MPRA
Munich Personal RePEc Archive

Individual wage and reservation wage: efficient estimation of a simultaneous equation model with endogenous limited dependent variables

Calzolari, Giorgio and Di Pino, Antonino
Universita' di Firenze, Italy, Universita' di Messina, Italy

23. September 2009

Individual Wage and Reservation Wage: Efficient Estimation of a Simultaneous Equation Model with Endogenous Limited Dependent Variables[(1)]

Giorgio Calzolari

Università di Firenze, Department of Statistics "G. Parenti", calzolar@ds.unifi.it

Antonino Di Pino

Università di Messina, Department "V. Pareto", dipino@unime.it

Abstract: We consider a simultaneous equation model with two endogenous limited dependent variables (individual wage and reservation wage) characterized by a selection mechanism determining a two-regimes endogenous-switching. We extend the *FIML* procedure proposed by Poirier-Ruud (1981) for a single equation switching model providing a stochastic specification for both equations and for the selection criterion. An accurate Monte Carlo experiment shows that the relative efficiency of the *FIML* estimator over to the Two-Stage procedure is remarkably high in presence of a high degree of endogeneity in the selection equation.

Keywords: Selection Bias, Endogenous Switching

1. Estimation of Regression Models with Two-Regimes Specification

We study a simultaneous equation model with two equations, Eq. (1) and Eq. (2), whose dependent variables (individual wage and reservation wage) are both partially observed or "limited" as a consequence of a selection mechanism that doesn't permit us to observe the two dependent variables together. The selection mechanism may be specified by a third equation, Eq. (3), whose dependent variable is a binary dummy "indicator" that produces, alternatively, two different regimes where the dependent variables of Eq. (1) and Eq. (2) are alternatively observable. Furthermore, we assume that the binary dummy indicator is not exogenous with respect to the two limited dependent variables. The binary indicator involving the two different regimes is given by the working status of the subject: its value is one if the subject works, and it is zero if the subject does not work. Furthermore, we can consider that the choice of a subject to work or not is influenced by both wage and reservation wage with opposite effects. One may adopt several ways to specify and to estimate simultaneously both wage and reservation wage equations. The two-regimes characteristic suggests to specify the model as an endogenous two-regimes "switching" regression model (Poirier and Ruud,

(1) A. Di Pino has benefited from financial support of MIUR - National Research Project (PRIN 2008-2010)

1981, *inter alia*). In this context, we may adopt two alternative approaches to estimate the model: i) a Two-Stage procedure generally utilized to estimate regression models with limited dependent variable (Heckman, 1976); ii) a *FIML* approach utilized to estimate regression models with endogenous switching (Poirier and Ruud, 1981). The Two-Stage is the most widely used by applied econometricians; it is simpler and provides a consistent parameters estimate. *FIML* is less popular in applied works, presumably for its computational complexity being a maximum likelihood method; besides consistency it also attains asymptotic efficiency, if the model is correctly specified. What is the loss of efficiency implied by the widely used Two-Stage method? In the next section, we provide a brief discussion on the stochastic specification of the model and on the *FIML* estimator characteristics. Then we show the performances of both Two-Stage and *FIML* estimators in a detailed set of Monte Carlo experiments. Moreover, the results of an empirical application will be briefly resumed.

2. The Model and the Estimation Procedures

Poirier-Ruud (1981) shows that we have an endogenous switching model only if the specification of the regression equation in two different regimes is related to the expected value of the dependent variable in both regimes. If applied to our model, it implies that the individual reservation wage level is given by the subject's evaluation of her/his own qualities and endowment, and it is not conditional on her/his non-working condition. Analogously, working income or wage depends on several exogenous factors which determine the working status, but does not depend on the employment condition of the subject. In this context, we utilize an estimation procedure based on a likelihood function given by the product of marginal probabilities of the subjects of perceiving a wage or, alternatively, of desiring a reservation wage. Let's start considering a three equations model. The purpose is to estimate simultaneously the individual wage, the reservation wage and the individual participation propensity in the labour market.

$$W = x'_1 \alpha + u \qquad \text{if } L = 1;\ 0 \ \textit{otherwise} \tag{1}$$

$$RW = x'_2 \beta + v \qquad \text{if } L = 0;\ 0 \ \textit{otherwise} \tag{2}$$

$$L^* = z'\gamma + \varepsilon \qquad \begin{cases} L = 1 & \textit{if } L^* > 0 \\ L = 0 & \textit{otherwise} \end{cases} \tag{3}$$

Moreover, if $L = 1$, $\varepsilon > -z'\gamma$; if $L = 0$, $\varepsilon \le -z'\gamma$

The variable W is a vector of n_1 elements, the number of pays perceived by the n_1 employed people. The variable RW is a vector of n_2 elements, the number of the reservation wages desired by the n_2 subjects unemployed and looking for a job. The binary variable L is a vector of $n = n_1 + n_2$ elements composed by n_1 unitary elements and n_2 elements equal to zero. Moreover, x'_1; x'_2 and z' are row-vectors, respectively, of three exogenous variables matrices $\mathbf{X_1}$, $\mathbf{X_2}$ and $\mathbf{Z}$. Some exogenous explanatory variables on the right hand side of each equation can be common to the three equations, but the following two identification conditions must be observed: i) at least one of the regressors of L (included in Z) must be independent with respect to W and RW, and ii) at least one of the regressors of, respectively, W and RW, must not appear in the equation of L. The error terms u, v, and ε are assumed normally distributed with covariance matrix given by: $\mathbf{\Sigma} = \begin{bmatrix} \sigma_u^2 & 0 & \sigma_{u\varepsilon} \\ 0 & \sigma_v^2 & \sigma_{v\varepsilon} \\ \sigma_{u\varepsilon} & \sigma_{v\varepsilon} & 1 \end{bmatrix}$. Taking into account the well-known properties of density functions and of conditional density functions: $\varphi(u,\varepsilon) = \varphi(u)\varphi(\varepsilon|u)$ and $\varphi(v,\varepsilon) = \varphi(v)\varphi(\varepsilon|v)$, the probability marginal distributions are:

$$P(u) = \int_{-z'\gamma}^{+\infty} \varphi(u,\varepsilon)d\varepsilon = \int_{-\infty}^{z'\gamma} \varphi(u,\varepsilon)d\varepsilon \quad \text{and} \quad P(v) = \int_{-\infty}^{-z'\gamma} \varphi(v,\varepsilon)d\varepsilon = \int_{z'\gamma}^{+\infty} \varphi(v,\varepsilon)d\varepsilon \qquad (4)$$

Furthermore, by substituting L = 1 or L = 0 into the third equation, we have, respectively:

$$P(u|u>0) = P(\varepsilon > -z'\gamma) = 1 - \Phi(-z'\gamma) = \Phi(z'\gamma) \quad \text{if } L = 1 \qquad (5a)$$

$$P(v|v \le 0) = P(\varepsilon \le -z'\gamma) = \Phi(-z'\gamma) = 1 - \Phi(z'\gamma) \quad \text{if } L = 0 \qquad (5b)$$

If we consider the error terms u and v normally distributed, the p.d.f. of u if $u|u>0$ and of v if $v|v \le 0$ are, respectively:

$$\varphi(u) = \frac{1}{\sigma_u} \frac{\varphi\left(\frac{W - x'_1\alpha}{\sigma_u}\right)}{\Phi(z'\gamma)} \quad \text{if } L = 1, \quad \text{and} \quad \varphi(v) = \frac{1}{\sigma_v} \frac{\varphi\left(\frac{RW - x'_2\beta}{\sigma_v}\right)}{1 - \Phi(z'\gamma)} \quad \text{if } L = 0 \qquad (6)$$

Furthermore, utilizing the conditional density of a bivariate normal distribution and assuming $\mathcal{L}(u,v,\varepsilon) = \prod_{L=1} \varphi(u,\varepsilon) \prod_{L=0} \varphi(v,\varepsilon) = \prod_{L=1} \varphi(u) \int \varphi(\varepsilon|u \ge 0) \prod_{L=0} \varphi(v) \int \varphi(\varepsilon|v \le 0)$ the Likelihood function to be maximized is:

$$\prod_{L=1} \left\{ \frac{1}{\sigma_u} \varphi\left(\frac{W - x'_1\alpha}{\sigma_u}\right) \Phi\left[\frac{\left(z'\gamma + \frac{\sigma_{u\varepsilon}}{\sigma_u^2}(W - x'_1\alpha)\right)}{\left(\sigma_u^2 - \sigma_{u\varepsilon}^2\right)^{\frac{1}{2}}}\right] \right\} \cdot \prod_{L=0} \left\{ \frac{1}{\sigma_v} \varphi\left(\frac{RW - x'_2\beta}{\sigma_v}\right) \cdot \right.$$

$$\left. \cdot \left[1 - \Phi\left[\frac{\left(z'\gamma + \frac{\sigma_{v\varepsilon}}{\sigma_v^2}(RW - x'_2\beta)\right)}{\left(\sigma_v^2 - \sigma_{v\varepsilon}^2\right)^{\frac{1}{2}}}\right]\right] \right\} \qquad (7)$$

The results of a Monte Carlo experiment show that both estimated coefficients and standard errors are asymptotically equal for *FIML* and *T-S* estimator when the covariances of the error terms, $\sigma_{u\varepsilon}$ and $\sigma_{v\varepsilon}$, are imposed equal to zero (simulating absence of endogeneity in the switching-model). Instead, we can observe that relative efficiency of *FIML* procedure is significant when the error terms correlation is close to 50% or more (cfr. Table 1).

Table 1 - *Montecarlo results of FIML and T-S estimations*

Error terms correlation (in percent)	Differential (in percent) of *T-S* estimator variance with respect to the *ML* estimator
10%	close to 0
50%	+ 10-20%
95%	+ 30-100%

For an empirical comparison, we apply both *FIML* and *T-S* estimators to the ISTAT Survey on Italian Graduates in 2001 dataset. In this context, wage equation, reservation wage equation, and participation equation are simultaneously estimated by applying both estimation procedures. The estimation results, not reported here for the sake of brevity, show that estimated coefficients and standard errors are very similar by utilizing both *FIML* and *T-S* procedures, and that the residual-based estimates of the error terms covariances, $\sigma_{u\varepsilon}$ and $\sigma_{v\varepsilon}$, are close to zero (endogeneity seems to be absent in the model).

References

Heckman J. J. (1978) "Dummy Endogenous Variables in a Simultaneous Equation System", *Econometrica*, 46, 931-959.

Heckman J. J.,Tobias J., Vytlacil E. (2003) "Simple Estimators for Treatment Parameters in a Latent Variable framework", *Review of Economics and Statistics,* 85 Issue 3, 748-755.

ISTAT (2004) *Survey on Italian Graduates in 2001*

Poirier D. J., Ruud P. A. (1981) "On The Appropriateness of Endogenous Switching", *Journal of Econometrics*, 16, 249-256.