Economics Bulletin

Volume 29, Issue 2

Marginal effects in the double selection regression model: an illustration for the wages of women in Spain

José de Hevia
Department of Economic Analysis, Universidad Rey Juan Carlos

María Arrazola
Department of Economic Analysis, Universidad Rey Juan Carlos

Abstract

In this article we obtain different marginal effects for continuous variables in the context of a double selection regression model, in which it is assumed that the model's disturbances have a normal distribution. Using data of Spanish women, we illustrate these effects by estimating a double selection regression model for the analysis of the economic return from education in the context of the Mincerian wage equation.

We are grateful to Comunidad de Madrid and Universidad Rey Juan Carlos for financial support to carry out this research (project URJC-CM-2007-CSH-1393).
Citation: José de Hevia and María Arrazola, (2009) "Marginal effects in the double selection regression model: an illustration for the wages of women in Spain", *Economics Bulletin*, Vol. 29 no.2 pp. 611-621.
Submitted: Sep 17 2008. **Published:** April 15, 2009.

1. Introduction

Sample selection problems have been extensively dealt with in Econometrics literature, both from a theoretical and an applied perspective, the pioneer works being those of Heckman (1976) and Lee (1976). Those authors have considered a regression model in which there is a single selection rule which determines whether or not the dependent variable of the model is observed. However, different authors have shown the importance of following a double selection approach in the empirical analysis of different phenomena. Catsiapsis and Robinson (1982) apply the case of two selection rules to the problem of estimating the schedule of financial aid (grants and scholarships) faced by US high school graduates contemplating investment in higher education. Fishe, Trost and Lurie (1981) employ a double selection model to study wages, college attendance and labour participation decisions for young women. Krishnan (1990) applies a double self-selection model to an Economics of Moonlighting study. Mohanty (2000) and (2001) demonstrates the interest of double selection in the analysis of male-female wage differences. Pfeiffer, F. and F. Reize (2000) apply a double selection model to compare firm survival and employment growth of business start-ups by unemployed and others. Serumaga-Zake and Naudé (2003) use a double selection scheme to estimate private returns to education in South Africa. Sorensen (1989) uses a bivariate probit selectivity model to estimate an earnings equation and to analyze the crowding hypothesis, according to which women are crowded into “female jobs”, resulting in lower wages for those jobs. Tunali (1986) employs a double selection model to analyze the migratory process of individuals.

In the context of these limited dependent variable models, there are different expected values and different marginal effects linked to those values, which are relevant in the analysis of the endogenous variable of the model (see, Cameron and Trivedi, 2005). The interest in calculating the different marginal effects is that it permits the analysis and comparison of the effects of the explanatory variables on the endogenous variable of the model for different groups of individuals.

For the single selection case there are some empirical works which analyze different marginal effects (see, Arrazola and Hevia, 2008; Hoffman and Kassouf, 2005, Saha et al., 1997). However, surprisingly, the marginal effects in the double selection model have not been tackled in Applied Econometrics literature. For that reason, in the second section of this article, different marginal effects in a double selection model have been obtained. In the third section, using data of Spanish women, we have illustrated these effects by estimating a double selection regression model for the analysis of the economic return to education.

2. Marginal effects in the double selection model

We have considered a general model with a double selection:

$$Y_1^* = X_1\beta_1 + U_1 \quad \begin{cases} Y_1 = 1 & if \quad Y_1^* > 0 \\ Y_1 = 0 & if \quad Y_1^* \le 0 \end{cases} \quad \text{first selection rule}$$

$$Y_2^* = X_2\beta_2 + U_2 \quad \begin{cases} Y_2 = 1 & if \quad Y_2^* > 0 \\ Y_2 = 0 & if \quad Y_2^* \le 0 \end{cases} \quad \text{second selection rule}$$

$$Y_3 = X_3\beta_3 + \sigma_3 U_3 \quad \text{Regression equation}$$

Y_3 is observed only if $Y_1 = Y_2 = 1$

Here X_m´s are 1 x K_m vectors of explanatory variables (m=1,2,3); β_m´s are K_m x 1 vectors of unknown coefficients; σ_3 is an unknown scale parameter and U_m´s are the disturbances with normal distribution, zero mean, and covariance matrix:

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho_{13} \\ \rho & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$$

The regression function can be written as:

$$E\left[Y_3 / X_3, \Gamma\right] = X_3\beta_3 + \sigma_3 E[U_3 / X_3, \Gamma] \tag{1}$$

where the conditioning argument Γ denotes the outcome of the selection rules, or the sample selection regime. Thus**:**

$$\begin{aligned} &\Gamma = \{\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4\} \\ &\Gamma_1 = \{(Y_1 = 1, Y_2 = 1,); (Y_1 = 0, Y_2 = 0); (Y_1 = 1, Y_2 = 0); (Y_1 = 0, Y_2 = 1)\} \\ &\Gamma_2 = \{(Y_1 = 1); (Y_1 = 0)\} \\ &\Gamma_3 = \{(Y_2 = 1); (Y_2 = 0)\} \\ &\Gamma_4 = \{(Y_1 = 0 \; or \; 1, Y_2 = 0 \; or \; 1)\} \end{aligned}$$

From (1) we obtain:

$$\begin{aligned} &E\left[Y_3 / X_3, \Gamma_1\right] = X_3\beta_3 + \sigma_3 E[U_3 / X_3, \Gamma_1] = X_3\beta_3 + \gamma_1\lambda_1 + \gamma_2\lambda_2 \\ &E\left[Y_3 / X_3, \Gamma_2\right] = X_3\beta_3 + \sigma_3 E[U_3 / X_3, \Gamma_2] = X_3\beta_3 + \gamma_1\lambda_3 \\ &E\left[Y_3 / X_3, \Gamma_3\right] = X_3\beta_3 + \sigma_3 E[U_3 / X_3, \Gamma_3] = X_3\beta_3 + \gamma_2\lambda_4 \\ &E\left[Y_3 / X_3, \Gamma_4\right] = E\left[Y_3 / X_3\right] = X_3\beta_3 \end{aligned}$$

in which:

$$\gamma_1 = \sigma_3\rho_{13} \qquad\qquad \gamma_2 = \sigma_3\rho_{23}$$

$$\lambda_1 = (2Y_1 - 1)\frac{\phi(X_1\beta_1)\Phi\left((2Y_2 - 1)\dfrac{X_2\beta_2 - \rho X_1\beta_1}{\sqrt{1-\rho^2}}\right)}{\Phi_J[(2Y_1 - 1)X_1\beta_1, (2Y_2 - 1)X_2\beta_2, (2Y_2 - 1)(2Y_1 - 1)\rho]}$$

$$\lambda_2 = (2Y_2 - 1)\frac{\phi(X_2\beta_2)\Phi\left((2Y_1 - 1)\dfrac{X_1\beta_1 - \rho X_2\beta_2}{\sqrt{1-\rho^2}}\right)}{\Phi_J[(2Y_1 - 1)X_1\beta_1, (2Y_2 - 1)X_2\beta_2, (2Y_2 - 1)(2Y_1 - 1)\rho]}$$

$$\lambda_3 = (2Y_1 - 1)\frac{\phi(X_1\beta_1)}{\Phi[(2Y_1 - 1)X_1\beta_1]}$$

$$\lambda_4 = (2Y_2 - 1)\frac{\phi(X_2\beta_2)}{\Phi[(2Y_2 - 1)X_2\beta_2]}$$

Where $\phi(.)$ and $\Phi(.)$, respectively, denote the standard univariate normal density and distribution functions and $\Phi_J(.)$ denotes the standard bivariate normal distribution function with correlation coefficient ρ.

Since Y_3 is observed, if, and only if, $Y_1=Y_2=1$, the estimation of the regression's parameters can be made starting from the expression (1) applied to that case and using for that purpose a two-step estimation procedure similar to that of Heckman for the case of a single selection rule (see Tunali, 1986).

Furthermore, in the context described, and as happens in models with a single selection rule, it could be of interest to calculate different marginal effects. In a double selection context, the effect of a change in a continuous variable X_j , which appears in X_1, X_2 and X_3, on the value of different conditional expectations could be calculated. Thus, for the general case presented in (1):

$$ME_j = \frac{\partial E[Y_3 / X_3, \Gamma]}{\partial X_j} = \beta_{3j} + \sigma_3 \frac{\partial E[U_3 / X_3, \Gamma]}{\partial X_j} \tag{2}$$

The marginal effect of X_j on Y_3 in the subpopulation determined by Γ consists of two components. There is a direct effect on the mean of Y_3, which is β_{3j}. In addition, if X_j appears in the bivariate probit model, it will influence Y_3 through its presence in the λ´s. In (2), depending on the argument Γ, we obtain different marginal effects:

$$ME_{1j} = \frac{\partial E[Y_3 / X_3, \Gamma_1]}{\partial X_j} = \beta_{3j} + \gamma_1 \frac{\partial \lambda_1}{\partial X_j} + \gamma_2 \frac{\partial \lambda_2}{\partial X_j} \tag{3}$$

$$ME_{2j} = \frac{\partial E[Y_3 / X_3, \Gamma_2]}{\partial X_j} = \beta_{3j} + \gamma_1 \frac{\partial \lambda_3}{\partial X_j} \tag{4}$$

$$ME_{3j} = \frac{\partial E[Y_3 / X_3, \Gamma_3]}{\partial X_j} = \beta_{3j} + \gamma_2 \frac{\partial \lambda_4}{\partial X_j} \tag{5}$$

$$ME_{4j} = \frac{\partial E[Y_3 / X_3, \Gamma_4]}{\partial X_j} = \frac{\partial E[Y_3 / X_3]}{\partial X_j} = \beta_{3j} \tag{6}$$

where

$$\frac{\partial \lambda_1}{\partial X_j} = \frac{(2Y_2 - 1)}{\sqrt{1-\rho^2}}\left(\beta_{2j} - \rho\beta_{1j}\right)\frac{\phi\left(\frac{X_2\beta_2 - \rho X_1\beta_1}{\sqrt{1-\rho^2}}\right)}{\Phi\left((2Y_2 - 1)\frac{X_2\beta_2 - \rho X_1\beta_1}{\sqrt{1-\rho^2}}\right)}\lambda_1 - \lambda_1\left(\beta_{1j}(\lambda_1 + X_1\beta_1) + \lambda_2\beta_{2j}\right)$$

$$\frac{\partial \lambda_2}{\partial X_j} = \frac{(2Y_1 - 1)}{\sqrt{1-\rho^2}}\left(\beta_{1j} - \rho\beta_{2j}\right)\frac{\phi\left(\frac{X_1\beta_1 - \rho X_2\beta_2}{\sqrt{1-\rho^2}}\right)}{\Phi\left((2Y_1 - 1)\frac{X_1\beta_1 - \rho X_2\beta_2}{\sqrt{1-\rho^2}}\right)}\lambda_2 - \lambda_2\left(\beta_{2j}(\lambda_2 + X_2\beta_2) + \lambda_1\beta_{1j}\right)$$

$$\frac{\partial \lambda_3}{\partial X_j} = -\beta_{1j}\lambda_3(X_1\beta_1 + \lambda_3)$$

$$\frac{\partial \lambda_4}{\partial X_j} = -\beta_{2j}\lambda_4(X_2\beta_2 + \lambda_4)$$

The interest of each of these marginal effects depends on the economic context to which the econometric model is applied. In the next section we present an illustrative example of the interest of these effects. Also, depending on whether ρ is zero or not, or on whether or not X_j appears in all the equations, particular expressions for the different marginal effects can be obtained. In most of the applied works in which sample selection models are used, it is usual to only calculate the effects of the explanatory variables on the endogenous variable for the whole population, i.e. ME_4. However, other effects such as ME_1, ME_2 y ME_3, are also of great interest.

3. An illustration for the wages of women in Spain

To illustrate the interest of the marginal effects proposed in the previous section, we have analyzed the return to education in the context of a Mincerian wage equation. For this wage equation, the obtention and interpretation of the marginal effects in a single selection context had already been analyzed (see Arrazola y Hevia, 2008, and Hoffmann and Kassouf, 2005). This article extends that analysis to the double selection context.

Following the work of Mohanty (2000), we consider that the observation of the wage offer is determined by a double selection scheme: individuals make the decision to participate or not in the labour market, but, however, they may or may not find a job. The wage offered to an individual is only observed in the case of the individual participating in the labour market and, also, finding a job. In this context, the double selection scheme can be described by the following equations:

Participation $$Y_1^* = \beta_{1S}S + Z_1\beta_{1'} + U_1 \qquad \begin{cases} Y_1 = 1 & if \quad Y_1^* > 0 \\ Y_1 = 0 & if \quad Y_1^* \le 0 \end{cases}$$

Wage-employment $$Y_2^* = \beta_{2S}S + Z_2\beta_{2'} + U_1 \qquad \begin{cases} Y_2 = 1 & if \quad Y_2^* > 0 \\ Y_2 = 0 & if \quad Y_2^* \le 0 \end{cases}$$

Wage offer equation $$\log w^* = \beta_{3S}S + Z_3\beta_{3'} + U_3$$

w^* is observed if and only if $Y_1 = Y_2 = 1$

With w^*= wage offer, S = years of education, Z_1 = other determinants of the participation, Z_2 = other determinants of the wage-employment status, Z_3= other determinants of the wage offer, β´s are unknown parameters and U_m´s possess the previously indicated properties.

Note that the variable S is present in all three equations of the model.

The literature on wages and economic returns to education has been mainly focussed on answering the question "What are the returns to education for the case of women?", namely "What is the value of β_{3S} ?" (see, for example, Serumaga-Zake and Naudé, 2003). This is a very interesting question. However, by using the previous model and the different marginal effects described, other equally interesting questions can be answered, such as "Are there any differences in the economic returns to education depending on the labour situation of women?" If the answer is yes, "Have working women greater or lesser returns to education than unemployed ones?" and "What happens in the case of inactive women?".

Using data from the enlarged survey of the European Household Panel for Spain in 2000 for women, the wage equation has been estimated by OLS following the Heckman two-step procedure. The selectivity terms were estimated in the first step by a bivariate probit model. Note that, in the labour market, it is impossible to distinguish individuals with $Y_1 =0$ and Y_2 =1 from those with $Y_1 =0$ and Y_2 =0, so a certain degree of partial observability is involved in estimating the bivariate probit model (see Mohanty, 2000). The three equations estimated are presented in Table I columns I-III. The sample is constituted by women aged between 16 and 65 who are not engaged in entrepreneurial activities. Data were taken from 13,147 women, 6,053 of whom participated in the labour market and 4,684 were wage-earners.

The variables included in the bivariate probit model aimed to cover the economic and social factors, which could have an influence on the decision to participate, or not, in the labour market, and, on the possibility of finding a job, or not. The wage equation includes as regressors: education, experience and its square, the selectivity terms obtained from the bivariate probit model, and a group of variables which indicate the individuals' region of residence. With respect to the identification of the model, note that more than the two exclusion restrictions which are necessary in order to break the functional form dependence have been incorporated.

From the results obtained in the estimation of the bivariate probit model (see Table I, columns I and II), it should be noted that ρ was statistically significant at the usual levels, underlining the importance of following a bivariate probit model in the analysis of the observation of the wage offer, and that the decision as to whether to participate in the labour market is not independent from the situation of finding a job or not. Additionally, the selection terms were statistically significant in the wage equation, which can be interpreted as being evidence of the existence of sample selection. It should also be pointed out that the variable "Education" was statistically significant in the three equations, with a positive effect on the probability of being active, of finding work and on the wage offered. In the previous context, it could be of interest to analyze the effect of education on the wage offer, i.e. the economic return to education. In this respect, at least four relevant effects linked to four different subpopulations can be found:

a) Economic return to education for employed women, i.e. when Y_1 =1 and Y_2 =1 (see equation (3)).

b) Economic return to education for unemployed women, i.e. when Y_1 =1 and Y_2 =0 (see equation (3))

c) Economic return to education for inactive women, i.e. when Y_1 =0 (see equation (4)).

d) Economic return to education for women (see equation (6)).

This last effect measures the mean effect of a variation in the education on wage offers for women regardless of their status in the labour market. The other three effects measure the mean effect of a variation in education on wage offers for different relevant subsamples from the population: employed, unemployed and inactive women. That disaggregation permits us to study to what extent the economic return to education varies in terms of the status of individuals in the labour market. As indicated previously, the literature on returns to education has concentrated almost exclusively on studying the last effect above (d), with the limitation that this signifies (see Arrazola and Hevia, 2008, for more details).

In another direction, and to highlight the interest of the results obtained in this article compared to those presented in Arrazola and Hevia (2008) and Hoffmann and Kassouf, (2005), it is important to point out that in a single selection context, the second and third effects could not be obtained separately since the model did not permit us to differentiate the case of the unemployed women from the case of the inactive women. Also, from an empirical point of view, it is important to analyze to what extent the model with a double selection is superior or not to the model with a single selection. In this respect, and following Mohanty (2001), Table I presents the results of the J-statistics for non-nested models which suggests that the specification with double selection cannot be rejected at all conventional levels of significance and that the wage equation with single selection (shown in Table I columns IV and V) is not appropriate in our case.

Table II shows the estimates of the marginal effects both for the double selection and single selection cases calculated from the results presented in Table I. Since *S* is part of the selection equations, the marginal effects vary as a function of the individuals' characteristics, so that in Table II the sample mean and the standard error for that mean are presented, except for the case of the economic return for women in which this parameter is obtained directly from the estimations of the wage equation presented in Table I.

For the double selection case, the results show that there were significant differences in the return to education between employed and unemployed or inactive women. However, there did not seem to be any large differences in the return to education between the unemployed and the inactive women. The results show that the employed status significantly increases the economic return to education. The usual empirical analyses on the economic return to education only pay attention to the return for women (7.7%), regardless of their status in the labour market, with the limitation in the analysis that this signifies.

In the single selection case, the return to education for employed women was lower than that of unemployed or inactive women, just the opposite of what happened for the double selection case. This contradiction shows the importance of estimating the marginal effects considering a double selection model if, as is the case, the latter is not rejected.

References

Arrazola, M. and J. de Hevia (2008) “Three measures of returns to education: An illustration for the case of Spain”, *Economics of Education Review*, **27**, 266-275.

Catsiapis, G. and C. Robinson (1982) “Sample selection bias with multiple selection rules”. *Journal of Econometrics* **18**, 351-368.

Cameron, A.C. and P.K. Trivedi (2005) *Microeconometrics. Methods and Applications*. Cambridge University Press, New York.

Fishe, R.P.H., P. Trost and P. Lurie (1981) “Labor Force earnings and college choice of young women: An examination of selectivity bias and comparative advantage”, *Economics of Education Review* **1**, 169-91.

Heckman, J. (1976) “The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simpler Estimator for Such Models”, *Annals of Economic and Social Measurement* **5**, 475-92.

Hoffmann R. and A.L. Kassouf (2005) “Deriving conditional and unconditional marginal effects in log earnings equations estimated by Heckman´s procedure”. *Applied Economics,* **37**, 1303-1311.

Krishnan, P. (1990) “The Economics of Moonlighting: A Double Selection Model”, *Review of Economics and Statistics* **72**, 361-67.

Lee, L.F. (1976) *Estimation of Limited Dependent Variable Models by Two Stage Methods,* Ph. D. Thesis, Department of Economics, University of Rochester.

Mohanty, M.S. (2000) “Determination of participation decision, hiring decision, and wages in a double selection framework: male-female wage differentials in the U.S. labour market revisited”, *Contemporary Economic Policy* **19(2)**, 197-212.

Mohanty, M.S. (2001) “Testing for the specification of the wage equation: double selection approach or single selection approach”, *Applied Economics*, **8**, 525-529.

Pfeiffer, F. and F. Reize (2000) “Business start-ups by the unemployed: an econometric analysis based on firm data”, *Labour Economics,***7**, 629-663.

Saha A., O. Capps and P.J. Byrne (1997) “Calculing marginal effects in models for zero expenditures in household budgets using a Heckman-type correction”, *Applied Economics,* **29**, 1311-1316.

Serumaga-Zake P.A.E. and W.A. Naudé (2003) “Private rates of return to education of Africans in South Africa for 1995: a Double Hurdle model”, *Development Southern Africa*, **20**(4), 515-28.

Sorensen, E. (1989) “Measuring the Pay Disparity between Typically Female Occupations and Other Jobs: A Bivariate Selectivity Approach”, *Industrial and Labour Relations Review* **42**, 624-39.

Tunali, L. (1986) “A General Structure for Models of Double Selection and an Application to a Joint Migration/Earning Process with Remigration”, *Research in Labour Economics* **8**, 235-82.

Table I. Probit results and wage equation with selection

	Double selection			Single selection	
	Bivariate Probit			Univariate Probit	
	Participation (I)	Wage-employment (II)	Wage equation (*) (III)	Wage Observation (IV)	Wage equation (*) (V)
Constant	-4.623 (0.133)	-2.475 (0.405)	5.593 (0.028)	-4.940 (0.144)	5.347 (0.050)
Education	0.084 (0.003)	0.052 (0.007)	0.077 (0.002)	0.089 (0.004)	0.083 (0.002)
Experience	-	-	0.029 (0.002)	-	0.033 (0.002)
Experience ^{2}	-	-	-0.0005 (0.00004)	-	-0.0005 (0.00004)
Age	0.257 (0.008)	0.163 (0.017)	-	0.231 (0.008)	-
Age^2	-0.003 (0.0001)	-0.002 (0.0002)	-	-0.003 (0.0001)	-
Caring duties	-0.187 (0.030)	-0.230 (0.049)	-	-0.189 (0.031)	-
Income of the rest of the household	-0.73×10^{-7} (0.40×10^{-8})	-	-	-0.42×10^{-7} (0.61×10^{-8})	-
Number of children	-0.138 (0.014)	-	-	-0.097 (0.016)	-
Marital status	-0.354 (0.033)	-	-	-0.296 (0.033)	-
Unemployment situation in the past five years	-	-0.932 (0.056)	-	0.236 (0.026)	-
Bivariate selectivity term (participation)	-	-	0.224 (0.016)	-	-
Bivariate selectivity term (wage-employment)	-	-	-0.224 (0.016)	-	-
Univariate selectivity term					0.181 (0.025)
χ^2_6 **variables of region of residence**	88.07	106.73	66.78	196.36	99.45
Disturbance correlation (ρ)	0.415 (0.104)				
Log likelihood function	-10,079.81		-2,000.52	-7,044.19	-2,090.19
J-statistic (p-value)					
H_0: Single selection H_1: Double selection			0.000		
H_0: Double selection H_1: Single selection			0.550		
N	13,147		4,684	13,147	4,684

Note: Standard errors in parentheses. (*) Dependent variable: log(hourly wage)

Table II. Marginal Effects

	Marginal effects	
	Mean	**Standard error**
Double selection		
Employed women	8.020%	0.0280
Unemployed women	6.959%	0.041
Inactive women	6.733%	0.038
Women	7.700%	
Single selection		
Employed women	6.130%	0.008
Unemployed or Inactive women	9.270%	0.018
Women	8.300%	