THE WAGE RATE ESTIMATION USING THE HECKMAN PROCEDURE

Ana Lúcia Kassouf*

Resumo

Este trabalho compara o procedimento de Heckman com um método tradicional de estimação de salários. As estimativas de salário podem ser usadas como custo de oportunidade do tempo dos indivíduos. Uma vez que grande parte dos adultos numa amostra não participam do mercado de trabalho, a simples estimação utilizando-se da técnica de mínimos quadrados ordinários produziria estimadores tendenciosos. Equações de participação no mercado de trabalho e de determinação de salários foram estimadas para mulheres e homens de 16 a 71 anos de idade. Quase todos os coeficientes foram significativos ao nível de 10% e apresentaram os sinais esperados. Quando o método tradicional foi utilizado para estimar o logaritmo do salário hora, observou-se uma tendenciosidade positiva (em valor absoluto) nas estimativas dos parâmetros, no caso da amostra de trabalhadores do sexo masculino e uma tendenciosidade negativa (em valor absoluto) para trabalhadores do sexo feminino.

Abstract

This paper estimates wage rates using the Heckman Procedure and a traditional method. The estimation becomes necessary when the analysis requires a measure of the individual's opportunity cost of time. Since a large fraction of the adults in the sample do not participate in the labor market, the simple estimation by ordinary least squares would cause bias in the coefficients. Labor force participation and the wage determination equations were estimated for men and women from 16 to 71 years old. Most of the coefficients were statistically significant at 10% level and presented the expected signs. It was found an upward biased (in absolute value) estimate for male workers and a downward biased (in absolute value) estimate for the female workers when using the traditional method of estimation.

1. Introduction.

Wage rates can be interpreted as the opportunity cost of time and can be used as determinants of changes in habits. More time-saving devices are used as the opportunity cost of time increases.

*Assistant professor in the Department of Economics and Rural Sociology at ESALQ, University of São Paulo.

The author would like to thank two anonymous reviewers and prof. Rodolfo Hoffmann for their helpful comments and suggestions.

The use of microwave ovens, the choice of having fewer children, the substitution from time-intensive traditional food, such as rice, to time-saving food, such as commercially supplied bread [Senauer, Sahn & Alderman (1986)], are examples of this behavior.

From the Becker (1965) model framework it is possible to observe that an increase in the wage rate raises the relative price of time-intensive activities and induces substitution against it. Kassouf (1994) used estimated wage rates as a measure of the mother's and father's opportunity cost of time in a child's health demand function. As the value of time of the mother, for example, increases, less time is devoted to child care, decreasing the child's health and nutrition. On the other hand, as the wage increases the income of the household increases, which has a positive effect on child's health.

Since a large number of individuals do not participate in the labor market, it becomes essential to estimate the wage rates as a measure of the value of time for those that are not employed based on data from those that are employed. Heckman (1974, 1980) developed a procedure to estimate the wage rates avoiding sample selectivity bias.

The objective of this paper is to present the econometric advantages of using the Heckman procedure to estimate wage rates. Also, to derive the opportunity cost of time by estimating labor force participation and wage determination equations as an empirical application of the Heckman procedure, and to compare the results with a traditional method of estimation.

Section 2 presents a methodology developed by Heckman to estimate the wage rates avoiding sample selectivity bias. Section 3 discusses the data and presents descriptive statistics for the variables employed in the study. Section 4 shows empirical results of the Heckman procedure and the traditional method applied to individual data. Section 5 presents the conclusions.

2. Econometric method.

The wage rates can be used to represent the opportunity cost of time. Usually not all the individuals in a survey are employed in the labor market and, therefore, estimated wage rates are calculated.

The simple estimation of the wage determination equation by ordinary least squares would cause bias in the coefficients due to the sample selectivity which exists when only individuals employed enter into the calculations. To see this problem, assume that L^* is a variable representing labor force participation, then we can write:

$$L_i^* = \gamma' Z_i + u_i$$

where Z_i is a vector of variables that determines participation in the job market.

Note that L* is not observed but we can observe L, such that,

$$\begin{aligned} L_i &= 1 \quad \text{if} \quad L_i^* > 0, \\ L_i &= 0 \quad \text{if} \quad L_i^* \leq 0 \end{aligned}$$

Let W represent the wage rates,

$$W_i = \beta' R_i + v_i$$

where R_i is a vector of variables that determines the wage rates.

We will observe W only when L^* is greater than zero. Assuming that u_i and v_i have a bivariate normal distribution with zero means, standard deviations σ_u and σ_v, and correlation ρ, then

$$\begin{aligned} E(W_i|W_i \text{ is observed}) &= E(W_i|L_i^* > 0) = E(W_i|u_i > -\gamma' Z_i) \\ &= \beta' R_i + E(v_i|u_i > -\gamma' Z_i) \\ &= \beta' R_i + \rho\sigma_v\lambda_i(\alpha_u) \qquad (1) \end{aligned}$$

where,

$$\lambda_i(\alpha_u) = \frac{\phi(\frac{\gamma' Z_i}{\sigma_u})}{\Phi(\frac{\gamma' Z_i}{\sigma_u})}$$

and ϕ and Φ are respectively, the probability density function and the cumulative distribution function for a normal distribution. The function $\lambda(\alpha_u)$ is called in the statistical literature, the inverse Mill's ratio [see, for example, Greene (1993)].

Therefore, from equation (1), the regression can be written as

$$W_i | L_i^* > 0 = \beta' R_i + \beta_\lambda \lambda_i(\alpha_u) + \varepsilon_i \tag{2}$$

Observing equation (2) it is easy to see that if we regress the wage rates on R by ordinary least squares, only for individuals that are working, the estimates of β will be inconsistent because we are omitting the variable $\lambda(\alpha_u)$.

However, ordinary least squares regression of the wages on R and λ, using data only for the individuals who are working, produces consistent estimates of β.

Based on this fact Heckman (1974 & 1980) proposed the following procedure: First, regress L on Z by maximum likelihood using a probit model, where L will take either the value 0 if the individual is not in the labor market or the value 1 if he or she has earnings from labor force employment. With the estimates of γ resulting from this model, it is possible to calculate the estimate of λ. Having the estimate of λ, one can regress W on R and $\hat{\lambda}$ using ordinary least squares and obtain consistent estimates of β. However, even if λ were observed, least squares would be inefficient because the disturbance, ϵ, is heteroskedastic since,

$$\text{var}(v | u > -\gamma' Z_i) = \sigma_v^2 [1 - \rho^2 \delta(\alpha_u)]$$

where,

$$\delta(\alpha_u) = \lambda(\alpha_u)(\lambda(\alpha_u) - \alpha_u)$$

and

$$\alpha_u = \frac{-\gamma' Z_i}{\sigma_u}$$

In the presence of heteroskedasticity, standard errors are incorrectly estimated by ordinary least squares and therefore, the t-statistics are no longer accurate.

White (1980) proposed a very simple method to estimate the correct covariance matrix in the presence of heteroskedasticity with the advantage that the form of the heteroskedasticity does not have to be known. The covariance matrix of the OLS estimator b for the model,

$$Y = X\beta + e$$

is given by

$$(X'X)^{-1}X'\Omega X(X'X)^{-1}$$

where

$$E(ee') = \Omega$$

Observe that the above covariance matrix reduces to $(X'X)^{-1}\sigma^2$ when the error term is assumed to have a constant variance.

White showed that a consistent estimate of the covariance matrix can be found using the least square residual, and is equal to,

$$(X'X)^{-1}X'\widehat{\Omega}X(X'X)^{-1} \tag{3}$$

where

$$\widehat{\Omega} = \text{diag}(\hat{e}_1^2, \hat{e}_2^2, \dots, \hat{e}_n^2)$$

The covariance matrix (3) is called the White Covariance Matrix and it is used to calculate the variance of the least square estimator in the presence of heteroskedasticity. Inferences based on the results of least squares are now possible.

3. Data.

The data set used to conduct this study is the 1989 National Health and Nutrition Survey, undertaken by the Brazilian Geographical and Statistical Institute (IBGE), Institute of Social Economic Planning (IPEA) and by the National Institute of Food and Nutrition (INAN).

The data were collected between July and September of 1989. Approximately 63,000 individuals were interviewed from 17,920 households. The sample selection was done primarily by defining the Northern, Northeastern, Southern, Southeastern and Central regions as well as the urban and rural areas to be represented by the survey. Independent samples were taken from each place resulting in 9 strata, since the North rural was excluded. On average, each stratum has information about 54 sectors, 1991 households and 7024 individuals. All the 486 (54 × 9) sectors, which are sample clusters, were selected with probabilities proportional to the number of households in 1980. In a second stage, 30 households on average, were selected from each sector with probability proportional to the number of households in the sector but inverse to the probability of the sector selection.

The data set provides information on monthly salary (in U.S. dollars) earned by individuals participating in the labor force a week before the interview, and any sort of payment in-kind received by the individual per month, which was already transformed into dollars. These variables were added to get the monthly wage rates. Information on the number of hours worked per week is also available. This variable was multiplied by 4 to obtain the number of hours worked per month. The monthly wage rates were divided by the number of hours worked per month to get the hourly wage rates.

The sample used in this study is composed by workers, male and female, age 16 to 71, from households with at least one child up to age 5. It excludes the Northern region.

The mean and standard deviation for each variable is given in table 1 for male and female that were participating in the labor force.

4. Results.

The wage rate estimation was done separately for men and women from age 16 to 71. This age range was chosen as a way to avoid large discrepancies, since very young workers received a very low wage and very old persons were already retired or also received low salaries. The data showed that 86.7 percent of the men and 33.0 percent of the women had earnings from labor force employment in the period of the survey. Therefore, a value of time for those that were not employed had to be estimated from a wage determination equation that was estimated for those who were employed.

4.1. Male wage rate estimation.

Heckman's procedure was applied to estimate the wage rates. First a probit equation was estimated for a sample of 6083 men from age 16 to 71. The dependent variable takes the value one if the men

Table 1.

Description of the variables, means and standard deviations, for female and male participants in the labor market

Variable	Description	Mean	S.D.
Male			
AGE	individual's age in years	33.13	11.60
EUCATION	individual's years of school	4.06	3.84
NUMBCHILD	number of children in the household	1.52	0.76
HEAD	=1 if individual is the head of the household	0.75	0.44
SON	=1 if individual is son in the household	0.18	0.38
RELATIVE	=1 if individual is a relative in the household	0.06	0.24
OTHER	=1 if individual is an employee or other	0.01	0.10
NEAST	=1 if individual resides in the Northeast region	0.29	0.46
CSES	=1 if the individual resides in the Central, Southeast or South Regions	0.70	0.46
WHITE	=1 if the individual is white	0.47	0.50
BLACK	=1 if the individual is black	0.06	0.23
ASIAN	=1 if the individual is Asian	0.002	0.05
PARDA	=1 if the individual is pardo	0.47	0.50
URBAN	=1 if individual resides in the urban area	0.47	0.50
RURAL	=1 if individual resides in the rural area	0.53	0.50
NLINCOME	individual's nonlabor income	12.19	155.87
WAGE	individual's wage in dollars per hour	1.13	2.20
LWAGE	logarithm of the WAGE variable	-0.49	1.00
Female			
AGE	individual's age in years	30.64	11.10
EUCATION	individual's years of school	4.47	3.87
NUMBCHILD	number of children in the household	1.52	0.77
HEAD	=1 if individual is the head of the household	0.06	0.23
WIFE	=1 if individual is the wife in the household	0.68	0.46
DAUGHTER	=1 if individual is daughter in the household	0.16	0.37
RELATIVE	=1 if individual is a relative in the household	0.07	0.26
OTHER	=1 if individual is a maid or other	0.02	0.14
NEAST	=1 if individual resides in the Northeast region	0.31	0.46
CSES	=1 if the individual resides in the Central, Southeast or South region	0.69	0.46
WHITE	=1 if the individual is white	0.48	0.50
BLACK	=1 if the individual is black	0.05	0.22
ASIAN	=1 if the individual is Asian	0.003	0.06
PARDA	=1 if the individual is pardo	0.47	0.50
URBAN	=1 if individual resides in the urban area	0.51	0.50
RURAL	=1 if individual resides in the rural area	0.49	0.50
NLINCOME	individual's non-labor income	4.64	36.34
WAGE	individual's wage in dollars per hour	0.83	1.65
LWAGE	logarithm of the WAGE variable	-0.84	1.07

had earnings from employment (5273 of them) and 0 otherwise (810 of them).

The variables assumed to affect the individuals' decision on participating in the labor force are age in years (AGE), age squared (AGE2), education in years (EDUCATION), education squared (EDUCATION2), race (WHITE, BLACK, PARDA and ASIAN, which is omitted), sector (URBAN and RURAL, which is omitted), the position of the individual in the household (HEAD, SON, RELATIVE, which is omitted, and OTHER), non-labor income (NLINCOME, which is the sum of rents, pensions and others), the number of children less than 6 years old (NUMBCHILD) and region (CSES, which includes Central, Southeast and South regions, and NORTHEAST, which is omitted). The regions were divided in two groups because the differences between Central, Southeastern and Southern regions were not significant in terms of earnings. However, there were significant differences between these regions and the Northeast, which is a poor region in Brazil.

The results of the probit equation for men are given in table 2. A likelihood ratio test (LRT), which tests if all the slope coefficients in the probit equation are zero, is highly significant and equal to 873.66. The critical value from the chi-squared table with 14 degrees of freedom at a 1% level of significance is 29.14. The model predicts 86.8 percent of the observations correctly, *i.e.*, 5280 observations out of 6083 that were predicted by the model coincided with the actual observed data. Inside the tables this measure was called "Prediction".

The significant coefficients in table 2 present the expected signs. The results indicate a parabolic relation for age showing that as a person gets older, more experience is acquired and more job opportunities appear until it reaches a point after which the older a person gets the less likely he is to participate in the labor force. This is shown by the positive coefficient for AGE and the negative coefficient for AGE2, respectively. The parabolic curve peaks between ages 36 and 37 which is midlife. Education had the expected positive coefficient indicating that more years of schooling increase the possibility of being employed. The quadratic term indicates that as years of school increase, the labor opportunities increase more than proportional to the increase in education. The head of the house-

Table 2.
Results of the probit analysis on labor force activity of men, ages 16-71.

Variable	Coefficient
Constant	-0.90
	(-1.30)
AGE	0.11
	(9.77)***
AGE2	-0.0015
	(-11.64)***
EDUCATION	0.033
	(4.23)***
EDUCATION2	0.000033
	(4.26)***
NUMBCHILD	0.030
	(1.00)
HEAD	0.62
	(7.34)***
SON	-0.34
	(-3.83)***
OTHER	0.43
	(1.90)*
CSES	0.29
	(5.46)***
WHITE	-0.26
	(-0.39)
BLACK	-0.0027
	(-0.00)
PARDA	-0.23
	(-0.35)
URBAN	-0.035
	(-0.71)
NLINCOME	-0.00004
	(-3.54)***
LRT	873.66***
Prediction	86.8%
Observ.	6083

The t-statistics are given in parentheses below coefficients.
*denotes significance at the 10% level.
**denotes significance at the 5% level.
***denotes significance at the 1% level.

hold is more likely to be working outside the house than the son of the head. The Central, Southeastern and Southern regions are more developed than the Northeastern region and therefore offer more job opportunities for individuals. Finally, men are less likely to be employed the higher is the income from non-labor sources such as rent or retirement income (NLINCOME).

On the basis of the coefficients of the probit equation given in table 2, the inverse Mill's ratio (LAMBDA) was calculated and used as a variable in the wage equation.

The dependent variable of the wage equation is the natural logarithm of the hourly wage rate for those men employed in the labor force. Hourly wages in contrast to monthly salaries are more homogeneous measures. The choice of transforming the hourly wages into logs was done based on the Box and Cox approach described in Weisberg (1985), which consists of calculating T^{ξ} for different values of ξ in the range $-2, +2$, where

$$T_i^{\xi} = \frac{W_i^{\xi} - 1}{\xi[GM(W)]^{\xi} - 1} \qquad \text{for} \qquad \xi \neq 0$$

or

$$T_i^{\xi} = GM(W)\ln(W_i) \qquad \text{for} \qquad \xi = 0$$

and W is the hourly wage rate and GM is the geometric mean.

The minimum residual sum of squares from the regression of T^{ξ} on the predictors, by ordinary least squares, for different values of ξ, indicates the best transformation of the hourly wage variable. A value $\xi = 0$ indicates a log transformation, if $\xi = 1$ the data do not need to be transformed and other values of ξ indicate a power transformation.

The results showed a value of ξ equal to zero as the one minimizing the residual sum of squares from the regression of T^{ξ} on the predictors (the predictors are the same as in table 3) indicating, therefore, that the hourly wage had to be transformed in the natural logarithm of the hourly wage.

The results of the wage equation estimated by ordinary least squares for the 5273 men participating in the labor force, using log of hourly wage as the dependent variable, are given in table 3.

Table 3.

Results of the wage equation for men with labor force earnings, age 16-71, using the Heckman procedure and the traditional method.

Variable	(1) Heckman Procedure	(2) Traditional Method
Constant	-2.57	-3.24
	(-8.27)***	(-14.33)***
AGE	0.054	0.085
	(5.35)***	(16.49)***
AGE2	-0.00054	-0.00094
	(-4.14)***	(-13.95)***
EDUCATION	0.072	0.077
	(8.50)***	(9.30)***
EDUCATION2	0.0035	0.0034
	(5.88)***	(5.73)***
URBAN	0.36	0.35
	(15.23)***	(15.19)***
CSES	0.26	0.30
	(9.16)***	(11.34)***
WHITE	0.31	0.30
	(1.36)	(1.46)
BLACK	0.049	0.060
	(0.21)	(0.29)
PARDA	0.20	0.19
	(0.88)	(0.94)
LAMBDA	-0.44	.
	(-3.82)***	.
Wald-test	3904.36 ***	.
F-test	.	437.97 ***
R^2	0.43	0.43
Observ.	5273	5273

The t-statistics are given in parentheses below coefficients.
*denotes significance at the 10% level.
**denotes significance at the 5% level.
***denotes significance at the 1% level.

The coefficient of determination (R^2) for the wage equation is .43 which is very high for a wage equation estimated with cross sectional data [Senauer, Sahn & Alderman (1986)].

The standard errors and t-tests presented in the first column of table 3 were recalculated using a covariance matrix estimator that is consistent in the presence of heteroskedasticity according to White (1980). However, no major differences were found between the standard errors estimated from the White covariance matrix and those estimated assuming that the errors have a constant variance. A Wald test was performed to test if the coefficients were statistically different from zero, since the F test could no longer be carried out as it relies on homoskedasticity. The value of the Wald test is 3904.36. The statistic is asymptotically distributed as a chi-squared with 10 degrees of freedom, and it is highly significant.

Column 2 of table 3 gives the coefficients obtained by not using the Heckman procedure, *i.e.*, by just regressing the logarithm of hourly wage, by ordinary least squares, on the right-hand-side variables, excluding lambda. Observe that practically all the significant coefficients in column 2 are larger in absolute value than the coefficients in column 1 (except EDUCATION2 and URBAN, whose coefficients are about the same). Since the main concern in this paper is to estimate the wages for the whole sample, it is interesting to compare the average of the estimated log wage hour using the Heckman procedure and the traditional method, which are, respectively, -0.45 and -0.55. The coefficient on inverse Mill's ratio variable (LAMBDA) is statistically significant, indicating that its inclusion was necessary to avoid a sample selection bias. Its negative sign indicates that unmeasured factors that increase the probability of participation decrease the wage rates. These facts lead us to conclude that the use of traditional methods of estimating the logarithm of hourly wage rates for those employed or not in the labor force, based only on a sample of working men give, in absolute value, an upward biased estimate of the true effect.

The significant coefficients from the wage equation (column 1 of table 3) have the expected signs. Observe that the AGE coefficient is positive and the AGE2 coefficient is negative as suggested by human capital theory, which states that earnings follow a parabolic curve, peaking somewhere in midlife, due to a depreciation of the workers' human capital in the form of taking more time to perform tasks, for example, as they age. The parabolic curve peaks at age 50 which is beyond midlife.

Pastore & Lopes (1973) analyzed the means of the monthly salaries for different age groups to observe the pattern followed in each specialization level in São Paulo industry. They concluded that for the nonspecialized labor, the salary peaks between ages 31 and 40, while for the medium and superior levels it peaks between 51 and 59 years.

Langoni (1972) found that, in Brazil, income increases with age, reaching a maximum point at 50-54 years in 1960 and 1969. Hoffmann (1993) using a sample of 1025 workers in the agricultural sector of the state of São Paulo extracted from PNAD 1985 observed that higher salaries are associated to workers from 40 to 49 years of age.

The results in table 3 also indicate that individuals with higher education, living in more developed regions and in the urban sector will get higher hourly wage rates.

Senna (1976) using a sample of 265,169 urban Brazilian male workers in 1970, regressed the logarithm of hourly wage on the number of years in school. He found the average rate of return in education, which is the slope coefficient, equals to 12.5%. Moreover, he observed that school explains 34% of the total wage variance. Pastore & Lopes (1973) observed that the mean monthly salary increases as the education level increases. An employee that has 8 years of education would receive 5 times more than the initial salary, high school level 5.4 and college level 7.5 times more. Hoffmann (1993) also found higher levels of education being associated to higher salaries.

4.2. Female wage rate estimation.

A similar procedure was used to estimate the wage rate for each woman, aged 16-71. All the variables assumed to affect the men's choice on participating in the labor force are also employed for women, except for the variables representing the relation of the woman in the household which are: the woman head (HEAD), the wife of the male head (WIFE), the daughter (DAUGHTER), another relative (RELATIVE), which was omitted, and other relations

(OTHER). Also, the estimated wage rate of the man, who was the head of the household (WAGEHEAD) was assumed to affect the choice of the woman between working outside the house or not. It is believed, however, that the wage of the primary woman does not affect the man's employment decision, and therefore, the estimated wage of the primary woman was not included in the men's labor participation probit equation.

Table 4 presents the results of the probit model where the dependent variable is either 1 or 0 depending on whether the woman participated in the labor force or not. There are 6633 women in the sample and only 2187 were employed in the labor force.

The results from table 4 are similar to the men's results. Observe that for women, the coefficient of the variable number of children less than 6 years old (NUMBCHILD) is negative, as expected, and highly significant. This fact is related to the reservation wage, which is the amount of extra earnings required by an individual (who is not working), to give up one unit of leisure. Thus, women with young children at home are likely to have higher reservation wages, and therefore, lower labor force participation rates than women with no children. Also, the woman head of the household is more likely to be employed than the wife of the male head. The peak of the parabolic curve for age occurs at 39 years. Observe that the opportunities for men to participate in the job market decreases earlier than for women.

Sedlacek & Santos (1991) studied the wives participation in the labor market using PNAD 1984 for the metropolitan regions of São Paulo, Rio de Janeiro and Recife. They used a probit equation with variables representing number of children, age, school and husband's income as exogenous variables. They found, as in the present study, that the higher is the women's education, the higher is the probability that they participate in the labor force. Moreover, age also presented a parabolic relation with a peak between 25 and 29 years. In São Paulo, the result was different showing a continuous decrease in the job participation as the woman gets older. Besides, they concluded that, the younger the children and the higher the number of kids, the lower is the mother labor participation.

Table 4.
Results of the probit analysis on labor force activity of women, ages 16-71.

Variable	Coefficient
Constant	-2.50
	(−7.35)***
AGE	0.11
	(12.09)***
AGE2	−0.0014
	(-11.47)***
EDUCATION	0.063
	(11.31)***
EDUCATION2	0.000064
	(11.38)***
NUMBCHILD	-0.064
	(-2.78)***
HEAD	0.56
	(5.82)***
WIFE	-0.30
	(-4.38)***
DAUGHTER	0.29
	(3.87)***
OTHER	1.49
	(10.38)***
CSES	0.012
	(0.28)
WHITE	-0.26
	(-0.87)
BLACK	-0.057
	(-0.19)
PARDA	-0.19
	(-0.65)
URBAN	0.37
	(9.02)***
NLINCOME	-0.00098
	(-2.24)**
WAGEHEAD	-0.042
	(-1.05)
LRT	995.34***
Prediction	72.6%
Oberv.	6633

The *t*-statistics are given in parentheses below coefficients.
*denotes significance at the 10% level.
**denotes significance at the 5% level.
***denotes significance at the 1% level.

Table 5.

Results of the wage equation for women with labor force earnings, age 16-71, using the Heckman procedure and the traditional method.

Variable	(1) Heckman Procedure	(2) Traditional Method
Constant	-4.05	-3.44
	(-10.59)***	(10.73)***
AGE	0.086	0.077
	(7.87)***	(8.78) ***
AGE2	-0.00094	-0.00083
	(-5.86)***	(-6.78) ***
EDUCATION	0.085	0.070
	(5.82)***	(5.05)***
EDUCATION2	0.0042	0.0044
	(4.35)***	(4.75)***
URBAN	0.29	0.17
	(6.25)***	(4.24)***
CSES	0.45	0.45
	(10.78) ***	(11.08)***
WHITE	0.030	0.089
	(0.09)	(0.32)
BLACK	-0.13	-0.14
	(-0.40)	(-0.49)
PARDA	-0.026	0.0038
	(-0.08)	(0.01)
LAMBDA	0.36	.
	(5.32)***	
R^2	0.46	0.45
F test	.	195.37***
Wald test	1696.40***	.
Observ.	2187	2187

The t-statistics are given in parentheses below coefficients.
***denotes significance at the 1% level.

The results for the wage equation estimated by ordinary least squares for all the women participating in the labor force including the variable lambda and not including it are given in columns 1 and 2 of table 5. The dependent variable is the logarithm of the hourly wage rate.

Table 5 shows that practically all the significant coefficients in column 2 are smaller in absolute value than the coefficients in column 1 (except EDUCATION2 and CSES whose coefficients are about the same). The average of the estimated logarithm of hourly wages using the Heckman procedure and the traditional method are, respectively, -1.44 and -1.06. The lambda coefficient is positive and highly significant. The positive sign indicates that unmeasured factors that raise the probability of participation also raise the wage rates. In this case it is possible to conclude that the use of traditional methods of estimating the logarithm of hourly wage rates for the whole sample, using only a sample of working women give, in absolute value, a downward biased estimate of the true effect.

The peak of the parabolic curve for age occurs between 45 and 46 in the earning equation (column 1 of table 5). This fact indicates that earnings start decreasing earlier for women than for men.

5. Conclusions.

A traditional method of estimating wage rate functions was compared to the Heckman procedure. This procedure eliminates possible bias caused by sample selectivity, since only people that are employed in the labor force enter in the calculations. The results showed for male workers that the traditional method gives an upward biased (in absolute value) estimate of the true effect of the variables, while for female workers it was observed a downward biased (in absolute value) estimate. Moreover, the coefficient on the variable Mill's Ratio was highly significant showing, once more, that the use of the Heckman procedure was necessary to avoid a sample selection bias.

The results showed a parabolic relation between age and labor force participation and between age and wage rates. The peak of the parabolic curve in the labor force participation equation occurred between 36 and 37 years for men and at 39 years for women. In the wage equation occurred at 50 years for men and between 45 and 46 years for women. It was also observed that as years of school increase, labor opportunities and wages increase. Higher nonlabor income decreases participation in the labor force, and women with young children at home have lower labor force participation rates.

(Received August 1993. Revised July 1994)

References

Becker, G. 1965. "A theory of the allocation of time." *The Economic Journal,* **75**: 493–517.

Berndt, E.R. 1991. *The Practice of Econometrics Classic and Contemporary.* New York: Addison-Wesley.

Greene, W. 1993. *Econometric Analysis.* New York: Macmillan Publishing Company.

Heckman, J. 1974. "Shadow prices, market wages, and labor supply." *Econometrica,* **1**: 679–694.

Heckman, J. 1980. "Sample selection bias as a specification error." In Smith, J.P., *Female Labor Supply: Theory and Estimation.* Princeton: Princeton University Press.

Hoffmann, R. 1993. "Distribuição de renda e pobreza na agricultura paulista." *São Paulo em Perspectiva,* **7**: 107–115.

INAN, IBGE & IPEA 1988. Pesquisa Nacional sobre Saúde e Nutrição.

Kassouf, A.L. 1994. "Demanda de saúde para crianças em diferentes regiões e setores do Brasil." *Revista Pesquisa e Planejamento Econômico.* (No prelo).

Langoni, C. 1972. "As fontes do crescimento econômico brasileiro." *Estudos Econômicos,* **2**: 3–34.

Pastore, J. & J.C. Lopes 1973. "A mão de obra especializada na indústria paulista." IPE (monografia).

Sedlacek, G. & E.C. Santos 1991. "A mulher cônjuge no mercado de trabalho como estratégia de geração de renda familiar." *Revista Pesquisa e Planejamento Econômico,* **21**: 449–470.

Senauer, B., D. Sahn, & H. Alderman 1986. "The effect of the value of time on food consumption patterns in developing countries: evidence from Sri Lanka." *Journal of Agricultural Economics,* **68**: 920–927.

Senna, J. 1976 "Escolaridade, experiência no trabalho e salários no Brasil." *Revista Brasileira de Economia,* **30**: 107–115.

Weisberg, S. 1985. *Applied Linear Regression.* New York: John Wiley & Sons.

White, H. 1980 "A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica,* **4**: 817–838.