PARAMETRIC AND SEMIPARAMETRIC ESTIMATION OF SAMPLE SELECTION MODELS: AN EMPIRICAL APPLICATION TO THE FEMALE LABOUR FORCE IN PORTUGAL

Danilo Coelho*, Helena Veiga† and Róbert Veszteg‡

February 28, 2005

Abstract

This paper analyses the Portuguese female labour market using a sample selection model. Our results are presented as a contrast to those in Martins (2001). We offer both parametric and semiparametric results applying a different approach: the normal kernel with local smoothing. Several estimates differ from Martins' in sign and/or in significance. According to our results, for example, the husband's wage has positive effect on the wife's participation in the labour market.

*Departament d'Economia i d'Història Econòmica, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain and Diretoria de Estudos Sociais, Instituto de Pesquisa Econômica Aplicada, Brasília - DF, 70076-900, Brasil; Danilo Coelho acknowledges the financial support from Capes, Brazilian Ministry of Education; Danilo.Coelho@idea.uab.es

†Departamento de Estadística, Universidad Carlos III de Madrid, c/ Madrid 126, 28903 Madrid, Spain; Helena Veiga acknowledges the financial support from Grant Sponsors: Program Praxis XXI (FCT) and FEP; mhveiga@est-econ.uc3m.es

‡Departamento de Economía, Universidad de Navarra, Edificio de Bibiliotecas (Entrada Este), 31080 Pamplona, Spain; Róbert Veszteg acknowledges the financial support from project BEC2003-01132; Robert.Veszteg@idea.uab.es

1 Introduction

The main issue in estimating labour supply is that no market wage is observable for those who decide not to participate in the market. Martins (2001) analyses female labour force in Portugal applying the sample selection model proposed by Heckman (1974, 1979) to solve this problem. The conventional sample selection model consists of two equations: in the example of female labour supply, the selection equations delivers the probability that a woman participates in the labour market, while the wage equation determines wages following the human capital approach. This model overcomes the bias that arises when female labour supply is estimated from data only for women who work. Moreover, since the originally used parametric methods gives inconsistent estimates when incorrect distributional assumptions are made, Martins (2001) offers both parametric and semiparametric estimates, and compares the two sets of results.

The purpose of this comment is to correct some errors in the estimation process that appear in Martins (2001) and provide new estimates. The first error is in the parametric probit estimation, as the presented results do not maximize the log-likelihood function. In the global maximum more variables become significant. As for the semiparametric estimation method, the kernel function used in Martins (2001) can take on both positive and negative values, which implies that the participation probability estimates may be outside the interval [0,1]. This makes obtaining the log-likelihood function impossible both theoretically and in practice. We have solved the problem by applying local smoothing in the kernel estimation, as suggested by Klein and Spady (1993).

2 Parametric Estimation

We use the same data set and model specification as in Martins (2001). For the precise definition and description of the variables used in this comment please check Martins (2001). The results are generated using Ox version 2.20 (see Doornik, 1999). The maximization method is the quasi-Newton method developed by Broyden, Fletcher, Goldfarb, Shanno (BFGS). We have tried several initial values and we report the results that correspond to the best log-likelihood value.

Table 1 reports our probit results and Martins' for comparison. The first seven variables in rows belong to the participation equation, while the next six belong to the wage equation. We can observe an important gain of significance among the coefficients of the wage equation. The variables PEXP2 and PEXPCHILD become significant at any usual level of significance. On the other hand, at 10% the coefficient of PEXP changes sign.

Our results are more in line with the assumptions of the model specification: "...the main alternative use for woman's time is child rearing (and the home activities related to this task)..." Martins (2001).

We also perform the specification tests suggested by Horowitz (1993) and Horowitz and Härdle (1994) in order to test the parametric model against a semiparametric alternative. The tests have been applied to the participation equation for windowwidths of 0.3, 0.5, 0.6, 0.8 and 1.0. These results reject the parametric model at significance levels of 90 and 95% as in Martins (2001).

3 Semiparametric estimation

In this part we use Klein and Spady's (1993) estimator. Since in the first step we are estimating the participation equation the predictions should lie in the [0,1] interval, otherwise for some parameter values it is impossible to compute the quasi-loglikelihood.[1] Condition (C.8a) in Klein and Spady (1993), which is necessary for the consistency and normality proof is violated by any non-negative kernel. The kernel applied in Martins (2001) satisfies condition (C.8a) but it can take negative values, which implies that the participation probabilities estimates may be outside the interval [0,1]. For this reason the alternative condition (C.8b) in Klein and Spady (1993) is considered using the normal kernel with local smoothing, setting $\varepsilon' = 0.5$.

Therefore the windowwidths for local smoothing have been computed according to

$$h_N \widehat{\lambda}_{zj} = [h_N][\widehat{\sigma}_z(\beta)][\widehat{L}_{zj}].$$

The component, $\widehat{\sigma}_z$, is the sample standard deviation of $v(x;\beta)$, the index function, conditioned on Z. For the final component, which reflects local smoothing, let $\widehat{l}_{zj} \equiv \widehat{g}_{zv}(v_j;\beta,\widehat{\lambda}_{zj};h_{NP})$ be a preliminary kernel density estimate without local smoothing with both h_{NP} and h_N set to 0.3. As suggested by condition (C.8b) in Klein ans Spady (1993) we set the value of parameter $\widehat{\lambda}_{zj}$ to 1.

The local smoothing parameter according to Klein and Spady (1993) can be defined by

$$\widehat{L}_{zj} \equiv \{[\widehat{l}_{zj} + (1 - \widehat{\tau}_{Pj})]/m\}^{-\frac{1}{2}}$$

where

$$\widehat{\tau}_{Pj} \equiv \tau(\widehat{l}_{zj}, h_{NP}^{\varepsilon'}),$$

the trimming function

$$\tau(t;\varepsilon) \equiv \{1 + \exp[(h_N^{\varepsilon/5} - t)/h_N^{\varepsilon/4}]\}^{-1}$$

[1] The original paper does not give indications how this problem was solved.

and m is the geometric mean of $\widehat{l}_{zj}$.

In the second step we perform the OLS method with White's heteroscedasticity consistent estimator. The parameter k, that represents the number of basis functions in the semiparametric method, is set to 2.

Table 2 reports the results of the two-step semiparametric estimation applying the same structure that appeared in Table 1.[2] In the participation equation the main difference occurred with the coefficient of HW. It shows that the husband's wage has a significant large positive effect on the participation probability. This can be justified by the fact that due to social reasons wives of men with higher wage incomes tend to become more independent and participate more in the labour market. As for the second step results, we can observe a gain of significance in the estimates of the wage equation. As compared to Martins' results, the variables related to the potential work experience (PEXP and PEXP2) turn out to be significant, while the effect that PEXPPCHILD and PEXP2CHILD have on the wage is of opposite sign. These results are in line with economic theory.

References

[1] Doornik, J.A. (1999) Object-Oriented Matrix Programming Using Ox, 3rd ed. London: Timberlake Consultants Press and Oxford: www.nuff.ox.ac.uk/Users/Doornik.

[2] Heckman, J. (1974) Shadow prices, market wages and labor supply, *Econometrica,* **42**, 679-694.

[3] Heckman, J. (1979) Sample selection as a specification error, *Econometrica* **47,** 153-161.,

[2] The constant reported for the wage equation includes the constant from the correction term. The constant of the wage equation is -0.569, using Andrews-Schafgans' consistent estimate.

[4] Fernández, I.A. and Rodríguez-Poo, J.M. (1997) Estimation and specification testing in female labor participation models: parametric and semiparametric methods, *Econometric Reviews,* **16,** 229-247.

[5] Martins, M.F.O. (2001) Parametric and semiparametric estimation of sample selection models: an empirical application to the female labour force in Portugal, *Journal of Applied Econometrics,* **16**, 23-39.

[6] Horowitz, J.L. (1993) Semiparametric estimation of a work-trip mode choice model, *Journal of Econometrics,* **58**, 49-70.

[7] Horowitz, J.L. and Härdle, W. (1994) Testing a parametric model against a semiparametric alternative, *Econometric Theory,* **10**, 821-848.

[8] Klein, R.W. and Spady, R.H. (1993) An efficient semiparametric estimator for binary response models, *Econometrica,* **61**, 387-421.

Table 1: Parametric estimates for the participation and wage equations

	Our results (2005)		Martins (2001)	
	Coefficients	t-ratio	Coefficients	t-ratio
CONSTANT	-2.639	-3.553	-0.570	-0.633
CHILD	-0.122	-4.714	-0.120	-4.444
YCHILD	-0.019	-0.275	-0.090	-1.184
HW	0.064	0.999	-0.100	-1.333
EDU	0.116	12.252	0.150	16.667
AGE	0.998	4.371	0.810	3.306
AGE2	-0.145	-5.106	-0.120	-4.000
CONSTANT	5.372	49.465	4.500	26.316
EDU	0.071	13.383	0.110	12.222
PEXP	-0.139	-1.876	0.130	1.781
PEXP2	0.064	4.413	-0.003	-0.188
PEXPCHILD	0.072	3.526	0.031	1.409
PEXP2CHILD	-0.018	-3.048	-0.010	-1.667
VAR	0.432	31.491	*	*
RO	-0.820	-5.062	0.350	1.606
Log L		-2470.600		-2488.500

Table 2: Two-step results for the participation and wage equations

	Our results (2005)		Martins (2001)	
	Coefficients	t-ratio	Coefficients	t-ratio
CHILD	-0.985	2.785	-0.097	-8.083
YCHILD	0.527	0.375	-0.018	-0.450
HW	12.444	38.596	-0.078	-2.600
EDU	7.792	32.276	0.086	7.167
AGE	1.000	-	1.000	-
AGE2	-7.175	-39.215	-0.145	-48.330
CONSTANT	-1.421	-0.865	4.700	47.000
EDU	0.250	7.846	0.100	25.000
PEXP	4.835	6.490	0.080	0.833
PEXP2	-1.058	-5.588	0.012	0.600
PEXPCHILD	-0.304	-4.173	0.043	2.150
PEXP2CHILD	0.068	3.284	-0.013	-1.857
Log L		-5547.614		