Semiparametric three Step Estimation Methods in Labor Supply Models

Ana I.Fernández
Departamento de Econometría
Universidad del País Vasco

Juan M.Rodríguez-Póo
Departamento de Economía
Universidad de Cantabria

Stefan Sperlich
Institut für Statistik und Ökonometrie
Humboldt-Universität zu Berlin

September 4, 1998

Abstract

The aim of this paper is to provide an alternative way of specification and estimation of a labor supply model. The proposed estimation procedure can be included in the so called predicted wage methods and its main interest is twofold. First, under standard assumptions in studies of labor supply, the estimator based on predicted wages is shown to be consistent and asymptotically normal. Moreover, we propose also a consistent estimator of the asymptotic variance covariance matrix estimator. Secondly, we introduce a semiparametric estimator based on marginal integration techniques that allows for nonlinear relationships between working hours and other explanatory variables. We show the asymptotic properties of this estimator and we compare the results empirically against those obtained in standard three step estimators based on predicted wages. [1]

Keywords: Semiparametric regression, Heckman estimator three step estimator, Additive Models, Marginal integration, Predicted wage methods.

[1]This research was financially supported by Sonderforschungsbereich 373, Deutsche Forschungsgemeinschaft and the TMR project *Savings and Pensions* by the European Comunity at Humboldt-Universität zu Berlin, and the Dirección General de Enseñanza Superior del Ministerio de Educación y Ciencia under research grants PB95-0346 and PB96-1469-C05-03. We thank Peter Hall and Stefan Profit for helpful discussion.

1 Introduction

The purpose of this paper is to provide an alternative way of specification and estimation of a standard model of labor supply. Our interest involves a structural labor supply model in which hours of work depend on the wage rate and other explanatory variables. The difficulty in estimating such system occurs because first, information is not available on the wage rate for those who do not work, and second, the wage rate is determined endogenously. To avoid the first problem, the estimation method must take into account the sample selection bias and the second problem is solved by specifying a relationship that considers the wage rate as a endogenous variable.

The estimation procedure proposed in this paper is a three step method based on the ideas developed by Heckman (1979,1993) and it can be included in the so called predicted wage methods (Wales and Woodland, 1980). Its main interest is twofold. First, in the standard econometric model that is traditionally assumed in studies of labor supply, the three step estimator based on predicted wages is shown to be consistent and asymptotically normal. Moreover, we provide also a consistent estimator of the asymptotic variance covariance matrix for this three step estimator. Second, the classical assumption of linearity between working hours and other explanatory variables is relaxed allowing for a semiparametric partial additive relationship. In particular, this more flexible specification can be of great interest when analyzing the relationship between hours and wage rates. This relationship has been traditionally assumed to be linear, but as Blundell and Meghir (1986) pointed out there exist very few theoretical foundations to support this hypothesis. The nonparametric additive components are estimated according to the method developed in Härdle, Huet, Mammen and Sperlich (1998) and it is based on marginal integration techniques (Linton and Nielsen, 1995). The resulting estimator turns out to be a semiparametric one, in the sense that the distribution of the random errors is assumed to be known (gaussian), but the index function is not specified to be linear. Within this set up, we show the root-n consistency and the asymptotic distribution of this three step semiparametric estimator and we also provide a consistent estimator of the asymptotic variance covariance matrix. This enables us to make comparisons between the three step fully parametric estimator and the semiparametric one, mainly, analyzing the relationship between hours and wage rates.

Note that our extension of the standard parametric model is quite different from the one proposed by Ahn and Powell (1993). They mainly focused on relaxing the assumptions on the model only in the first step. They neither calculated the standard deviation of the second step nor considered the third step at all. But on the other hand, in the first step their approach is more general than ours and a combination of both would be certainly worth to be considered.

In the next section of the paper we specify a simultaneous equation structural model of labor supply. We also recall the basic principles of three step estimation methods based on predicted wages and we establish the main statistical results in the fully parametric context. In section 3 we introduce the semiparametric three step estimator and we establish its main asymptotic properties. Section 4 presents an extensive application based on a spanish labor force data and in section 5 we conclude.

2 The Structural Model of Labor Supply

In this section we start by considering a structural econometric model of labor supply. To this end, we previously specify the relationship among wages, hours of work (or participation) and other explanatory variables, and next, we will introduce formally the assumptions that are necessary to obtain the statistical properties of the three step estimator based on predicted wages. We will finally propose a consistent estimator of the asymptotic variance covariance matrix of the previous estimator.

Let us consider a labor supply in which both the wage rate and hours of work are endogenously determined. The extended model can be expressed as follows. Hours of work are a function of a vector of explanatory variables x which includes the log of the wage rate as its first element $\{w\}$. In addition it is assumed that w is a function of another vector of exogenous variables z. Thus, we have for the individual i

- the hours equation,

$$(1) \qquad y_i = \begin{cases} h(w_i, x_i; \beta) + u_{1i} & if \quad h(w_i, x_i; \beta) + u_{1i} > 0 \\ 0 & otherwise \end{cases}$$

- and the wage equation.

$$(2) \qquad w_i = g(z_i; \gamma) + u_{2i} \qquad if \quad y_i > 0$$

As it has been remarked in Wales and Woodland (1980), there exist two problems in the estimation of the structural parameters of the previous simultaneous equation system. First, both equations are subject to sample selection bias and second, the model is a set of two simultaneous equations in which the wage rate, which is an explanatory variable in the hours equation, is correlated with the hours equation disturbance.

In order to solve these problems and estimate the structural parameters of the previous model, some further hypotheses are needed.

(A.1) The values $\{w_i, x_i, z_i\}_{i=1}^N$ are realizations from measurable i.i.d random variables, where $W \in I\!R$, $X \in I\!R^{p+d}$ and $Z \in I\!R^r$. Moreover, $\{y_i\}_{i=1}^N$ are realizations from a truncated random variable, and ζ is a binary variable that takes the value 1 as $y > 0$ and 0 otherwise.

(A.2) The random variables have bounded support and $2+\epsilon$ finite moments for some $\epsilon > 0$.

(A.3) The data satisfy the restrictions (1) and (2).

(A.4) Moreover,

$$
\begin{aligned}
h(w_i, x_i; \beta) &= \beta_w w_i + x_i^T \beta, \\
g(z_i; \gamma) &= z_i^T \gamma
\end{aligned}
\tag{3}
$$

where the vector z contains at least one variable not contained in x.

(A.5) Define the parameter vector $\theta = (\beta_w, \beta, \gamma)$ and the parameter space $\Theta = I\!B_w \times I\!B \times \Gamma$. Then $\theta \in \Theta$, Θ is a compact set, and θ_0 is an interior point of Θ.

(A.6) $\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \sim N(0, \Sigma)$ where $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$.

Taking into account these assumptions, in order to estimate the structural parameters of the labor supply model several procedures have been proposed in the literature (see Wales and Woodland, 1980). Among them, the methods based on predicted wage rates have been followed by several authors (see Boskin, 1973; Hall, 1973 and Rosen, 1976). Traditionally, this method has consisted on a three step procedure that is implemented as follows.

In the first step then, we estimate the parameters of the reduced form model for the hours equation by using a probit maximum likelihood procedure. The reduced form model is

$$
\zeta_i = \begin{cases} 1 & if \quad z_i^T \alpha + x_i^T \beta + v_{1i} > 0 \\ 0 & otherwise \end{cases}
\tag{4}
$$

The relationship between the structural and the reduced form parameters is $\alpha = \beta_w \gamma$ and $v_{1i} = \beta_w u_{2i} + u_{1i}$. The variable ζ_i is equal to 1 iff $y_i > 0$ and 0 otherwise, and therefore, under assumptions *(A.1)* to *(A.6)* the likelihood function has the following form

$$
L(\alpha, \beta) = \prod_{i=M+1}^{N} \left(1 - F\left(\frac{z_i^T \alpha + x_i^T \beta}{\sigma_v}\right)\right) \prod_{i=1}^{M} F\left(\frac{z_i^T \alpha + x_i^T \beta}{\sigma_v}\right),
\tag{5}
$$

where M is the number of individuals for those who both wages and number of working hours are observed, N is the number of individuals in the sample, $F(\bullet)$ is the cumulative normal distribution function and $\sigma_v^2 = \beta_w^2\sigma_2^2 + \sigma_1^2 + 2\beta_w\rho\sigma_1\sigma_2$. Maximum likelihood estimates of the reduced form parameters $\hat{\alpha}$ and $\hat{\beta}$ can be estimated by introducing the identifying restriction $\sigma_v = 1$. The estimators $\hat{\alpha}$ and $\hat{\beta}$ must fulfill the following restriction

$$(6) \qquad \frac{1}{N}\sum_{i=1}^{N} m\left(\tilde{x}_i;\delta\right) = \frac{1}{N}\sum_{i=1}^{N} \frac{\zeta_i - F\left(\tilde{x}_i^T\delta\right)}{F\left(\tilde{x}_i^T\delta\right)\left(1-F\left(\tilde{x}_i^T\delta\right)\right)} f\left(\tilde{x}_i^T\delta\right)\tilde{x}_i = 0$$

where $\tilde{x}_i^T = \begin{pmatrix} x_i^T & z_i^T \end{pmatrix}$, $\delta = \begin{pmatrix} \alpha^T & \beta^T \end{pmatrix}^T$ and $f(\bullet)$ stands for the gaussian density.

In the second step, the log wage equation is estimated by least square methods correcting the sample selection bias by the Mill's ratio. More exactly, the selection bias corrected equation will be

$$(7) \qquad w_i = z_i^T\gamma_1 + \gamma_2\lambda\left(z_i^T\hat{\alpha} + x_i^T\hat{\beta}\right) + w_i - E\left(w_i|w_i>0\right), \quad i = 1,\cdots,M,$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the maximizers of (5) at this stage, $\hat{\gamma} = \begin{pmatrix} \hat{\gamma}_1^T & \hat{\gamma}_2 \end{pmatrix}^T$ must fulfill the following condition

$$(8) \qquad \frac{1}{N}\sum_{i=1}^{N} g\left(\tilde{z}_i;\hat{\delta},\gamma\right) = -\frac{2}{N}\sum_{i=1}^{N}\zeta_i\tilde{z}_i\left(w_i - \tilde{z}_i^T\gamma\right) = 0$$

where $\tilde{z}_i^T = \left[z_i^T \quad \lambda\left(\tilde{x}_i^T\hat{\delta}\right)\right]$and $\lambda(\bullet) = f(\bullet)/F(\bullet)$ is the Mill's ratio. In the third step the structural parameters of the hours equation are estimated. In order to do this, we construct the predicted wages unconditionally for all individuals in the sample, i.e.

$$(9) \qquad \hat{w}_i = z_i^T\hat{\gamma} \qquad i = 1,\cdots,N.$$

Recall that $\hat{\gamma}$ are either O.L.S. or feasible G.L.S. of the log wage equation. Then, by substituting the predicted wages in the hours equation (1) it is possible to estimate the structural parameters by Tobit maximum likelihood. For this unconditional predicted wages the likelihood function has the following expression

$$(10) \qquad L\left(\beta_w,\beta\right) = \prod_{i=1}^{M} f\left(\frac{y_i - \beta_w\hat{w}_i - x_i^T\beta}{\sigma_1}\right)\prod_{i=M+1}^{N}\left(1 - F\left(\frac{\beta_w\hat{w}_i + x_i^T\beta}{\sigma_1}\right)\right).$$

Other possible ways to calculate predicted wages are available. If the structural model is recursive ($\rho = 0$), then the predicted wages for participants are the observed ones, and for nonparticipants, the predicted ones. The problem is that if $\rho \neq 0$, then $\hat{w}$ is endogenous and therefore the Tobit maximum likelihood estimators are inconsistent. This is also the problem when the predicted wages are generated conditionally by using the Mill's ratio. In this case, the estimators of the hours equation are also inconsistent since the criteria that determines the truncation of both structural equations is the same. The estimators derived from the maximization of (10) will fulfill the following equation condition

$$(11)\frac{1}{N}\sum_{i=1}^{N} h\left(\mathbf{x}_i;\hat{\delta},\hat{\gamma},\tau\right) = -\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\zeta_i}{f\left(\mathbf{x}_i^T\tau\right)} f'\left(\mathbf{x}_i^T\tau\right)\mathbf{x}_i + \frac{1-\zeta_i}{1-F\left(\mathbf{x}_i^T\tau\right)} f\left(\mathbf{x}_i^T\tau\right)\mathbf{x}_i\right] = 0$$

where $\mathbf{x}_i^T = \left[z_i^T\hat{\gamma} \quad x_i^T\right]$ and $\tau = \left(\beta_w^T \quad \beta^T\right)^T$. $f'(\bullet)$ stands for the first derivative of the standard normal density function with respect to the whole argument.

It should be recognized that the usual Tobit standard errors are not appropriated for this estimators since predicted rather than actual wage rates are used for non workers. We show now, under fairly general assumptions, the consistency and asymptotic normality of the three step estimators derived previously. Moreover, we provide a consistent estimator of the asymptotic variance-covariance matrix of the three step estimator. Before to give the theoretical results, we will introduce some notation. Let

$$m(x,z) = m(x,z;\gamma_0), \quad g(x,z) = g(x,z;\delta_0,\gamma_0), \quad h(x,z) = h(x,z;\delta_0,\gamma_0,\tau_0),$$

$$M_\gamma = E\left[\nabla_\gamma m(x,z)\right], \quad G_\delta = E\left[\nabla_\delta g(x,z)\right], \quad G_\gamma = E\left[\nabla_\gamma g(x,z)\right],$$

$$H_\gamma = E\left[\nabla_\gamma h(x,z)\right], \quad H_\tau = E\left[\nabla_\tau h(x,z)\right], \quad \text{and } \psi(x,z) = -M_\gamma^{-1} m(x,z).$$

Theorem 1 *Assume conditions (A.1) to (A.6) hold, then*

$$\sqrt{N}\left(\hat{\tau}-\tau_0\right) \xrightarrow{D} N\left(0, V\left(\tau_0\right)\right)$$

where

$$V(\tau_0) = H_\tau^{-1} E\left[h(x,z) + H_\gamma G_\gamma^{-1}\left(g(x,z) + G_\delta\psi(x,z)\right)\right]\left[h(x,z) + H_\gamma G_\gamma^{-1}\left(g(x,z) + G_\delta\psi(x,z)\right)\right]^T H_\tau^{-1}.$$

Moreover, if

$$\hat{V}(\hat{\tau}) = \hat{B}_N(x_i,z_i)^{-1}\hat{A}_N(x_i,z_i)\hat{B}_N(x_i,z_i)^{-1}$$

where

$$\hat{A}_N(x_i,z_i) = \frac{1}{N}\sum_{i=1}^{N}\left\{h\left(x_i,z_i,\hat{\delta},\hat{\gamma},\hat{\tau}\right) + \hat{A}_{1N}(x_i,z_i)\right\}\left\{h\left(x_i,z_i,\hat{\delta},\hat{\gamma},\hat{\tau}\right) + \hat{A}_{1N}(x_i,z_i)\right\}^T,$$

$$\hat{A}_{1N}(x_i,z_i) =$$

$$\left[\frac{1}{N}\sum_{i=1}^{N}\nabla_\gamma g\left(x_i,z_i,\hat{\delta},\hat{\gamma}\right)\right]^{-1}\left\{g\left(x_i,z_i,\hat{\delta},\hat{\gamma}\right) + \left[\frac{1}{N}\sum_{i=1}^{N}\nabla_\delta g\left(x_i,z_i,\hat{\delta},\hat{\gamma}\right)\right]\hat{A}_{2N}(x_i,z_i)\right\},$$

$$\hat{A}_{2N}(x_i,z_i) = -\left[\frac{1}{N}\sum_{i=1}^{N}\nabla_\delta m\left(x_i,z_i,\hat{\delta}\right)\right]^{-1} m\left(x_i,z_i;\hat{\delta}\right)$$

and

$$\hat{B}_N(x_i, z_i) = \frac{1}{N}\sum_{i=1}^{N} \nabla_\gamma h\left(x_i, z_i, \hat{\delta}, \hat{\gamma}, \hat{\tau}\right),$$

then

$$\hat{V}(\hat{\tau}) \longrightarrow_p V(\tau_0)$$

as N tends to infinity.

However, in order to identify the structural model described in (1) and (2) it has been necessary to introduce some restrictions that are hardly supported by economic theory but necessary to identify and estimate some structural parameters (Mroz, 1987). In particular, in this paper we are concerned about the assumption of linearity between the hours of work and the log-wages. This is implicitly assumed in the econometric model that incorporates assumptions *(A.1)* to *(A.6)* and, in fact, this has been the model that has been more commonly used in econometric applications of labor supply (Wales and Woodland, 1980). However, Blundell and Meghir (1986) discuss some models proposed by economic theory that do not imply linearity between hours of work and log-wages. On these grounds, we propose a three step estimator of the structural parameters of a labor supply model that do not rely on the assumption of linearity between hours of work and log of wages. Moreover, it is also possible to estimate with flexible methods the relationships between the endogenous variable and other explanatory variables in the hours equation. Although we will not discuss the implications of misspecification on distributional assumptions in this type of models, we remark that it is also a relevant issue that has been studied in deep by Vijverberg (1991).

3 A Semiparametric Approach to the Heckman Estimator

In this section we will start by relaxing some of the assumptions that have been discussed in the previous section. More concretely, assumptions *(A.4)* and *(A.5)* will be replaced by the following weaker conditions

(A.4')

$$E(\zeta | W = w_i, X = x_i) = h(w_i, x_i; \beta) = F\left(\beta^T s_i + \eta(t_i, w_i)\right) \qquad i = 1, \cdots, N$$

and $g(z_i; \gamma) = z_i^T \gamma$ for $i = 1, \cdots, M$, where $F(\bullet)$ is the cumulative normal density function, and η is an unknown function from $\mathbb{R}^{d+1}$ in $\mathbb{R}$.

(A.5')

$$\eta\left(t_i, w_i\right) = \alpha + \sum_{j=1}^{d} \eta_j(t_{ij}) + \eta_{d+1}(w_i) \qquad i = 1, \cdots, N$$
$$\eta_{d+1}(z_i^T \gamma + u_{2i}) = \sum_{k=1}^{r} \varphi_k(z_{ik}) + \varphi_{r+1}(u_{2i}) \qquad i = 1, \cdots, M,$$

where φ is an unknown function.

(A.6') Let us define as define by Ψ the set of all real valued functions that are four times continuously differentiable in $I\!R$.. Then, $\{\eta_j(\bullet)\}_{j=1}^{d+1}$ and $\{\varphi_k(\bullet)\}_{k=1}^{r+1} \in \Psi$.

Assumption *(A.4')* allows for possible non-linear relationships among the variables and assumption *(A.5')* introduces the restriction of additivity in the nonparametric function $\eta(\bullet)$. Additivity provides several advantages as avoiding the curse of dimensionality (Stone, 1986) and a better interpretability. Additionally we also assume the following

(A.7) $X = (S, T)$, where $T \in I\!R^d$ are absolutely continuous random variables and $S \in I\!R^p$ are dummy variables. $\{s_i, t_i\}_{i=1}^{N}$ are realizations from (S, T). S and T have compact support D_S and D_T. The support D_T is of the form $D_{T,1} \times D_{T,-1}$ with $D_{T,1} \subset I\!R$ and $D_{T,-1} \subset I\!R^{d-1}$. T has a twice continuously differentiable density f_T with $\inf_{t \in D_T} f_T\left(t\right) > 0$.

(A.8) For the η's we set the following, $E_{t_j}[\eta_j(t_j)] = 0$ for all $j = 1, \ldots, d$ and $E[\eta(T)] = \alpha$. For the φ's we need $E_{z_k}[\varphi(z_k)] = 0$ for all k. Additionally we set $E_{w_i}[\eta_{d+1}(w_i)] = 0$

Assumption *(A.7)* is considered without loss of generality, all variables could be continuous and the results would still hold. Finally, assumption *(A.8)* is necessary for the sake of identification.

Taking into account the structural model, equations (1) and (2), and the assumptions *(A.1)* to *(A.3)* , *(A.4')*, *(A.5')*, and *(A.6)* to *(A.8)*, we now develop the three step semiparametric estimator and its statistical properties.

3.1 First step: Generalized Additive Partial Linear Model

In the first step, we estimate the selection bias correcting factor, the so-called Mills ratio λ by estimating a reduced form model that comes out from substituting equation (2) in (1).

Using assumptions *(A.4')*, *(A.5')* and *(A.6)* then the reduced form model is a Generalized Additive Partial Linear Model (GAPLM)

$$E[\zeta|X=x, Z=z] = F\left\{\beta^T s_i + \sum_{j=1}^{d} \eta_j(t_{ij}) + \sum_{k=1}^{r} \varphi_k(z_{ik})\right\} \tag{12}$$

The parameter vector β, and the nonparametric additive components $\{\eta_j(t)\}_{j=1}^d$ and $\{\varphi_k(z)\}_{k=1}^r$ are estimated by the method proposed by Härdle, Huet, Mammen and Sperlich (1998). To make the exposition more clear, in equation (12) we can rewrite $\sum_{j=1}^d \eta_j(t_{ij}) + \sum_{k=1}^r \varphi_k(z_{ik})$ as $\sum_{l=1}^{d+r} \psi_l(v_{il})$. The probably so far best known nonparametric estimation procedure for these models has been the backfitting algorithm, see Hastie and Tibshirani (1990). An alternative is the so-called marginal integration estimator. The idea of marginal integration was first introduced by Tjøstheim and Auestad (1994) and Linton and Nielsen (1995). To estimate a particular additive component ψ_k one uses a multidimensional preliminar estimator and then integrates out all covariates v_j except v_k.

To get the pre-estimate of ψ, we use the method of Severini and Staniswalis (1994) which considered a model of the form (12). Their approach is based on an iterative application of smoothed local and un-smoothed global likelihood functions. In particular, this method allows an $\sqrt{n}$ estimation of the parametric component. Afterwards, we apply the integration idea on ψ to obtain estimates for ψ_j, $j = 1, \ldots, d+r$ (see Härdle, Huet, Mammen and Sperlich, 1998).

The reason why we prefer in our application the marginal integration approach to the backfitting procedure is that the first mentioned is indeed estimating the marginal effect of the particular input variable, even when the assumption of additivity is violated. In contrast, the backfitting is looking for an optimal fit of the regression problem and thus the estimation of marginal effects can be hard to interprete if strong separability of the input variables is not valid (cf. Sperlich, Linton, Härdle, 1997). Further, no theory for GAPLM with the backfitting algorithm has been developed until recently, see e.g. Linton, Mammen and Nielsen (1997).

Given (s_i, v_i), with $\mu_i = F\{s_i^T\beta + \alpha + \psi_1(v_{i1}) + \ldots + \psi_{d+r}(v_{id+r})\}$ the (conditional) likelihood of y_i for the binary case is given by

$$Q(\mu_i; y_i) = y_i \log \mu_i + (1-y_i)\log(1-\mu_i) \ , \tag{13}$$

and the conditional likelihood function

$$\mathcal{L}(\psi^+, \beta) = \sum_{i=1}^{n} Q(\mu_i; y_i) \ , \tag{14}$$

where $\psi^+(v)$ is the additive function $\alpha + \psi_1(v_1) + \ldots + \psi_{d+r}(v_{d+r})$.

Without loss of generality, we describe now how to estimate the component ψ_1. For a vector $u \in I\!R^{d+r}$ we denote the vector $(u_2, \ldots, u_{d+r})^T$ by $u_{\underline{1}}$, respectively $v_{i\underline{1}} = (v_{i2}, \ldots, v_{id+r})^T$. Further, for a kernel function L defined on $I\!R^{d+r-1}$ we put $L_g(v) = g^{-(d+r-1)} L(g^{-1}v)$ and for a kernel function K defined on $I\!R$ we put $K_h(v) = h^{-1}K(h^{-1}v)$. For L we take the product kernel $L = \prod_{j=2}^{d+r} L_j$. The bandwidth g is related to the smoothing in direction of the nuisance covariates, the bandwidth h to the direction of interest (here direction 1). We also make use of the smoothed likelihood defined by

$$\mathcal{L}^S(\psi^+, \beta) = \int \sum_{i=1}^n K_h(v_1 - v_{i1}) L_g(v_{\underline{1}} - v_{i\underline{1}}) Q\left[F\left\{s_i^T\beta + \psi^+(v)\right\}; y_i\right] dt. \tag{15}$$

Following Severini and Staniswalis (1994) and Härdle, Huet, Mammen and Sperlich (1998) we put for $\beta \in B \subset I\!R^p$

$$\widehat{\psi}_\beta(v) = \arg\min_\psi \sum_{i=1}^n K_h(v_1 - v_{i1}) L_g(v_{\underline{1}} - v_{i\underline{1}}) Q\left[G\{s_i^T\beta + \psi\}; y_i\right], \tag{16}$$

$$\widehat{\beta} = \arg\min_{\beta \in B} \mathcal{L}(\widehat{\psi}_\beta, \beta), \tag{17}$$

$$\widehat{\psi} = \widehat{\psi}_{\widehat{\beta}}. \tag{18}$$

The first equation (16) can also be written as $\widehat{\psi}_\beta = \arg\min_\psi \mathcal{L}^S(\psi, \beta)$. The estimate $\widehat{\psi}$ is a multivariate kernel estimate of ψ^+ which makes no use of the additive structure but serves as a pre-estimate as mentioned above. This procedure can be performed by an iterative alternating Newton Raphson type algorithm.

We now apply the marginal integration method. Because of the identifiability conditions, $\psi_1(v_1)$ is equal to $\int w_{\underline{1}}(v)\psi^+(v_1, v)dv$ or up to a constant, where $w_{\underline{1}}$ is a weight function verifying $\int w_{\underline{1}}(v)dv = 1$.

Thus, we put

$$\widetilde{\psi}_1(v_1) = \frac{\frac{1}{n}\sum_{i=1}^n w_{\underline{1}}(v_{i\underline{1}})\widehat{\psi}(v_1, v_{i\underline{1}})}{\frac{1}{n}\sum_{i=1}^n w_{\underline{1}}(v_{i\underline{1}})}. \tag{19}$$

and by centering $\widetilde{\psi}_1$ to zero, we get an estimate for ψ_1. Here again, introduction of a weight function w_1 may be useful to avoid problems at the boundary. The additive constant α is simply estimated by

$$\hat{\alpha} = \frac{1}{n}\sum_{i=1}^n \widehat{\psi}(v_i). \tag{20}$$

The final additive estimate of $\psi^+(v)$ will then be given by $\hat{\alpha} + \widehat{\psi}_1(v_1) + \ldots + \widehat{\psi}_d(v_d)$.

In the paper of Härdle, Huet, Mammen and Sperlich (1998) the consistency for these estimators (including $\hat{\beta}$) is proved and the asymptotic distribution is developed. Since in the next subsection we will present some asymptotic results about the two step semiparametric estimator, we introduce the following assumptions about the behavior of the kernel and the bandwidths

(A.9) The kernel L is a product kernel $L(\nu) = L_1(\nu_1) \times ... \times L_{d-1}(\nu_{d-1})$. The kernels L_j are symmetric probability densities with compact support. The kernel K is a symmetric probability density with compact support. Moreover they are both twice continuously differentiable.

(A.10) The bandwidths h and g tend to zero and $Nh^{2(d-1)}(\log N)^{-2}$tends to infinity.

3.2 Second step: Semiparametric least squares estimation

The log-wage equation can be estimated recalling that under the previous assumptions

$$E[W|\zeta_i = 1, Z = z_i, X = x_i] = \gamma_1^T z_i + \gamma_2 \lambda \left(s_i^T \beta^* + \sum_{l=1}^{d+r} \psi_l(v_{il}) \right) \qquad i = 1, \cdots, M,$$

where λ_i stands for the Mill's ratio. Then,

$$w_i = \gamma_1^T z_i + \gamma_2 \lambda \left(s_i^T \beta^* + \sum_{l=1}^{d+r} \psi_l(v_{il}) \right) + \epsilon_i \qquad i = 1, \cdots, M, \tag{21}$$

where $\epsilon_i = w_i - E[W|\zeta_i = 1, Z = z_i, X = x_i]$. In the previous equation, the estimation of the parameter vector $\gamma^T = (\gamma_1^T \quad \gamma_2)^T$ is unfeasible since the parameters of the index, β^* and $\{\psi_l(\bullet)\}_{l=1}^{d+r}$ are unknown. In this case, β^* stands for the parameter vector β with the normalization $\sigma_1^2 + V(\varphi_{k+1}(u_{2i})) = 1$. Heckman (1979), proposed in a fully parametric setting, to replace them by consistent estimators, proceeding in the same way then obtain the following regression equation

$$w_i = \gamma_1^T z_i + \gamma_2 \lambda \left(s_i^T \hat{\beta}^* + \sum_{l=1}^{d+r} \hat{\psi}_l(v_{il}) \right) + \epsilon_i \qquad i = 1, \cdots, M. \tag{22}$$

The parameter vector $(\gamma_1^T \quad \gamma_2)^T$ can be estimated by ordinary least squares. In this case, $\hat{\gamma} = (\hat{\gamma}_1^T \quad \hat{\gamma}_2)^T$ must fulfill the following set of conditions

$$\frac{1}{N} \sum_{i=1}^{N} g\left(\tilde{z}_i; \hat{\beta}, \hat{\psi}(v_i), \gamma\right) = -\frac{2}{N} \sum_{i=1}^{N} \zeta_i \tilde{z}_i \left(w_i - \tilde{z}_i^T \gamma\right) = 0 \tag{23}$$

where $\tilde{z}_i^T = \left[z_i^T \quad \lambda\left(s_i^T \hat{\beta}^* + \sum_{l=1}^{d+r} \hat{\psi}_l(v_{il})\right) \right]$, $\hat{\psi}(v_i) = \sum_{l=1}^{d+r} \hat{\psi}_l(v_{il})$ and $\lambda(\bullet) = f(\bullet)/F(\bullet)$ is the Mill's ratio. We show in the Appendix the following asymptotic result for $\hat{\gamma}$.

Theorem 2 *Under the assumptions (A.1)-(A.3), (A.4')-(A.6'), and (A.7)-(A.10) then*

$$\sqrt{N}\left(\hat{\gamma}-\gamma_0\right) \xrightarrow{D} N\left(0, M^{-1} S M^{-1}\right),$$

where

$$M = E\left[\nabla_\gamma g\left(x, z; \beta_0, \psi_0\left(\nu\right), \gamma_0\right)\right],$$

and

$$S = E\left\{g\left(x, z; \beta_0, \psi_0\left(\nu\right), \gamma_0\right) - E\left[g\left(x, z; \beta_0, \psi_0\left(\nu\right), \gamma_0\right)\right]\right\} \times \\ \left\{g\left(x, z; \beta_0, \psi_0\left(\nu\right), \gamma_0\right) - E\left[g\left(x, z; \beta_0, \psi_0\left(\nu\right), \gamma_0\right)\right]\right\}^T.$$

Moreover, if

$$\hat{M}_N = \frac{1}{N}\sum_{i=1}^{N}\left\{\nabla_\gamma g\left(x_i, z_i; \hat{\beta}, \hat{\psi}\left(\nu_i\right), \hat{\gamma}\right)\right\}\left\{\nabla_\gamma g\left(x_i, z_i; \hat{\beta}, \hat{\psi}\left(\nu_i\right), \hat{\gamma}\right)\right\}^T$$

and

$$\hat{S}_N = \frac{1}{N}\sum_{i=1}^{N}\left\{g\left(x_i, z_i; \hat{\beta}, \hat{\psi}\left(\nu_i\right), \hat{\gamma}\right) - \frac{1}{N}\sum_{j=1}^{N} g\left(x_j, z_j; \hat{\beta}, \hat{\psi}\left(\nu_j\right), \hat{\gamma}\right)\right\} \times \\ \left\{g\left(x_i, z_i; \hat{\beta}, \hat{\psi}\left(\nu_i\right), \hat{\gamma}\right) - \frac{1}{N}\sum_{j=1}^{N} g\left(x_j, z_j; \hat{\beta}, \hat{\psi}\left(\nu_j\right), \hat{\gamma}\right)\right\}^T$$

then

$$\hat{M}_N^{-1}\hat{S}_N\hat{M}_N^{-1} \longrightarrow_p M^{-1} S M^{-1}$$

as N tends to infinity.

3.3 Third step: Tobit local maximum likelihood estimator

In this section, we replace all the wages for the predicted wages obtained from the previous step. i.e. $w_i = \hat{\gamma}^T z_i$ for $i = 1, ..., N$. Let us consider the structural model that is represented in equation (10) and in addition consider the previous set of assumptions. Then the likelihood function for the i-th observation in this model is

$$\ln L = \frac{-\ln(2\pi)}{2} + \frac{-\ln(\sigma^2)}{2} - \frac{\{y_i - \gamma^T s_i - \eta(t_i)\}^2}{2\sigma^2} - \ln\left\{1 - F\left(\frac{-\gamma^T s_i - \eta(t_i)}{\sigma}\right)\right\}.$$

Remark: In the set of continuous explanatory variables t_i we have included the predicted wages $\hat{w}_i$.

We estimate γ, σ and η in the following way:

1. $\hat{\eta}_{\gamma,\sigma} = \arg\max_{\eta} \mathcal{L}^S\left(\eta, \gamma, \sigma\right)$,

2. $\widehat{(\gamma, \sigma)} = \arg\max_{\gamma,\sigma} \mathcal{L}\left(\hat{\eta}_{\gamma,\sigma}, \gamma, \sigma\right)$,

3. $\hat{\eta} = \hat{\eta}_{\hat{\gamma},\hat{\sigma}}$,

where

$$\mathcal{L}(\eta,\gamma,\sigma) = \sum_{i=1}^{n} \frac{-\ln(2\pi)}{2} + \frac{-\ln(\sigma^2)}{2} - \frac{\{Y_i - \gamma^T s_i - \eta(t_i)\}^2}{2\sigma^2} - \ln\left\{1 - F\left(\frac{-\gamma^T s_i - \eta(t_i)}{\sigma}\right)\right\}, \tag{24}$$

$$\mathcal{L}^S(\eta(\cdot),\gamma,\sigma) = \int \sum_{i=1}^{n} \left[\frac{-\ln(2\pi)}{2} + \frac{-\ln(\sigma^2)}{2} - \frac{\{Y_i - \gamma^T s_i - \eta(t)\}^2}{2\sigma^2} - \ln\left\{1 - F\left(\frac{-\gamma^T s_i - \eta(t)}{\sigma}\right)\right\}\right] K_h(t - t_i)dt. \tag{25}$$

Remark: Since by assumption *(A.4')* the function $\eta(\bullet)$ in (24) is unknown, in order to estimate it we use the local likelihood method proposed by Staniswalis (1989).
To apply the Newton Raphson Algorithm we have to determine the gradients and Hesse-matrices. In Appendix II, we give the explicit expressions for the algorithm.

4 The Application

4.1 Model and Data

The source of the data is the *Encuesta de Población Activa (EPA)* , the Spanish Labor Force Surveys. These surveys have been carried out on a quarterly basis since 1975 and are collected by the National Bureau of Statistics (INE). They cover approximately 60,000 households and contain information about 150,000 individuals that are older than 16 years. It provides information at different levels of disaggregation both at national and regional level. From these surveys, in the second quarter of 1990, the National Bureau of Statistics randomly selected a cross-section of 4,989 individuals and additional information about some variables that were considered relevant for labor market participation analysis were provided. In this paper we consider a subsample of 1010 individuals participating in the labor market, 612 workers and 398 non workers.

The variables included in this data set are defined in Table 1 including some basic statistics.

Further, we certainly have the information whether a person has a job (JOB) or not.

In this application we are considering the problem of estimating the conditional expectation of being employed. As discussed in the introduction, we have two problems; for including the wages we have to predict them for non workers, and additionally, estimating

Variable	Description	Whole Sample	Workers
SEXM	dummy, 1 if male	0.680 (0.466)	0.625 (0.484)
AGE1	dummy, age 16 to 19	0.131 (0.338)	0.111 (0.314)
AGE2	dummy, age 20 to 25	0.265 (0.441)	0.256 (0.437)
AGE3	dummy, age 26 to 35	0.278 (0.448)	0.261 (0.439)
AGE4	dummy, older than 45	0.138 (0.345)	0.143 (0.351)
EDUC1	elementary school	0.350 (0.477)	0.339 (0.474)
EDUC2	high school	0.115 (0.320)	0.106 (0.308)
EDUC3	university	0.064 (0.245)	0.039 (0.194)
URATE	unemployment rate of the district	0.171 (0.069)	0.171 (0.071)
SINGLE	dummy, 1 if single	0.689 (0.463)	0.725 (0.446)
NOHH	dummy, 1 if person is not head of household	0.703 (0.456)	0.616 (0.486)
WAGE	earnings per hour	292.735 (313.237)	483.108 (264.402)

Table 1: *Comparative Statistics of the explanatory variables; mean and standard deviation (in brackets).*

the wage equation we are touched by the sample selection problem. Therefore we apply the three step Heckman estimation procedure.

We did the estimation for two competing models, a standard parametric (Model I) and a semiparametric one (Model II), as we described them in section 3. The only difference between the models is that in Model II we allow for nonlinearities of in the influence of the continuous variables URATE in step 1 and URATE, ln(WAGE) in step 3.

We proceed as follows. We regress in step

1. the variable JOB against AGE1, AGE2, AGE3, AGE4, EDUC1, EDUC2, EDUC3, SEXM, SINGLE, NOHH, SEXM*SINGLE, URATE and a constant (CONST) by

$$\text{(Model I)} \qquad E[\zeta|X=x] = F\left(x^T\beta_I\right) \text{ , respectively by}$$
$$\text{(Model II)} \qquad E[\zeta|S=s,T=t] = F\left\{s^T\beta_{II} + \eta(t)\right\},$$

with $X = (S^T, T)^T$ denoting all input variables, T =URATE;

2. ln(WAGE) against AGE1, AGE2, AGE3, AGE4, EDUC1, EDUC2, EDUC3, SEXM, SEXM*SINGLE, URATEthe Mills ratio and a constant (CONST) by

$$E[\ln(\text{WAGE})|X = x, \Lambda = \lambda] = x^T\beta + \alpha\lambda,$$

again with X denoting all input variables and $\Lambda\lambda$ the Mills ratio;

3. JOB against AGE1, AGE2, AGE3, AGE4, EDUC1, EDUC2, EDUC3, SEXM, SINGLE, NOHH, URATE, ln(WAGE) and (CONST) by

$$\begin{array}{ll} \text{(Model I)} & E[\zeta|X = x] = F\left(x^T\beta_I\right) \text{ , respectively by} \\ \text{(Model II)} & E[\zeta|S = s, T = t] = F\left\{s^T\beta_{II} + \eta_1(t_1) + \eta_2(t_2)\right\}, \end{array}$$

$X = (S^T, T)^T$ denoting all input variables, t_1 =URATE, t_2 =ln(WAGE).

In step 1 and 3 we had $n = 1010$ observations, in step 2 $n = 612$, see Table 1.

In step 2 we regress ln(WAGE) versus ln(X) since we are interested in the partial increase proportional to wage, i.e. $\frac{\partial \text{WAGE}}{\partial X_j}\frac{1}{\text{WAGE}} = \frac{\Delta \ln(\text{WAGE})}{\Delta X_j}$.

We also compare the two cases of (a) taking in step 3 predicted wages for all persons in the sample versus (b) taking the real (observed) wages for workers. Please note that in the analytic part of our paper we worked out explicitly the first case whereas the second can be done straight forward.

For the nonparametric estimation we applied in all steps the quartic kernel, $K(u) = \frac{15}{16}(1-u^2)^2\ I\{|u| \le 1\}$. Due to the necessary undersmoothing discussed in section 3, in the first step bandwidth $h = 0.8*$ stdev(URATE) has been used, where stdev($\cdot$) is the standard deviation of the corresponding input. In step three we used $h_1 = (1.5, 1.75)*$ stdev(URATE,ln(WAGE)) when estimating the parametric (linear) part and $h_2 = (1.0, 1.25)*$ stdev(URATE,ln(WAGE)), $g = h_1$ to estimate η_1, η_2. For the explanation of h and g, see section 3.

In Table 2 we present the results the estimation results for Model I and II in the first two steps. For the nonparametric part in Model I see Figure 1, (f1,1). Because step 1 is only for determining the Mills ratio, we skip a detailed discussion of its numerical results. We only notice that all coefficients have the expected sign due to economic theory, the results for the dummy variables between Model I and II differ mainly concerning the significance but not much in the values of coefficients. Further, the influence of URATE seems to be strongly nonlinear.

	step 1		step 2	
Variable	Model I	Model II	Model I	Model II
Constant	1.034 (0.213**)	1.003 -	6.305 (0.074**)	6.393 (0.065**)
AGE1	−0.520 (0.183**)	−0.450 (0.112**)	−0.302 (0.097**)	−0.329 (0.097**)
AGE2	−0.267 (0.163*)	−0.259 (0.099**)	0.046 (0.075)	0.044 (0.060)
AGE3	−0.188 (0.146*)	−0.192 (0.088**)	0.006 (0.067)	0.010 (0.052)
AGE4	−0.439 (0.161**)	−0.443 (0.096**)	0.071 (0.070)	0.068 (0.055*)
EDUC1	0.034 (0.105)	−0.033 (0.066)	−0.027 (0.051)	−0.016 (0.046)
EDUC2	−0.118 (0.142)	−0.178 (0.090**)	0.210 (0.070**)	0.214 (0.063**)
EDUC3	−0.531 (0.182**)	−0.510 (0.115**)	0.645 (0.109**)	0.601 (0.101**)
SEXM	−0.219 (0.181)	−0.247 (0.112**)	0.081 (0.075)	0.060 (0.065)
SINGLE	0.696 (0.199**)	0.690 (0.120**)	−	−
NOHH	−0.895 (0.130**)	−0.910 (0.078**)	−	−
SEXM*SINGLE	−0.102 (0.212)	−0.089 (0.131)	−1.111 (0.066**)	−0.093 (0.059*)
U-RATE	−0.504 (0.614)	−	0.073 (0.280)	0.071 (0.215)
Mills	−	−	−0.480 (0.116**)	−0.299 (0.062**)

Table 2: *Estimation results for step 1 and 2 for the parametric part. Standard deviations are given in brackets. Asterisks indicate significance at 10 (*), respectively 5 (**) percent level.*

For the wage equation (step 2) we again have quite similar results for the different Models, except for the Mills ratio. Notice that age, education, sex and family status have significant influence with expected signs. They confirm that very young age and low education level have a negative influence on the earnings per hour.

The fact that URATE is perfectly insignificant could indicate that pay policy and wage negotiations are still nationwide in Spain and are not affected by the labor market in the particular district. The Mills ratio in both models is strongly significant, so we indeed deal with a big selection bias in the wage equation. For the semiparametric case (Model II) it is much smaller.

Unfortunately, in both models the wage regression have a R^2 of only about 21%, i.e. we are not explaining much of the variance of ln(WAGE). This lead us to the comparison of two cases ((a) and (b)) in step 3. In Figure 1 we have plotted the estimates of the influence functions for URATE, and ln(WAGE) for both cases. The problem in case (b), when we take the observed wages for workers and predicted wages only for the non workers is, that due to the small R^2 we have predicted wages only in a much smaller (but high-level) range than the range is for observed wages.

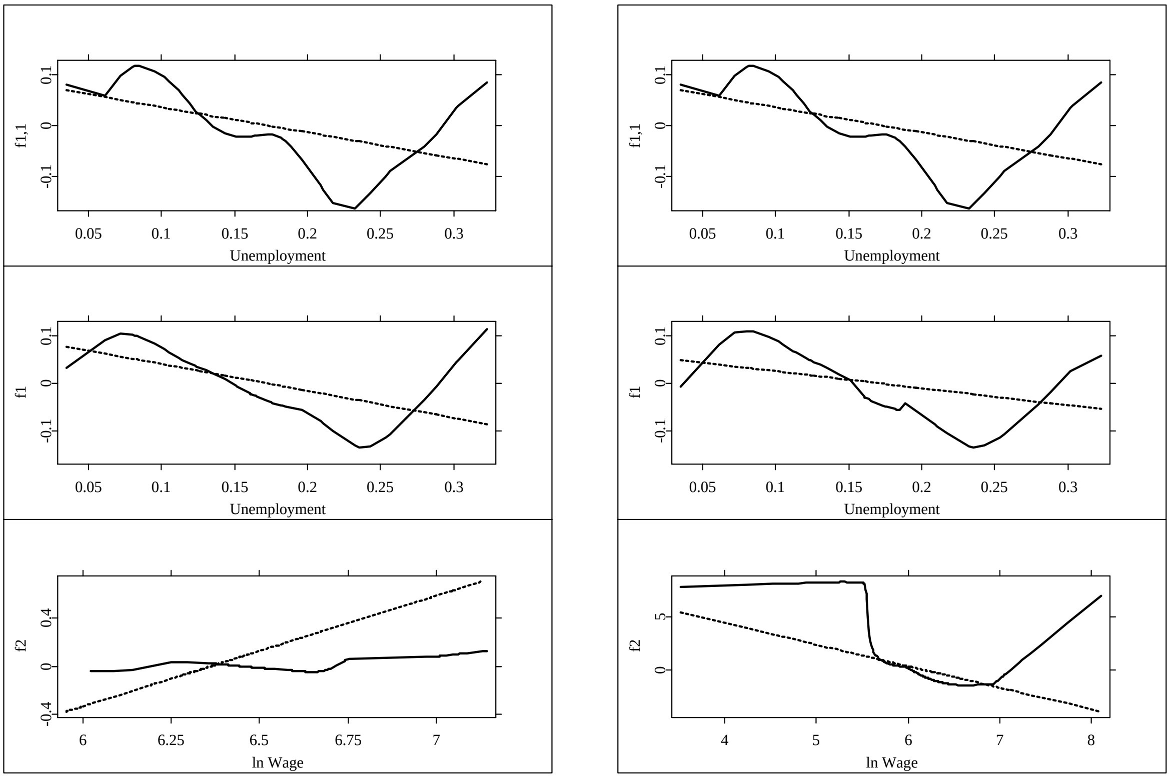

Figure 1: *Nonparametric and parametric estimates for URATE and ln(WAGE). At top (f1,1) for step 1, below (f1,f2) for step 3, left case (a), right case (b).*

Consequently all small wages in that sample belong to workers and vice versa we get a strongly negative estimate for the influence of ln(WAGE) on having a job. Therefore, and to be consistent in the inputs, we rely more on case (a) where we take the predicted wages for all people and thus avoid the problem of having two quite different variations in the same predictor variable.

In Table 3 we show the estimation results in step 3, case (a) for the parametric part. First, to manifest the difference between a standard probit, as often done in the economic literature, and the probit with corrected standard deviations, we present both for Model I. We can see that the corrected ones are about 5 to 15% bigger than the uncorrected ones.

The coefficient estimates for Model I and II again are quite similar. Certainly, including

Variable	Model I	Model I*	Model II
Constant	-4.715 (14.001)	(12.133)	1.044 -
AGE1	-0.244 (0.829)	(0.615)	-0.477 (0.133**)
AGE2	-0.310 (0.201*)	(0.181**)	-0.257 (0.098**)
AGE3	-0.193 (0.168)	(0.145*)	-0.198 (0.088**)
AGE4	-0.504 (0.213**)	(0.204**)	-0.444 (0.096**)
EDUC1	0.059 (0.134)	(0.118)	-0.035 (0.066)
EDUC2	-0.309 (0.474)	(0.421)	-0.185 (0.097**)
EDUC3	-1.119 (1.351)	(1.228)	-0.660 (0.260**)
SEXM	-0.292 (0.123**)	(0.094**)	-0.308 (0.058**)
SINGLE	0.696 (0.190**)	(0.199**)	0.621 (0.067**)
NOHH	-0.895 (0.130**)	(0.130**)	-0.894 (0.075**)
ln(WAGE)	0.912 (2.209)	(1.907)	–
U-RATE	-0.570 (0.652)	(0.631)	–

Table 3: *Estimation results for step 3, case (a) for the parametric part. Standard deviations are given in brackets. Column Model I* is giving the uncorrected standard deviations. Asterisks indicate significance at 10 (*), respectively 5 (**) percent level. For Model II we give here the uncorrected standard deviations neglecting the first two steps.*

ln(WAGE), the significance for the other explanatory variables is shrinking (compare step 1) in the parametric Model I. But still age (AGE2, AGE4), SEX, SINGLE and NOHH is highly significant with expected signs. URATE is only significant at a level of about 21%. Notice that this statement holds only for the linear influence of URATE. Looking at Figure 1, we see a clearly nonlinearity for URATE while for ln(WAGE) insignificance seems to be real and not just caused by a misspecification of its functional form.

Nevertheless, zooming this result, see Figure 2, we get the impression that in Spain there is a small upper-middle class. We have many jobs where people earn a small salary, but for some people with the adequate abilities to earn a lot there is an increasing probability to get such a job.

Looking at the estimated influence of URATE, the functional form is a little bit harder to

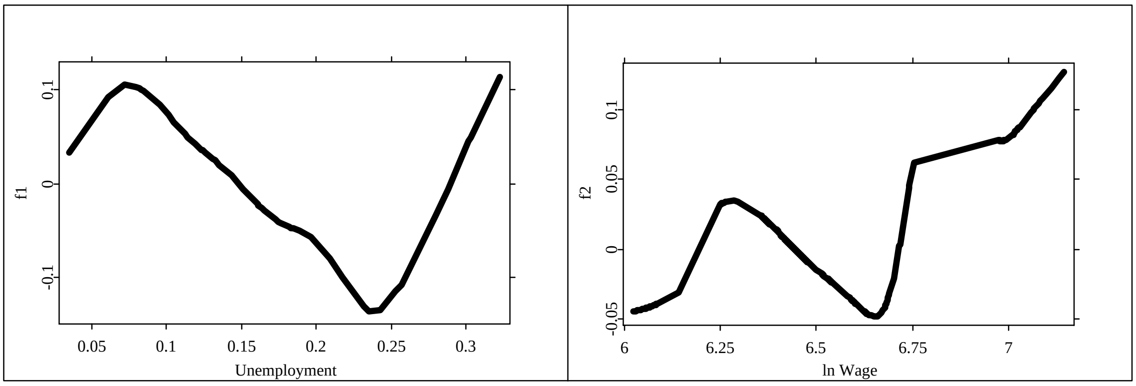

Figure 2: *Nonparametric estimates for URATE and ln(WAGE) for step 3, case (a).*

understand. Let E be the employed people, LF the labor force (all participants), U the unemployed people, u the percentage of U to LF, and P the probability to be employed (that is what we estimate).

We have $P = E/LF = E/(E+U) = E/(E+u*LF)$ and consider

$$\frac{\partial P}{\partial u} = \frac{-E*LF}{(E+u*LF)^2} ,$$

compare with parametric results in Table 3.

Now we consider this with respect to a possibly changing Labor Force. This consideration makes sense, since there are many situations where at the same time many people give up to register, e.g. in a recession when nobody is believing in his chance.

$$\frac{\partial P/\partial u}{\partial LF} = \frac{(-E(E+u*LF)+2E*u*LF)}{(E+u*LF)^3} .$$

Here you can see how for certain changes in LF or u this probability P can be positive, respectively negative. This way we can understand the nonlinear function, especially the increase for high unemployment u.

5 Conclusions

The purpose of this paper has been to provide an extension of specification and estimation a standard model of labor supply which is known as the three step Heckman algorithm. It can thus be included in the so called predicted wage methods.

We first give consistent estimates for the standard deviations in the known parametric standard model for the third step. We then introduce an alternative model specification using latest developments from non- and semiparametric statistics. We present the estimation procedure for all steps and consistent estimates for the standard deviations of the parametric part in the first and second step.

In a detailed application example we demonstrate the handling and performance of our new estimation method. We further compare the results with the standard parametric approach, and discuss the differences and possible ways of interpretation.

Appendix I: Proof of Theorems 1 and 2

Proof of Theorem 1

Before to prove the results we introduce the following lemma.

Lemma 1 (Newey and McFadden, 1994) *If z_i is i.i.d., $a(z,\theta)$ is continuous at θ_0 with probability one, and there is a neighborhood N of θ_0 such that $E\left[\sup_{\theta\in N}\|a(z,\theta)\|\right]<\infty$, then for any $\tilde{\theta}\rightarrow_p\theta_0$, $n^{-1}\sum_{i=1}^{n}a\left(z_i,\tilde{\theta}\right)\rightarrow_p E\left[a(z,\theta_0)\right]$.*

The proof can be found in Newey and McFadden (1994), p. 2156.

We will show first the convergence in distribution result. In order to do so, first note that

$$\sqrt{N}\left(\hat{\delta}-\delta_0\right)=\frac{1}{\sqrt{N}}\sum\bar{\psi}\left(x_i,z_i\right) \tag{26}$$

where $\bar{\psi}(x_i,z_i)=-\left[\frac{1}{N}\sum_{i=1}^{N}\nabla_\delta m\left(x_i,z_i,\bar{\delta}\right)\right]^{-1}m(x_i,z_i)$ and $\bar{\delta}$ is a mean value. This is because $\hat{\delta}$ makes (6) equal to zero, with probability one, and the mean value theorem. Furthermore, Assumptions (A.1) to (A.6) ensure both that $\hat{\delta}$ is a consistent estimator for δ_0 (see Newey and McFadden, 1994: Theorem 2.5, p. 2131), and $E\left[\sup_{\delta\in N_\delta}\|m(x,z;\delta)\|\right]<\infty$, therefore, Lemma 1 applies for $\bar{\delta}$ between $\hat{\delta}$ and and δ_0 and then

$$\sqrt{N}\left(\hat{\delta}-\delta_0\right)=\frac{1}{\sqrt{N}}\sum\psi\left(x_i,z_i\right)+o_p(1) \tag{27}$$

where now, $\psi(x_i,z_i)=-\left[E\left[\nabla_\gamma m(x,z)\right]\right]^{-1}m(x_i,z_i)$. Proceeding in the same way as before for equation (8), it is also possible to show that

$$\begin{aligned}\sqrt{N}\left(\hat{\gamma}-\gamma_0\right)=-\left[\tfrac{1}{N}\sum_{i=1}^{N}\nabla_\gamma g\left(x_i,z_i,\hat{\delta},\bar{\gamma}\right)\right]^{-1}\times\\ \left\{\tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}g(x_i,z_i)+\left[\tfrac{1}{N}\sum_{i=1}^{N}\nabla_\delta g\left(x_i,z_i,\bar{\delta},\gamma_0\right)\right]\sqrt{N}\left(\hat{\delta}-\delta_0\right)\right\}\end{aligned} \tag{28}$$

But then, if we substitute (27) into (28) and we apply again Lemma 1 we obtain

$$\sqrt{N}\left(\hat{\gamma}-\gamma_0\right)=G_\gamma^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{g\left(x_i,z_i\right)+G_\delta\psi\left(x_i,z_i\right)\right\}+o_p(1) \tag{29}$$

The same can be done for equation (11) and then

$$\sqrt{N}\left(\hat{\tau}-\tau_0\right)=-H_{N\tau}\left(\hat{\delta},\hat{\gamma},\bar{\tau}\right)^{-1}\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}h\left(x_i,z_i\right)-H_{N\tau}\left(\hat{\delta},\hat{\gamma},\bar{\tau}\right)^{-1}H_{N\delta}\left(\bar{\delta},\gamma_0,\tau_0\right)\sqrt{N}\left(\hat{\delta}-\delta_0\right)$$

$$-H_{N\tau}\left(\hat{\delta},\hat{\gamma},\bar{\tau}\right)^{-1}H_{N\gamma}\left(\delta_0,\bar{\gamma},\tau_0\right)\sqrt{N}\left(\hat{\gamma}-\gamma_0\right)$$

$$-H_{N\tau}\left(\hat{\delta},\hat{\gamma},\bar{\tau}\right)^{-1}H_{N\delta\gamma}\left(H_{N\gamma}\left(\delta_0,\bar{\gamma},\tau_0\right)\right)N\left(\hat{\delta}-\delta_0\right)\left(\hat{\gamma}-\gamma_0\right) \tag{30}$$

where

$$H_{N\tau}\left(\hat{\delta},\hat{\gamma},\bar{\tau}\right)=\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\nabla_\tau h\left(x_i,z_i,\hat{\delta},\hat{\gamma},\bar{\tau}\right)$$

$$H_{N\delta}\left(\bar{\delta},\gamma_0,\tau_0\right)=\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\nabla_\delta h\left(x_i,z_i,\bar{\delta},\gamma_0,\tau_0\right)$$

$$H_{N\gamma}\left(\delta_0,\bar{\gamma},\tau_0\right)=\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\nabla_\gamma h\left(x_i,z_i,\delta_0,\bar{\gamma},\tau_0\right)$$

$$H_{N\delta\gamma}\left(\bar{\delta},\bar{\gamma},\tau_0\right)=N^{-\frac{3}{2}}\textstyle\sum_{i=1}^{N}\nabla_\delta\nabla_\gamma h\left(x_i,z_i,\bar{\delta},\bar{\gamma},\tau_0\right)$$

Now, substituting back both equations (27) and (29) into (30) and applying Lemma 1 then we obtain

$$\sqrt{N}\left(\hat{\tau}-\tau_0\right)=-H_\tau^{-1}$$

$$\times\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\left\{h\left(x_i,z_i\right)+H_\delta\psi\left(x_i,z_i\right)+H_\gamma G_\gamma^{-1}\left(g\left(x_i,z_i\right)+G_\delta\psi\left(x_i,z_i\right)\right)\right\}+o_p(1) \tag{31}$$

and taking into account assumption (A.2), we can apply both the Lindeberg-Levy CLT and the Slutzky theorem (see Serfling, 1980: pp.19 and 28) and the proof is done.

The proof of consistency of the variance covariance matrix $\hat{V}\left(\hat{\tau}\right)$ is immediate by applying Lemma 1 to the different terms.

□

Proof of Theorem 2

In order to develop the proof of this theorem we need two previous lemmas.

Lemma 2 *Assume conditions (A.1) to (A.3), (A.4') to (A.7'), and (A.8)-(A.10) hold, then*

$$\nu_N(\beta,\psi) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} g(x_i, z_i; \beta, \psi(v_i), \gamma_0) - \frac{1}{\sqrt{N}}\sum_{i=1}^{N} E\left[g(x_i, z_i; \beta, \psi(v_i), \gamma_0)\right] \tag{32}$$

is stochastically equicontinuous at (β_0, ψ_0).

Proof of Lemma 2

In order to show stochastic equicontinuity, we must prove that for all $\epsilon > 0$ and $\eta > 0$, there exists $\delta > 0$ such that

$$\overline{\lim}_{N\longrightarrow\infty}\mathbf{P}\left(\sup_{\rho((\beta,\psi),(\beta_0,\psi_0))<\delta} \left|\nu_N(\beta,\psi) - \nu_N(\beta_0,\psi_0)\right| > \eta\right) < \epsilon \tag{33}$$

Where the supremum is taken over $(\beta,\psi) \in I\!B \times \Psi$ for the metric $\rho((\beta,\psi),(\beta_0,\psi_0)) = \|\beta-\beta_0\|_S + \sup_v |\psi(v) - \psi_0(v)|$. Substituting (32) into (33) and rearranging terms we have that

$$\begin{aligned}\nu_N(\beta,\psi) - \nu_N(\beta_0,\psi_0) &= \tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\left\{g(x_i,z_i;\beta,\psi(v_i),\gamma_0) - g(x_i,z_i;\beta_0,\psi_0(v_i),\gamma_0)\right\}\\ &+\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N} E\left\{g(x_i,z_i;\beta,\psi(v_i),\gamma_0) - g(x_i,z_i;\beta_0,\psi_0(v_i),\gamma_0)\right\},\end{aligned} \tag{34}$$

then by the triangle inequality

$$\mathbf{P}\left(\sup\left|\nu_N(\beta,\psi) - \nu_N(\beta_0,\psi_0)\right| > \eta\right)$$

$$\begin{aligned}&\leq \mathbf{P}\left(\sup\left|\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\left\{g(x_i,z_i;\beta,\psi(v_i),\gamma_0) - g(x_i,z_i;\beta_0,\psi_0(v_i),\gamma_0)\right\}\right| > \eta\right)\\ &+\mathbf{P}\left(\sup\left|\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N} E\left\{g(x_i,z_i;\beta,\psi(v_i),\gamma_0) - g(x_i,z_i;\beta_0,\psi_0(v_i),\gamma_0)\right\}\right| > \eta\right)\end{aligned} \tag{35}$$

Taking a Taylor expansion around β_0 and ψ_0 and using Freshet derivatives as in Newey (1994) then it can be shown that

$$\begin{aligned}&\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\left\{g(x_i,z_i;\beta,\psi(v_i),\gamma_0) - g(x_i,z_i;\beta_0,\psi_0(v_i),\gamma_0) - G(\beta-\beta_0,\psi(v_i)-\psi_0(v_i))\right\}\\ &\leq C_1\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\left\{(\beta-\beta_0)^T s_i + \psi(v_i) - \psi_0(v_i)\right\}^2 \leq C_2\left\{\|\beta-\beta_0\|_S^2 + \left(\sup_v|\psi(v)-\psi_0(v)|\right)^2\right\}\end{aligned} \tag{36}$$

for some positive constants C_1 and C_2. Moreover, according to equation (23), $G\left(\beta-\beta_0, \hat{\psi}(v_i) - \psi_0(v_i)\right)$ is equal to

$$\begin{aligned}&\left\{-2\zeta_i\begin{bmatrix} z_i \\ \lambda_i' \end{bmatrix}\left[w_i - z_i^T\gamma_{10} - \gamma_{20}\lambda_i\right] + 2\zeta_i\begin{bmatrix} z_i \\ \lambda_i \end{bmatrix}\lambda_i'\gamma_{20}\right\} \times \left\{(\beta-\beta_0)^T s_i + \psi(v_i) - \psi_0(v_i)\right\}\\ &\equiv W_i(z_i,s_i,v_i)\left\{(\beta-\beta_0)^T s_i + \psi(v_i) - \psi_0(v_i)\right\}\end{aligned} \tag{37}$$

Notice that λ and λ' are evaluated at β_0 and ψ_0.

Since for $\delta > 0$ we have $\rho\left(\left(\hat{\beta},\hat{\psi}\right),(\beta_0,\psi_0)\right) < \delta$ then the left hand side of (35) is bounded by

$$\begin{aligned}(38)\quad &\mathbf{P}\left(\sup\left|\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N W_i\left(z_i,s_i,v_i\right)\left\{\left(\beta-\beta_0\right)^T s_i+\psi(v_i)-\psi_0(v_i)\right\}\right|>\eta\right)\\ &+\mathbf{P}\left(\sup\left|\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N E\left[W_i\left(z_i,s_i,v_i\right)\left\{\left(\beta-\beta_0\right)^T s_i+\psi(v_i)-\psi_0(v_i)\right\}\right]\right|>\eta\right),\end{aligned}$$

and,

$$\begin{aligned}(39)\quad &\overline{\lim}_{N\longrightarrow\infty}\mathbf{P}\left(\sup\left|\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N W_i\left(z_i,s_i,v_i\right)\left\{\left(\beta-\beta_0\right)^T s_i+\psi(v_i)-\psi_0(v_i)\right\}\right|>\eta\right)\\ &\leq\overline{\lim}_{N\longrightarrow\infty}\mathbf{P}\left(\left\|\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N W_i\left(z_i,s_i,v_i\right)\right\|>\tfrac{\eta}{\delta}\right)<\epsilon\end{aligned}$$

The same holds for the second term of equation (38). Under assumptions (A.1) and (A.2) then by the Levy CLT $\frac{1}{\sqrt{N}}\sum_{i=1}^N W_i\left(z_i,s_i,v_i\right)=O_p(1)$ and therefore by choosing δ small we close the proof.

□

The proof of this theorem is based in the results developed by Andrews (1994). In equation (23) using the mean value theorem we obtain

$$(40)\quad \sqrt{N}\left(\hat{\gamma}-\gamma_0\right)=-\left[\frac{1}{N}\sum_{i=1}^N\nabla_\gamma g\left(x_i,z_i,\hat{\beta},\hat{\psi}(v_i),\bar{\gamma}\right)\right]^{-1}\times\frac{1}{\sqrt{N}}\sum_{i=1}^N g\left(x_i,z_i;\hat{\beta},\hat{\psi}(v_i),\gamma_0\right),$$

where $\bar{\gamma}$ lies in the segment between γ_0 and $\hat{\gamma}$. Then by rearranging terms

$$\begin{aligned}\sqrt{N}\left(\hat{\gamma}-\gamma_0\right)=&-\left[\tfrac{1}{N}\textstyle\sum_{i=1}^N\nabla_\gamma g\left(x_i,z_i,\hat{\beta},\hat{\psi}(v_i),\bar{\gamma}\right)\right]^{-1}\times\left\{\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N g\left(x_i,z_i;\beta_0,\psi_0(v_i),\gamma_0\right)\right.\\ &\left.+\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N g\left(x_i,z_i;\hat{\beta},\hat{\psi}(v_i),\gamma_0\right)-\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^N g\left(x_i,z_i;\beta_0,\psi_0(v_i),\gamma_0\right)\right\}\end{aligned}\quad(41)$$

We claim the following results,

$$(42)\quad \frac{1}{N}\sum_{i=1}^N\nabla_\gamma g\left(x_i,z_i,\hat{\beta},\hat{\psi}(v_i),\bar{\gamma}\right)\longrightarrow_p M,$$

where $M=E\left[\nabla_\gamma g\left(x,z,\beta_0,\psi_0(v),\gamma_0\right)\right]$.

$$(43)\quad \frac{1}{\sqrt{N}}\sum_{i=1}^N g\left(x_i,z_i;\beta_0,\psi_0(v_i),\gamma_0\right)-\frac{1}{\sqrt{N}}\sum_{i=1}^N E\left[g\left(x_i,z_i;\beta_0,\psi_0(v_i),\gamma_0\right)\right]\Longrightarrow N\left(0,S\right),$$

as N tends to infinity.

$$(44)\quad \frac{1}{\sqrt{N}}\sum_{i=1}^N g\left(x_i,z_i;\hat{\beta},\hat{\psi}(v_i),\gamma_0\right)-\frac{1}{\sqrt{N}}\sum_{i=1}^N g\left(x_i,z_i;\beta_0,\psi_0(v_i),\gamma_0\right)\longrightarrow_p 0,$$

where S is the asymptotic variance. Under this results the theorem will have been proved since by (42) and (44) then (41) becomes

$$\begin{aligned}(45)\qquad \sqrt{N}\left(\hat{\gamma}-\gamma_0\right) &= M^{-1}\Big(o_p(1)+\lim \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big\{g\left(x_i,z_i;\beta_0,\psi_0(v_i),\gamma_0\right) \\ &\qquad -E\left[g\left(x_i,z_i;\beta_0,\psi_0(v_i),\gamma_0\right)\right]\Big\}\Big).\end{aligned}$$

Finally, substitute (43) into (45), apply the Slutzky theorem, and we obtain

$$\sqrt{N}\left(\hat{\gamma}-\gamma_0\right)\xrightarrow{D} N\left(0,M^{-1}SM^{-1}\right)$$

This closes the proof. Now we proceed to prove the results contained in equations (42), (43) and (44).

In order to show (44), we can rewrite the left hand side of this equation as equal to

$$(46)\qquad \nu_N\left(\hat{\beta},\hat{\psi}\right)-\nu_N\left(\beta_0,\psi_0\right)-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}E\left[g\left(x_i,z_i;\hat{\beta},\hat{\psi}(v_i),\gamma_0\right)\right],$$

Since $\nu_N(\beta,\psi)$ is stochastically equicontinuous (see Lemma 2), $\hat{\beta}$ and $\hat{\psi}$ are consistent estimators of β_0 and ψ_0 (see Haerdle, Huet, Mammen and Sperlich, 1998) with respect to the pseudo-metric $\rho\left((\beta,\psi),(\beta_0,\psi_0)\right)$ and $P\left(\left(\hat{\beta},\hat{\psi}\right)\in I\!B\times\Psi\right)\longrightarrow 1$ then it is possible to show that

$$(47)\qquad \nu_N\left(\hat{\beta},\hat{\psi}\right)-\nu_N\left(\beta_0,\psi_0\right)\longrightarrow_p 0,$$

see Andrews (1994), p. 49. Moreover

$$(48)\qquad \frac{1}{\sqrt{N}}\sum_{i=1}^{N}E\left[g\left(x_i,z_i;\hat{\beta},\hat{\psi}(v_i),\gamma_0\right)\right]\longrightarrow_p 0$$

To show this, rewrite the left hand side of equation (48) as

$$\begin{aligned}(49)\qquad &\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}E\left[g\left(x_i,z_i;\hat{\beta},\hat{\psi}(v_i),\gamma_0\right)\right]\times I\left(\rho\left(\left(\hat{\beta},\hat{\psi}\right),(\beta_0,\psi_0)\right)<\epsilon\right) \\ &+\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}E\left[g\left(x_i,z_i;\hat{\beta},\hat{\psi}(v_i),\gamma_0\right)\right]\times I\left(\rho\left(\left(\hat{\beta},\hat{\psi}\right),(\beta_0,\psi_0)\right)\geq\epsilon\right),\end{aligned}$$

then (48) holds because of the consistency result and the fact that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}E\left[g\left(x_i,z_i;\beta,\psi(v_i),\gamma_0\right)\right]=0\qquad\text{for all }\beta\text{ and }\psi$$

in a neighborhood of β_0 and ψ_0. Under assumption (A.2), we can apply the Lindeberg-Levy CLT and then equation (43) is proved.

We finally we show (42). In order to do so, note first that

$$(50)\qquad \frac{1}{N}\sum_{i=1}^{N}\nabla_\gamma g\left(x_i,z_i,\hat{\beta},\hat{\psi}(v_i),\bar{\gamma}\right)=\frac{2}{N}\sum_{i=1}^{N}\zeta_i\tilde{z}_i\tilde{z}_i^T$$

where $\tilde{z}_i^T = \left[z_i^T \quad \lambda\left(s_i^T\hat{\beta} + \hat{\psi}(v_i)\right)\right]$. Then it is sufficient to show that

$$\sup_{(\beta,\psi)\in B\times\Psi}\left|\frac{1}{N}\sum_{i=1}^{N}\nabla_\gamma g\left(x_i, z_i, \beta, \psi(v_i)\right) - E\left[\nabla_\gamma g\left(x, z, \beta_0, \psi_0(v)\right)\right]\right| = o_p(1) \tag{51}$$

Given that both $\hat{\beta}$ and $\hat{\psi}$ are consistent estimates of β_0, ψ_0 , assumptions *(A.1)* and *(A.2)* and Lemma 4.3 from Newey and McFadden (1994), p. 2156, we have

$$\frac{1}{N}\sum_{i=1}^{N}\nabla_\gamma g\left(x_i, z_i, \hat{\beta}, \hat{\psi}(v_i)\right) = E\left[\nabla_\gamma g\left(x, z, \beta_0, \psi_0(v)\right)\right] + o_p(1) \tag{52}$$

Then, using Lemma 2, Theorem 1 from Andrews (1992), p. 244 applies and the proof is done.

□

Appendix II: Newton-Raphson algorithm for local likelihood

We start with calculating $\partial\mathcal{L}^S(\eta_j,\gamma,\sigma)/\partial\eta_j$ and $\partial^2\mathcal{L}^S(\eta_j,\gamma,\sigma)/\partial\eta_j^2$, where η_j is the function $\eta(\cdot)$ at point t_j. For ease of notation we set $u_{ij} = \{-\gamma^T s_i - \eta_j\}\sigma^{-1}$ and get

$$\frac{\partial\mathcal{L}^S(\eta_j,\gamma,\sigma)}{\partial\eta_j} = \frac{1}{\sigma}\sum_{i=1}^{n}A_n\left(u_{ij}\right)K_h(t_j - t_i), \tag{53}$$

$$\frac{\partial^2\mathcal{L}^S(\eta_j,\gamma,\sigma)}{\partial\eta_j^2} = \frac{-1}{\sigma^2}\sum_{i=1}^{n}\left[1 + \frac{u_{ij}f(u_{ij})}{1-F(u_{ij})} + \frac{f^2(u_{ij})}{\{1-F(u_{ij})\}^2}\right]K_h(t_j - t_i). \tag{54}$$

where

$$A_n\left(u_{ij}\right) = \frac{Y_i}{\sigma} + u_{ij} - \frac{f(u_{ij})}{1-F(u_{ij})}$$

Moreover, we have used $\partial f(-u/s)/\partial u = -us^{-2}f(-u/s)$.

Next, we calculate $\partial\mathcal{L}(\eta_{\gamma,\sigma},\gamma,\sigma)/\partial\gamma$ and $\partial^2\mathcal{L}(\eta_{\gamma,\sigma},\gamma,\sigma)/\partial\gamma^2$ and denote $\eta_i = \eta_{\gamma,\sigma}(t_i)$, $\tilde{s}_{ii} = s_i + \partial\eta_i/\partial\gamma$:

$$\frac{\partial\mathcal{L}(\eta_{\gamma,\sigma},\gamma,\sigma)}{\partial\gamma} = \frac{1}{\sigma}\sum_{i=1}^{n}\left[\frac{Y_i}{\sigma} + u_{ii} - \frac{f(u_{ii})}{1-F(u_{ii})}\right]\tilde{s}_{ii}. \tag{55}$$

For the Hesse-matrix we neglect the dependency of $\tilde{s}_{ii}$ on γ and get

$$\frac{\partial\mathcal{L}^2(\eta_{\gamma,\sigma},\gamma,\sigma)}{\partial\gamma^2} = \frac{-1}{\sigma^2}\sum_{i=1}^{n}\left[1 + \frac{f(u_{ii})u_{ii}}{1-F(u_{ii})} - \frac{f^2(u_{ii})}{\{1-F(u_{ii})\}^2}\right]\tilde{s}_{ii}\tilde{s}_{ii}^T.$$

Here we used $\partial f(u_{ii})/\partial\gamma = f(u_{ii})u_{ii}\tilde{s}_{ii}$.

A little bit more complicated is to get $\partial\mathcal{L}(\eta_{\gamma,\sigma},\gamma,\sigma)/\partial\sigma$ and $\partial^2\mathcal{L}(\eta_{\gamma,\sigma},\gamma,\sigma)/\partial\sigma^2$; set $\eta'_\sigma = \partial\eta_{\gamma,\sigma}/\partial\sigma$:

$$\frac{\partial\mathcal{L}(\eta_{\gamma,\sigma},\gamma,\sigma)}{\partial\sigma} = \frac{-1}{\sigma}\sum_{i=1}^n B_n\left(u_{ii}\right). \tag{56}$$

Where

$$B_n(u_{ii}) = 1 - \left(\frac{Y_i}{\sigma} + u_{ii}\right)^2 - \left\{\frac{Y_i}{\sigma} + u_{ii}\right\}\eta'_\sigma + \frac{f(u_{ii})(u_{ii} + \eta'_\sigma)}{1 - F(u_{ii})}$$

For the Hesse-matrix we again neglect the dependency of η'_σ on σ and so get with $B_n(u_i i)$ from (56)

$$\frac{\partial^2\mathcal{L}(\eta_{\gamma,\sigma},\gamma,\sigma)}{\partial\sigma^2} = \sum_{i=1}^n \sigma^{-2}B_n\left(u_{ii}\right) + \sigma^{-1}\Bigg[2\sigma^{-1}\{\frac{Y_i}{\sigma} + u_{ii}\}\{-\eta' - (\frac{Y_i}{\sigma} + u_{ii})\} +$$
$$\frac{\eta'_\sigma}{\sigma}\Big\{-\eta'_\sigma - (\frac{Y_i}{\sigma} + u_{ii})\Big\} - \frac{f'(u_{ii})(u_{ii} + \eta'_\sigma)}{1 - F(u_{ii})} - \frac{f(u_{ii})u'_{ii}}{1 - F(u_{ii})} - \frac{f^2(u_{ii})u'_{ii}(u_{ii} + \eta'_\sigma)}{\{1 - F(u_{ii})\}^2}\Bigg],$$

with

$$u'_{ii} = \frac{\partial u_{ii}}{\partial\sigma} = -\sigma^{-1}\left(u_{ii} + \eta'_\sigma\right)$$

and

$$f'(u_{ii}) = \frac{\partial f(u_{ii})}{\partial\sigma} = \frac{f(u_{ii})}{\sigma}\big(-1 + \eta'_\sigma u_{ii} + u_{ii}^2\big).$$

The question is how to get $\eta'_\gamma = \partial\eta/\partial\gamma$ and $\eta'_\sigma = \partial\eta/\partial\sigma$. For the likelihood maximizing η_j expression (53) is equal to zero. First we derive it with respect to γ:

$$\begin{aligned}
\frac{\partial^2\mathcal{L}^S(\eta_j,\gamma,\sigma)}{\partial\eta_j\partial\gamma} &= -\sigma^{-2}\sum_{i=1}^n\left[1 + \frac{f(u_{ij})u_{ij}}{1 - F(u_{ij})} + \Big\{\frac{f(u_{ij})}{1 - F(u_{ij})}\Big\}^2\right]K_h(t_j - t_i)\tilde{s}_{ii} = 0 \\
\Longleftrightarrow \eta'_\gamma &= \frac{-\sum_{i=1}^n\left[1 + \frac{f(u_{ij})u_{ij}}{1-F(u_{ij})} + \Big\{\frac{f(u_{ij})}{1-F(u_{ij})}\Big\}^2\right]K_h(t_j - t_i)s_i}{\sum_{i=1}^n\left[1 + \frac{f(u_{ij})u_{ij}}{1-F(u_{ij})} + \Big\{\frac{f(u_{ij})}{1-F(u_{ij})}\Big\}^2\right]K_h(t_j - t_i)}.
\end{aligned}$$

Second we derive this expression with respect to σ:

$$\begin{aligned}
\frac{\partial^2\mathcal{L}^S(\eta_j,\gamma,\sigma)}{\partial\eta_j\partial\sigma} &= -\sigma^{-2}\sum_{i=1}^n\Bigg[A_n\left(u_{ij}\right) + \eta'_\sigma + \frac{Y_i}{\sigma} + u_{ij} - \frac{f(u_{ij})}{1 - F(u_{ij})} + \eta'_\sigma\frac{f(u_{ij})u_{ij}}{1 - F(u_{ij})} \\
&\quad + \frac{f(u_{ij})u_{ij}^2}{1 - F(u_{ij})} - \frac{f^2(u_{ij})u_{ij}}{\{1 - F(u_{ij})\}^2} - \eta'_\sigma\frac{f^2(u_{ij})}{\{1 - F(u_{ij})\}^2}\Bigg]K_h(t_j - t_i), \\
\Longleftrightarrow \eta'_\sigma &= \frac{-\sum_{i=1}^n\left[A_n\left(u_{ij}\right) + \frac{Y_i}{\sigma} + u_{ij} - \frac{f(u_{ij})}{1-F(u_{ij})} + \frac{f(u_{ij})u_{ij}^2}{1-F(u_{ij})} - \frac{f^2(u_{ij})u_{ij}}{\{1-F(u_{ij})\}^2}\right]K_h(t_j - t_i)}{\sum_{i=1}^n\left[1 + \frac{f(u_{ij})u_{ij}}{1-F(u_{ij})} - \frac{f^2(u_{ij})}{\{1-F(u_{ij})\}^2}\right]K_h(t_j - t_i)}.
\end{aligned}$$

Finally we need the mixed derivatives $\partial\mathcal{L}(\eta_{\gamma,\sigma},\gamma,\sigma)/\partial\gamma\partial\sigma$ and $\partial\mathcal{L}(\eta_{\gamma,\sigma},\gamma,\sigma)/\partial\sigma\partial\gamma$.

$$\frac{\partial\mathcal{L}(\eta_{\gamma,\sigma},\gamma,\sigma)}{\partial\gamma\partial\sigma} =$$
$$-\sigma^{-2}\sum_{i=1}^n\left[\eta'_\sigma + 2u_{ij} + \frac{\sigma f'_\sigma(u_{ij})}{1 - F(u_{ij})} - \frac{f^2(u_{ij})(\eta'_\sigma + u_{ij})}{\{1 - F(u_{ij})\}^2} - \frac{f(u_{ij})}{1 - F(u_{ij})}\right]\tilde{s}_{ii}$$

with $\eta'_\sigma = \partial \eta_{\gamma,\sigma}/\partial\sigma$, $\tilde{s}_{ii}$ as above and

$$f'_\sigma(u_{ij}) = \frac{\partial f(u_{ij})}{\partial \sigma} = \frac{f(u_{ij})}{\sigma}\Big(-1 + \eta'_\sigma u_{ij} + u_{ij}^2\Big),$$

$$\begin{aligned}\frac{\partial \mathcal{L}(\eta_{\gamma,\sigma},\gamma,\sigma)}{\partial\sigma\partial\gamma} &= -\sigma^{-2}\sum_{i=1}^n \Big[2u_{ij} + \eta'_\sigma + \\ &\quad \frac{u_{ij} f(u_{ij})(u_{ij}+\eta'_\sigma)}{1-F(u_{ij})} - \frac{f(u_{ij})}{1-F(u_{ij})} - \Big\{\frac{f(u_{ij})}{1-F(u_{ij})}\Big\}^2 (u_{ij}+\eta'_\sigma)\Big]\tilde{s}_{ii}.\end{aligned}$$

Here, we have neglected the dependency of η'_σ on γ and the dependency of η'_γ on σ.

The Hesse-matrix for $\mathcal{L}^S$ is simply given by (54).

The Hesse-matrix for $\mathcal{L}$ is given by

$$H_{\mathcal{L}} = \begin{pmatrix} \frac{\partial^2\mathcal{L}}{\partial\gamma^2} & \frac{\partial^2\mathcal{L}}{\partial\gamma\partial\sigma} \\ \frac{\partial^2\mathcal{L}}{\partial\sigma\partial\gamma} & \frac{\partial^2\mathcal{L}}{\partial\sigma^2} \end{pmatrix}.$$

References

Ahn, H. and Powell, J.L. (1993) Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics,* **58**: 3-29.

Andrews, D.W.K. (1994) Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity. *Econometrica,* **62**: 43-72.

Boskin, M.J. (1973) The economics of Labor Supply. in G. Cain and H. Watts (eds), *Income, Maintenance and Labor Supply.* New-York.

Blundell, R. and Meghir, C. (1986) Selection Criteria for a Microeconometric Model of Labour Supply. *Journal of Applied Econometrics,* **1**: 55-80.

Hall, R.E. (1973) Wages, Income, and Hours of Work in the U.S. Labour Force. in G. Cain and H. Watts (eds), *Income, Maintenance and Labor Supply.* New-York.

Hastie, T. J. and R. J. Tibshirani. (1990) Generalized Additive Models. *Chapman and Hall: London.*

Härdle, H., Huet, S., Mammen, E. and Sperlich, S. (1998) Semiparametric additive indices for binary response and generalized additive models. *Reprint, Humboldt-Universität zu Berlin, Germany.*

Heckman, J.J. (1979) Sample Selection Bias as a Specification Error. *Econometrica,* **47**: 153-161.

Heckman, J.J. (1993) What has been learned about labor supply in the past twenty years? *American Economic Review, Papers and Proceedings,* **83**: 116-121.

Linton, O. and Nielsen, J.P. (1995) A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, **82**: 93-101.

Linton, O. Mammen, E. and Nielsen, J.P. (1997) The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Preprint, Yale University.*

Newey, W.K. and McFadden, D. (1994) Large sample estimation and hypothesis testing. in R.F. Engle and D. McFadden (eds.), *Handbook of Econometrics,* **4:**2113-2241.

Mroz, T.A. (1987) The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions. *Econometrica,* **:** 765-799.

Rosen (1976) Taxes in a Labor Supply Model with Joint Wage-Hours determination. *Econometrica*, **44**: 485-507.

Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*, New York: Wiley.

Severini, T.A. and Staniswalis, J.G. (1994) Quasi-Likelihood estimation in semiparametric models. *J. Amer. Statist. Assoc.*, **89** : 501-511.

Sperlich, S., Linton, O.B. and Härdle, W. (1997) A simulation comparison between integration and backfitting methods of estimating separable nonparametric regression models. *Discussion Paper 66, sfb373, Humboldt-Universität Berlin, Germany* .

Staniswalis, J.G.. (1989) The kernel estimate of a regression function in likelihood-based models. *Journal of the American Statistical Association*, **84**: 276-283.

Stone, C.J. (1986) The dimensionality reduction principle for generalized additive models. *Ann. Statist.*, **14**: 590-606.

Tjøstheim, D. and Auestad, B.H. (1994) Nonparametric identification of nonlinear time series: projections. *J. American Statistical Association*, **89**: 1398-1409.

Vijverberg, W.P. (1991) Selectivity and Distributional Assumptions in Static Labor Supply Models. *Southern Economic Journal,* **57**: 822-840.

Wales, T.J. and Woodland, A.D. (1980) Sample Selectivity and the Estimation of labor Supply Functions. *International Economic Review,* **21**: 437-468.