$$z_{ij}^* = \mathbf{w}'_{ij}\boldsymbol{\gamma}_j + \xi_{ij}, \tag{2}$$

$$z_{ij} = \begin{cases} 1, & \text{if } z_{ij}^* \geq 0, \\ 0, & \text{if } z_{ij}^* < 0, \end{cases} \tag{3}$$

$$y_{ij} = y_{ij}^{*} z_{ij}, \tag{4}$$

where *j* denotes the selection equations $(j = 1, ..., m)$ and *i* denotes individual observation $(i = 1, ..., n)$. The *m*-selection equations act simultaneously over all the observable response variables (q_{ij}) and interact between themselves through the multivariate error correlation structure. In Tauchmann's formulation, because of the weightings $y_{ij}^{*} z_{ij}$, the variables (y_{ij}) are not observable for $y_{ij} = 0$, however in ATE evaluation the impact variable (y_{ij}) is available for $y_{ij} = 1$ and $y_{ij} = 0$.

The errors structure linking the errors $\varepsilon'_j = [\varepsilon_{i1}, ..., \varepsilon_{im}]$ and $\xi'_j = [\xi_{i1}, ..., \xi_{im}]$ can be represented by the following covariance matrix:

$$VAR(\varepsilon, \xi) = \begin{bmatrix} \Sigma(\varepsilon, \varepsilon) & \Sigma'(\varepsilon, \xi) \\ \Sigma(\varepsilon, \xi) & \Sigma(\varepsilon, \xi) \end{bmatrix}. \tag{5}$$

As noted by Tauchmann, 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 Mills 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:

$$y_{ij} = z_{ij}\mathbf{x}'_{ij}\boldsymbol{\beta} + z_{ij}\sum_{h=1}^{H} \beta_{\lambda, ih}\, \psi_{ih}\, \phi(\mathbf{w}'_{hj}\, \boldsymbol{\beta}_h) \frac{\Phi_{h-1}(\widetilde{A}_{hj}, \widetilde{R}_{hj})}{\Phi_h(\cdot)} + z_{ij}\varepsilon_{ij}, \tag{6}$$

$$\psi_{ih} = 2z_{ih} - 1, \tag{7}$$

$$\widetilde{A}_{hj} = \frac{\psi_{lj}(\mathbf{w}'_{lj}\boldsymbol{\gamma}_l - \rho_{lh}^{\mu\mu}\mathbf{w}'_{hj}\boldsymbol{\gamma}_l)}{\sqrt{1 - (\rho_{lh}^{\mu\mu})^2}}, \quad h = 1, ..., H; \quad l \neq h, \tag{8}$$

where $\psi_{ih}\ (i \neq h)$ are used as the elements for a diagonal matrix Ψ and $\widetilde{R}_{hj}$ defines a partial correlation matrix $R_{hj} = \underset{\sim}{\rho}[\mu_j \,|\, \mu_{hj}]$, with elements $\widetilde{R}_{hj} = \Psi_{hj} R_{hj} \Psi_{hj}$. In (6) the normal multivariate distributions with dimension *h* and $h-1$ are denoted by, respectively, Φ_h and Φ_{h-1}. The parameters estimation of (6) provided by OLS leads to unbiased estimators, but the standard errors are still inconsistent. In the absence of the analytic expression for the covariance parameter estimator matrix a bootstrap approach is done for the standard errors[1].

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:

$$z_{i1}^* = \mathbf{w}_{i1}'\gamma_1 + \xi_{i1}, \tag{9}$$

$$z_{i2}^* = \mathbf{w}_{i2}'\gamma_2 + \xi_{i2}, \tag{10}$$

$$z_{i1} = \begin{cases} 1, & \text{if } z_{i1}^* \geq 0, \\ 0, & \text{if } z_{i1}^* < 0, \end{cases} \tag{11}$$

$$z_{i2} = \begin{cases} 1, & \text{if } z_{i2}^* \geq 0, \\ 0, & \text{if } z_{i2}^* < 0, \end{cases} \tag{12}$$

$$y_i = \delta_1 z_{i1} + \delta_2 z_{i2} + \mathbf{x}_i'\boldsymbol{\beta} + \beta_{\lambda 1}\lambda_{i1} + \beta_{\lambda 2}\lambda_{i2} + \varepsilon_i, \tag{13}$$

where z_{i1}^* and z_{i2}^* denote the latent correlated selection process which generates the dichotomous variables in (11) and (12). In (13) the dependent variable simultaneously affected by y_{i1}^* and y_{i2}^* is denoted by q_{ij}:

$$\phi_2(z_{i1}^*, z_{i2}^*, \rho) = \frac{1}{2\pi\sqrt{1-\rho[\xi_1; \xi_2]^2}} \exp -\left[\frac{z_{i1}^{*2} + z_{i2}^{*2} - 2\rho[\xi_1; \xi_2] z_{i1}^* z_{i2}^*}{2\{1-\rho[\xi_1; \xi_2]^2\}}\right]. \tag{14}$$

---

[1]In Section 4, we briefly present the methodology of Murphy and Topel [5]. But in the case of the bivariate selection mechanism presented in this paper still is cumbersome to derive it. For this reason we had made the choice of a bootstrap estimation.

The bivariate inverse Mills expression are developed using the following ancillary functions, which follows a notation close to Greene [3, p. 787]:

$$q_{i1} = 2z_{ij} - 1; \quad q_{i2} = 2z_{ij} - 1; \quad \rho_{*j} = q_{i1}\, q_{i2}\, \rho[\xi_1; \xi_2], \tag{15}$$

$$P(Z_1 = z_1, Z_2 = z_2 \,|\, \mathbf{w}_1; \mathbf{w}_2) = \Phi_2(q_{i1}\, s_{i1}, q_{i2}\, s_{i2}, \rho_{*j}), \tag{16}$$

$$k_{ij} = q_{ij} s_{ij}, \tag{17}$$

$$s_{ij} = w_{ij} \gamma_j, \tag{18}$$

$$g_{i1} = \phi(k_{i1}) \Phi([k_{i2} - \rho_{*j} k_{i1}] / \sqrt{1 - \rho_{*j}^2}), \tag{19}$$

$$g_{i2} = \phi(k_{i2}) \Phi([k_{i1} - \rho_{*j} k_{i2}] / \sqrt{1 - \rho_{*j}^2}). \tag{20}$$

Separately dividing the terms (19) and (20) by the bivariate cumulative normal density leads to the inverse Mills ratio. In case of a univariate selection process, as in ATE model, there were only two expressions for the inverse Mills ratio. In the present case, there are eight possibilities, that correspond to all combinations of z_{i1} and z_{i2}.

$$\begin{matrix} z_{i1} = 0 \\ z_{i2} = 0 \end{matrix} \Rightarrow \begin{cases} \lambda_{i1} = \phi(-\gamma_1 w_1) \dfrac{\Phi([-\gamma_2 w_2 + \rho\gamma_1 w_1]/\sqrt{1-\rho^2})}{\Phi_2(-\gamma_1 w_1, -\gamma_2 w_2, \rho)}, \\ \lambda_{i2} = \phi(-\gamma_2 w_2) \dfrac{\Phi([-\gamma_1 w_1 + \rho\gamma_2 w_2]/\sqrt{1-\rho^2})}{\Phi_2(-\gamma_1 w_1, -\gamma_2 w_2, \rho)}, \end{cases} \tag{21}$$

$$\begin{matrix} z_{i1} = 1 \\ z_{i2} = 0 \end{matrix} \Rightarrow \begin{cases} \lambda_{i1} = \phi(\gamma_1 w_1) \dfrac{\Phi([-\gamma_2 w_2 + \rho\gamma_1 w_1]/\sqrt{1-\rho^2})}{\Phi_2(\gamma_1 w_1, -\gamma_2 w_2, -\rho)}, \\ \lambda_{i2} = \phi(-\gamma_2 w_2) \dfrac{\Phi([\gamma_1 w_1 - \rho\gamma_2 w_2]/\sqrt{1-\rho^2})}{\Phi_2(\gamma_1 w_1, -\gamma_2 w_2, -\rho)}, \end{cases} \tag{22}$$

$$\begin{matrix} z_{i1} = 0 \\ z_{i2} = 1 \end{matrix} \Rightarrow \begin{cases} \lambda_{i1} = \phi(-\gamma_1 w_1) \dfrac{\Phi([\gamma_2 w_2 - \rho\gamma_1 w_1]/\sqrt{1-\rho^2})}{\Phi_2(-\gamma_1 w_1, \gamma_2 w_2, -\rho)}, \\ \lambda_{i2} = \phi(\gamma_2 w_2) \dfrac{\Phi([-\gamma_1 w_1 + \rho\gamma_2 w_2]/\sqrt{1-\rho^2})}{\Phi_2(-\gamma_1 w_1, \gamma_2 w_2, -\rho)}, \end{cases} \tag{23}$$

$$\begin{matrix} z_{i1} = 1 \\ z_{i2} = 1 \end{matrix} \Rightarrow \begin{cases} \lambda_{i1} = \phi(\gamma_1 w_1) \dfrac{\Phi([\gamma_2 w_2 - \rho\gamma_1 w_1]/\sqrt{1-\rho^2})}{\Phi_2(\gamma_1 w_1, \gamma_2 w_2, \rho)}, \\ \lambda_{i2} = \phi(\gamma_2 w_2) \dfrac{\Phi([\gamma_1 w_1 - \rho\gamma_2 w_2]/\sqrt{1-\rho^2})}{\Phi_2(\gamma_1 w_1, \gamma_2 w_2, \rho)}. \end{cases} \tag{24}$$

Applying mathematics expectation in equation (13) gives the following terms:

$$E[y_i \mid z_{i1} = 0,\ z_{i2} = 0] = \mathbf{x}_i'\boldsymbol{\beta} + \beta_{\lambda 1}\lambda_{i1} + \beta_{\lambda 2}\lambda_{i2}, \tag{25}$$

$$E[y_i \mid z_{i1} = 1,\ z_{i2} = 0] = \delta_1 + \mathbf{x}_i'\boldsymbol{\beta} + \beta_{\lambda 1}\lambda_{i1} + \beta_{\lambda 2}\lambda_{i2}, \tag{26}$$

$$E[y_i \mid z_{i1} = 0,\ z_{i2} = 1] = \delta_2 + \mathbf{x}_i'\boldsymbol{\beta} + \beta_{\lambda 1}\lambda_{i1} + \beta_{\lambda 2}\lambda_{i2}, \tag{27}$$

$$E[y_i \mid z_{i1} = 1,\ z_{i2} = 1] = \delta_1 + \delta_2 + \mathbf{x}_i'\boldsymbol{\beta} + \beta_{\lambda 1}\lambda_{i1} + \beta_{\lambda 2}\lambda_{i2}. \tag{28}$$

Subtracting equations (25) through (28) results in the following interpretations of the bivariate average treatment effect (BATE):

(i) **BATE1.** The effect of both treatments $(z_{i1} = 1,\ z_{i2} = 1)$ against having none of them $(z_{i1} = 0$ and $z_{i2} = 0)$.

(ii) **BATE2.** The effect of only the first on treatment $(z_{i1} = 1,\ z_{i2} = 0)$ against having none of them $(z_{i1} = 0,\ z_{i2} = 0)$.

(iii) **BATE3.** The effect of the second treatment $(z_{i1} = 0, z_{i2} = 1)$ against none of them $(z_{i1} = 0, z_{i2} = 0)$.

(iv) **BATE4.** The effect of both treatments $(z_{i1} = 1,\ z_{i2} = 1)$ against having only the first one $(z_{i1} = 1,\ z_{i2} = 0)$.

(v) **BATE5.** The effect of both treatments $(z_{i1} = 1, z_{i2} = 1)$ against the only second $(z_{i1} = 0, z_{i2} = 1)$.

(vi) **BATE6.** The effect of only the first treatment $(z_{i1} = 1,\ z_{i2} = 0)$ against only the second one $(z_{i1} = 0,\ z_{i2} = 1)$.

(vii) **BATE7.** The effect of only the second treatment $(z_{i1} = 0, z_{i2} = 1)$ against only the first one $(z_{i1} = 1, z_{i2} = 0)$.

$$BATE1 = E[y_i \mid z_{i1} = 1, z_{i2} = 1] - E[y_i \mid z_{i1} = 0, z_{i2} = 0]$$

$$= \delta_1 + \delta_2 + \beta_{\lambda 1}[\lambda_1^{(1,1)} - \lambda_1^{(0,0)}] + \beta_{\lambda 2}[\lambda_2^{(1,1)} - \lambda_2^{(0,0)}], \quad (29)$$

$$BATE2 = E[y_i \mid z_{i1} = 1, z_{i2} = 0] - E[y_i \mid z_{i1} = 0, z_{i2} = 0]$$

$$= \delta_1 + \beta_{\lambda 1}[\lambda_1^{(1,0)} - \lambda_1^{(0,0)}] + \beta_{\lambda 2}[\lambda_2^{(1,0)} - \lambda_2^{(0,0)}], \quad (30)$$

$$BATE3 = E[y_i \mid z_{i1} = 0, z_{i2} = 1] - E[y_i \mid z_{i1} = 0, z_{i2} = 0]$$

$$= \delta_2 + \beta_{\lambda 1}[\lambda_1^{(0,1)} - \lambda_1^{(0,0)}] + \beta_{\lambda 2}[\lambda_2^{(0,1)} - \lambda_2^{(0,0)}], \quad (31)$$

$$BATE4 = E[y_i \mid z_{i1} = 1, z_{i2} = 1] - E[y_i \mid z_{i1} = 1, z_{i2} = 0]$$

$$= \delta_2 + \beta_{\lambda 1}[\lambda_1^{(1,1)} - \lambda_1^{(1,0)}] + \beta_{\lambda 2}[\lambda_2^{(1,1)} - \lambda_2^{(1,0)}], \quad (32)$$

$$BATE5 = E[y_i \mid z_{i1} = 1, z_{i2} = 1] - E[y_i \mid z_{i1} = 0, z_{i2} = 1]$$

$$= \delta_1 + \beta_{\lambda 1}[\lambda_1^{(1,1)} - \lambda_1^{(0,1)}] + \beta_{\lambda 2}[\lambda_2^{(1,1)} - \lambda_2^{(0,1)}], \quad (33)$$

$$BATE6 = E[y_i \mid z_{i1} = 1, z_{i2} = 0] - E[y_i \mid z_{i1} = 0, z_{i2} = 1]$$

$$= \delta_1 - \delta_2 + \beta_{\lambda 1}[\lambda_1^{(1,0)} - \lambda_1^{(0,1)}] + \beta_{\lambda 2}[\lambda_2^{(1,0)} - \lambda_2^{(0,1)}], \quad (34)$$

$$BATE7 = E[y_i \mid z_{i1} = 0, z_{i2} = 1] - E[y_i \mid z_{i1} = 1, z_{i2} = 0]$$

$$= \delta_2 - \delta_1 + \beta_{\lambda 1}[\lambda_1^{(0,1)} - \lambda_1^{(1,0)}] + \beta_{\lambda 2}[\lambda_2^{(0,1)} - \lambda_2^{(1,0)}], \quad (35)$$

where $\lambda_1^{(0,0)}, \lambda_2^{(0,0)}, \lambda_1^{(1,0)}, \lambda_2^{(1,0)}, \lambda_1^{(0,1)}, \lambda_2^{(0,1)}, \lambda_1^{(1,1)}, \lambda_2^{(1,1)}$ denote the bivariate inverses Mills ratios expressions from (21) to (24).

3.2. Robust covariance matrix and Murphy and Topel theorem

The robust covariance matrix is based on Murphy and Topel [5] methodology for two-step estimations in the case of non-linear models.

Consider the regression $y = h(\mathbf{x}, \boldsymbol{\beta}, \mathbf{w}, \boldsymbol{\gamma}) + \varepsilon$ with covariance matrix $\mathbf{V}_b$. In the first step, the parameters of $\boldsymbol{\gamma}$ are estimated with $\mathbf{x}$ as the explanatory variables. In the second step, estimation of the regression of $y = h(\mathbf{x}, \boldsymbol{\beta}, \mathbf{w}, \hat{\boldsymbol{\gamma}})$ is done, where $\hat{\boldsymbol{\gamma}}$ is the first round estimate of γ. The parameters in $\boldsymbol{\gamma}$ have a covariance matrix given by $\mathbf{V}_c$ and assuming asymptotic normality this matrix is unbiased $(\hat{\mathbf{V}}_c)$. The step below gives the second stage covariance matrix estimate: $\hat{\mathbf{V}}_b$. Let $\hat{\boldsymbol{\beta}}$ be an unbiased estimator of $\boldsymbol{\beta}$, with estimated covariance matrix $\hat{\mathbf{V}}_b$. Let $s^2\hat{\mathbf{V}}_b$ be the estimated matrix of $\sigma^2\mathbf{V}_b = \sigma^2(\mathbf{X}^{0'}\mathbf{X}^0)^{-1}$, where $\mathbf{X}^0$ is the matrix of pseudo-regressions evaluated at the values: $\mathbf{x}_i^0 = \partial h(\mathbf{x}, \boldsymbol{\beta}, \mathbf{w}, \hat{\boldsymbol{\gamma}}) / \partial\beta$. Under the necessary conditions assumed for the non-linear least square estimation, the second-step estimation of $\boldsymbol{\beta}$ is consistent and asymptotically normally having covariance matrix given by:

$$\hat{\mathbf{V}}_b = \frac{1}{n}[\sigma^2\mathbf{V}_b + \mathbf{V}_b(\mathbf{C}\mathbf{V}_c\mathbf{C}' - \mathbf{C}\mathbf{V}_c\mathbf{R}' - \mathbf{R}\mathbf{V}_c\mathbf{C}')\mathbf{V}_b], \tag{36}$$

where

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

and

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

The $\mathbf{R}$ matrix depends on the score function of the bivariate probit likelihood.

In the BATE, the first derivatives of the (21) to (24) are necessary in order to get the **C** matrix. In Appendix A, we show some analytics problems in a more detailed manner. Given the trouble calculation for those derivatives, 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. There is no ordinary direct formula for the variance estimates of the second stage model. One way to surpass this problem is to use simulation strategies to investigate the properties from the estimators. A nonparametric bootstrap analysis was used where the main concern was to generate robust standard errors and confidence intervals for the estimates. These methods are weak consistent for the parameters estimates what is sufficient for most statistical problems (Shao and Tu [6]).

To avoid problems in estimation of the standard errors and confidence limits (CLs) in cases where the population could be generated from a highly skewed distribution we used 5.000 replicas[2]. There are a few bootstrap confidence sets in the literature, such as, 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 it is not an easy task to implement them. We chose the hybrid method (Shao and Tu [6]) which has the same accuracy as the traditional normal approximation when a considerable size of resample is used. The exact[3] 1-2α confidence interval for θ of the hybrid bootstrap method is presented below:

$$IC_{(1-2\alpha,\, \hat{\theta}_n)}[\hat{\theta}_n - n^{-\varsigma}H_{Boot}^{-1}(1-\alpha),\, \hat{\theta}_n - n^{-\varsigma}H_{Boot}^{-1}(\alpha)], \tag{37}$$

where ς is a fixed constant, where H_{Boot} comes from

$$H_{Boot}(x) = P_*\{n^{\varsigma}(\hat{\theta}_n^* - \hat{\theta}_n) \leq x\}. \tag{38}$$

4. Application Using PINTEC Data

Three data base were used, on firm level for the year 2008: the Brazilian Annual Survey of Industry (PIA/IBGE-A), the Brazilian Innovation Survey

---

[2]For the SE and CL of the estimates, two SAS macros, %BOOT and %BOOTIC, respectively.
[3]A confidence set is exact when its confidence coefficient is exactly equal to its nominal level.

(PINTEC/IBGE-B) and the Annual Relation of Social Information (RAIS/MTE-C). We describe below the variables.

Product Innovation (*Product*) (B): Dummy variable indicating that the firm has introduced a new product for the firm;

Process Innovation (*Process*) (B): Dummy variable indicating that the firm has introduced a new process for the firm;

National: Dummy variable for no foreign controller capital;

Firm Size (*l*)(A): Total numbers of employees;

Total Revenue: Total revenue;

Labor Productivity (*y*/*l*)(A): Total revenue over number of employees (*l*);

Market Share (*Share*) (A): Total number of employees from the firm over total number of employees of the sector;

R&D (B): Total spending on R&D activities. Includes intra and extramural R&D;

R&D Effort (B): Ratio between R&D and total sales;

Schooling (*Skill*) (C): Weighted average of employees schooling;

Gross fix capital stock (*k*)(A/C): Capital stock measured by perpetual inventory method, according to Alves and Silva [1];

Turnover rate (rot) (C): Rate of employees that leaves the firm on the next year;

Geographic Localization (*locus*) (B): Category for geographic Brazilian regions;

Economic Classification (*ocde*[4]) (B): (i) Extractive Industry; (ii) High Technology; (iii) Medium-High Technology; (iv) Medium-Low Technology; (v) Low Technology; and (vi) Services;

---

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

Cooperation for innovation (*coop*) (B): Category of cooperation for innovation in several levels of partnership.

Age of the firm (*age*) (C): Age in years of the firm.

4.1. Results

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 innovation in order to successfully internalize the economics results.

Considering the correlation between the two of innovative activities we present the bivariate probabilistic models in (39). The second step regression model is represented in equation (40).

$$[prod_{imkl};\ proc_{imkl}] = \Phi[\gamma_0 + \gamma_1 RDeffort_i + \gamma_2 Skill_i + \gamma_3 Share_i + \gamma_4 National_i + \gamma_m coop_{im} + \eta_k locus_{ik} + \mu_l ocde_{il}], \quad (39)$$

$$\ln[\gamma/l_{imkl}] = \eta_0 + \delta_1 prod_{imkl} + \delta_2 proc_{imkl} + \eta_1 \ln(l_{imkl}) + \eta_2 \ln(k_{imkl}) + \eta_3 \ln(AGE_{imkl}) + \beta_{\lambda 1}\lambda^{(1)}_{imkl} + \beta_{\lambda 2}\lambda^{(2)}_{imkl} + \xi_{imkl}. \quad (40)$$

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 estimated correlation was 71.6%, so we do have a bivariate process between these two variables. The average schooling of the labor force had a positive and significant high impact over both the product and process innovation. The same is true for the market concentration.

Table 2. Second-step estimation: linear regression

Variable	Estimate	Boot. SE	*t*-value	*p*-value	Variable	Estimate	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-BATE

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.

In Table 3, we see that the effect of innovating in both product and process is negative relative to the effect of innovating only in process (BATE4 = –3,043). This does not mean that there is any negative effect from the both types of innovation, as shown in BATE2 and BATE3. The negative effect in BATE4 means that, given the innovation process that has already been taken by the firm, it is not worth to do both process and product innovation. This benefit would be possible only if a bivariate treatment analysis (BATE) is used.

5. Conclusions

The bivariate analysis of treatment (BATE) was a great analytic potential compared with the univariate analysis of treatment (ATE), because in the BATE model it is considered a bivariate 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.

Appendix A - Derivatives for the R Matrix

In this appendix, we present some of the derivatives that are necessary to derive the **C** matrix in Subsection 3.2, specifically in the case where $(z_{i1} = 0;\ z_{i2} = 0)$. Considering the inverse Mills ratio in Subsection 3.1 we have:

$$\lambda_{i1}(\gamma_1, \gamma_2) = \phi(-\gamma_1 w_1) = \frac{\Phi\left(\dfrac{-\gamma_2 w_2 + \rho\gamma_1 w_1}{\sqrt{1-\rho^2}}\right)}{\Phi_2(-\gamma_1 w_1, -\gamma_2 w_2, \rho)}.$$

Considering the partial derivatives of $\lambda_{i1}(\gamma_1, \gamma_2)$ in relation to γ_1 we have:

$$\frac{\partial}{\partial\gamma_1}\lambda_{i1} = \frac{\partial}{\partial\gamma_1}\left\{\phi(-\gamma_1 w_1)\times\Phi\left(\frac{-\gamma_2 w_2 + \rho\gamma_1 w_1}{\sqrt{1-\rho^2}}\right)\times\frac{1}{\Phi_2(-\gamma_1 w_1, -\gamma_2 w_2, \rho)}\right\}$$

$$= \frac{\partial}{\partial\gamma_1}\left\{\underbrace{H(\gamma_1)\times G(\gamma_1)}_{Z(\gamma_1)}\times K(\gamma_1)\right\} = Z'(\gamma_1)K(\gamma_1) + Z(\gamma_1)K'(\gamma_1)$$

$$= \{H'(\gamma_1)G(\gamma_1) + H(\gamma_1)G'(\gamma_1)\}K(\gamma_1) + \{H(\gamma_1)G(\gamma_1)\}K'(\gamma_1)$$

$$= H'(\gamma_1)G(\gamma_1)K(\gamma_1) + H(\gamma_1)G'(\gamma_1)K(\gamma_1) + H(\gamma_1)G(\gamma_1)K'(\gamma_1),$$

where $H(\gamma_1)$, $G(\gamma_{i1})$, $K(\gamma_{i1})$, $H'(\gamma_1)$, $G'(\gamma_{i1})$ and $K'(\gamma_{i1})$ are given by:

$$H(\gamma_1) = \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(-\gamma_1 w_1)^2\right],$$

$$H'(\gamma_1) = \frac{-2\gamma_1 w_1^2}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(-\gamma_1 w_1)^2\right],$$

$$G(\gamma_1) = \Phi\left(\frac{-\gamma_2 w_2 + \rho\gamma_1 w_1}{\sqrt{1-\rho^2}}\right),$$

$$G(\gamma_1)' = \frac{\rho w_1}{\sqrt{1-\rho^2}} \times \phi\left(\frac{-\gamma_2 w_2 + \rho\gamma_1 w_1}{\sqrt{1-\rho^2}}\right),$$

$$K(\gamma_1) = [\Phi_2(-\gamma_1 w_1, -\gamma_2 w_2, \rho)]^{-1},$$

$$K'(\gamma_1) = \frac{-\mathbf{w}_{i1}\phi(-\mathbf{w}'_{i1}\gamma_1)\Phi([\mathbf{w}'_{i2}\gamma_2 - \rho\mathbf{w}'_{i1}\gamma_1]/\sqrt{1-\rho^2})}{\Phi_2(-\gamma_1 w_1, -\gamma_2 w_2, \rho)}.$$

References

[1] P. F. Alves and A. M. Silva, Estimativa do Estoque de Capital das Empresas Industriais Brasileiras, Discussion Text 1325, IPEA, Brasília, 2008.

[2] G. De Luca and F. Peracchi, A Sample Selection Model for Unit and Item Nonresponse in Cross-sectional Surveys, CEIS Research Paper 99, 2007. http://ssrn.com/abstract=967391

[3] W. Greene, Econometric Analysis, Prentice Hall, 6th ed., 2007.

[4] J. Heckman, Dummy endogenous variables in a simultaneous equation system, Econometrica 46 (1976), 931-959.

[5] K. M. Murphy and R. H. Topel, Estimation and inference in two-step econometric models, J. Business and Economic Statistics 3 (1985), 370-379.

[6] J. Shao and D. Tu, The Jackknife and Bootstrap, Springer Series in Statistic, 1995.

[7] H. A. Tauchmann, Note on Consistency of Heckman-type Two-step Estimators for the Multivariate Sample-selection Model, RWI: Discussion Paper No. 40, 2006. http://dx.doi.org/10.2139/ssrn.900079