Semiparametric Instrumental Variable Estimation in an Endogenous Treatment Model[1]

Roger Klein and Chan Shen

December, 2013

[1]This version represents work in progress, and will be replaced by a complete draft when available.

Abstract

In this paper we propose an instrumental variables (IV) estimator for a semiparametric outcome model with endogeous discrete treatment variables. The main contribution of our paper is that the identification, consistency and asymptotic normality of our estimator all hold even under misspecification of the treatment model. As expected from Newey and McFadden (1994), the covariance matrix for the parameters and functions of interest does not depend on estimation uncertainty of the instruments for the endogenous treatments. Further, we extend our method to the nonparametrtic case with both continuous and discrete exogenous variables. We prove identification, consistency and asymptotic normality and provide uniform convergence results for estimated functions of interest.

1 Introduction

The purpose of this paper is to develop a robust IV estimator for a class of semiparametric treatment models where the treatments are discrete endogenous variables that flexibly interact with other explanatory variables. We use the term robust to indicate that the method formulated here is robust to imposing incorrect restrictions on the treatment model.

Endogeneity is a very important issue in econometric analysis, and there is a large literature on this issue (see e.g. Heckman (1978) and Hausman (1983)). To deal with endogeneity, the IV estimator is widely employed in the empirical literature (e.g. Card (2001)). In part, the appeal of this estimator in linear systems is its very favorable property of robustness against misspecification in modelling the instruments. In nonlinear parametric models, Amemiya (1974,1977) developed an optimal instrument which depends on an unknown conditional expectation of the derivative of a residual function. Employing nonparametric expectations while retaining the parametric outcome model structure, Newey (1990) provided a way to implement the optimal instrument and showed that the estimator is consistent, asymptotically normal and efficient. The literature continued to progress to fully nonparametric models based on either generalizations of two-stage-least-squares or control approaches. For example, Newey and Powell (2003) developed a two-stage series estimator with a detailed discussion on identification. Das (2005) and Cai et. al. (2005) developed a two-stage estimator for a model that is linear in endogenous variables with an associated multiplicative nonparametric impact or marginal effect response functions. Newey, Powell, and Vella (1999) and Imbens and Newey (2009) developed control estimators for a nonparametric triangular system where both the treatment and outcome variables of interest are continuous. It is important to note that in the case of nonparametric models, the issue of imposing incorrect parametric restrictions does not arise.

There is also a growing literature on estimating semiparametric models with endogeneity (see Blundell and Powell (2001) for a survey). For example, Ai and Chen (2003, 2007) proposed a sieve minimum distance estimator for a wide class of semiparametric models that accommodates endogeneity. Meanwhile, there is a vast literature on semiparametric models for specific model structures. To our knowledge, these methods are based on either a control method, an extremum estimator approach, or a generalized two-stage-least-squares procedure. For example, Blundell and Powell (2004) developed a control estimator for estimating a binary response model when one or more of the explanatory variables are endogenous; Rothe (2009) formulated a maximum-likelihood type estimator for this model using a double index formulation to deal with endogeneity; while Lewbel (2000) developed a two-stage type estimator. However, there is an important class of semiparametric endogenous treatment models where misspecification of the treatment model is an important problem that needs to be addressed. This class of semiparametric endogenous treatment model has a discrete endogenous treatment and a continuous outcome. The purpose of this paper is to fill this gap and provide a robust IV estimator for this model.

2 The Model and the Estimator

The model we consider is one with discrete endogenous treatments and a continuous outcome. There are many empirical applications with such a model structure. For example, in health economics, the health outcome (e.g. survival) can be continuous and the treatment could be no treatment, chemotherapy only, radiation therapy only, or both chemotherapy and radiation therapy. The treatment could also be smoking choices: no smoking, smoking 1-5 cigarettes per week, smoking 6-10 cigarettes, etc. There are also many examples in labor economics such as job training program choice with wages being the outcome.

Denote Y_i as the continuous outcome and $\mathcal{W}_i$ as a function of the i.i.d. exogenous variables $X_i \equiv [X_{1i}...X_{di} \;\; X_{Di}]$ where $X_{1i}...X_{di}$ are continuous variables, while X_{Di} is a vector of discrete variables. In the semiparametric case, $\mathcal{W}_i$ may be the single index $V_i(\theta_0) = X_{1i} + X_{IIi}\theta_0$, where $X_{IIi} \equiv [X_{2i}...X_{di} \;\; X_{Di}]$ or a vector of indices. In the nonparametric case, $\mathcal{W}_i$ is simply X_i. With $L+1$ possible treatment options, the endogenous treatment variable $\mathfrak{T}_i$ could take on values $0, ..., L$. Writing the treatment indicator as $T_{il} = 1\{\mathfrak{T}_i = l\}$, $l = 1, ..., L$, and letting $\mathfrak{T}_i = 0$ be the reference (e.g. no treatment) option, the model we consider is:

$$Y_i = g(\mathcal{W}_i, \mathfrak{T}_i) + \varepsilon_i,$$

where ε_i is the error component with conditional expectation zero given the exogenous variables.

Because of the discreteness of the treatment, we can rewrite the model in the following form:

$$
\begin{aligned}
Y_i &= \sum_{l=1}^{L} g(\mathcal{W}_i, l)T_{il} + g(\mathcal{W}_i, 0)\left[1 - \sum_{l=1}^{L} T_{il}\right] + \varepsilon_i \qquad (1)\\
&= \sum_{l=1}^{L}[g(\mathcal{W}_i, l) - g(\mathcal{W}_i, 0)]T_{il} + g(\mathcal{W}_i, 0) + \varepsilon_i \\
&\equiv \sum_{l=1}^{L} M_l(\mathcal{W}_i)T_{il} + B(\mathcal{W}_i) + \varepsilon_i.
\end{aligned}
$$

This model structure is the same as in Das (2005) and Cai et. al. (2005) in the nonparametric case. For this case, they propose a two-stage estimator where the endogenous treatment variables are replaced by estimates of their conditional expectations. In contrast, here we develop an IV estimator that retains consistency in the semiparametric case when the treatment model is misspecified. While these approaches can coincide in linear models, they differ in the present nonlinear context. While we cover the nonparametric case, our main contribution is in the semiparametric case where we establish that identification, consistency and asymptotic normality of our estimator all hold even under misspecification of the treatment model. This result mirrors that in linear models where this robustness property is well-known. Such misspecification is ruled out in the nonparametric framework where there are no incorrect parametric restrictions. For the nonparametric estimator presented here, we cover the case where the explanatory exogenous variables are continuous, discrete, or combinations of both.

2.1 Semiparametric Model and Estimator

For the semiparametric model, we let $W_i \equiv V_j(\theta_0)$ be an index and begin with a localized instrumental variables estimation of $M_l(V_j(\theta))$ and $B\left(V_j(\theta)\right)$.[1] We use $\hat{P}_l$, a semiparametric estimator for the l^{th} treatment probability, as the corresponding treatment instrument and note that $\hat{P}_l$ may be misspecified. For example, suppose that we incorrectly impose a single index restriction on the treatment probability model when there is a second index driving heteroskedasticity. Then we would be estimating the incorrect probability $\tilde{P}$,

[1] It is relatively straightforward to let the outcome equation depend on a vector of indices. Here, we present the single index case.

in that $\hat{P}$ would not converge to the true P. Using localized linear projections, we provide identification results under this type of misspecification.

With $\hat{P}_{il}, l = 1, ..., L,$ as the i^{th} observation on the instrument for the l^{th} treatment, let:

$$\hat{Q}_i = t_i \left\{ \begin{array}{ccccc} \hat{P}_{i1} & \hat{P}_{i2} & ... & \hat{P}_{iL} & 1 \end{array} \right\}_{1x(L+1)}, \tag{2}$$

where t_i is a trimming function to control g_V (the density for V) and the density denominators in $\hat{P}_{il}$. Then, the estimators $\hat{M}_l(V_j(\theta))$ and $\hat{B}\left(V_j(\theta)\right)$ satisfy

$$\sum_{i=1}^{N} \hat{Q}_i' \left[Y_i - \hat{M}_1\left(V_j\left(\theta\right)\right) T_{i1} - ...\hat{M}_l\left(V_j\left(\theta\right)\right) T_{iL} - \hat{B}\left(V_j\left(\theta\right)\right) \right] K(V_j, V_i) = 0$$

where $l = 1, ..., L$ and $K(v, V_i) = \frac{1}{h} k(\frac{V_i - v}{h})$ is a kernel weight, with $k(\cdot)$ a symmetric density and $\displaystyle\int t^2 k(t) dt$ is bounded.[2]. Notice that this IV estimator estimates the M and B functions at each index value V_j by essentially using observations in a small neighborhood of V_j. The kernel weights ensure this localization. By repeating the above estimation strategy for given θ, we obtain $\hat{M}_l\left(V_j\left(\theta\right)\right)$ and $\hat{B}\left(V_j\left(\theta\right)\right)$ for $j = 1, ..., N$ and $l = 1, ..., L$. The key to this localized argument is that in estimating the unknown functions at a point, they are treated as if they were parameters (see Fan and Gijbels (1996) for related discussions).

To provide the estimator satisfying the above moment conditions, under index assumptions, write the model as

$$Y = R\alpha_0(v) + \varepsilon + \Delta(v),$$

[2] In semiparametric models with a linear index, for identification one of the index variables must be continuous. In this case, the index will be continuous even if some of the index variables are discrete. Therefore continuous kernels of the form assumed here are appropriate. To deal with discrete regressors in the nonparametric case, we formulate a different kernel that is in part based on indicators to handle discrete explanatory variables.

where:

$$
\begin{aligned}
R &= \{T \quad \mathbf{1}\}_{Nx(L+1)}, \, R_i = \left\{ \begin{array}{ccccc} T_{i1} & T_{i2} & \ldots & T_{iL} & \mathbf{1} \end{array} \right\}_{1x(L+1)} \\
\alpha_0(v) &\equiv \left\{ \begin{array}{l} M(v)_{Lx1} \\ B(v) \end{array} \right\}_{(L+1)x1} \\
\Delta(v) &\equiv \left\{ \begin{array}{l} \Delta_1(v) \\ \ldots \\ \Delta_N(v) \end{array} \right\}_{Nx1}, \Delta_i(v) \equiv R_i \left[\alpha_0(V_i) - \alpha_0(v)\right]
\end{aligned}
$$

With X_{1i} being continuous, let $V_i = X_{1i} + X_{IIi}\theta_0$, $X_{IIi} \equiv [X_{2i} \ldots X_{di} \quad X_{Di}]$,and $D_N(v,\theta)$ be an NxN diagonal matrix having i^{th} diagonal element $K(v, V_i(\theta))$. Throughout, v is in a compact subset of the support for V_i where $g_V(v) \neq 0$. Define $\hat{Q}$ as the matrix with i^{th} component $\hat{Q}_i$ as in (2). Then, in the semiparametric case, our estimator for the α function is given as:

$$
\hat{\alpha}(v,\theta) \equiv \hat{\Omega}(v,\theta)^{-1} \frac{\hat{Q}' D_N(v,\theta) Y}{N}
$$

where $\hat{\Omega}(v,\theta)_{(L+1)x(L+1)} \equiv \frac{\hat{Q}' D_N(v,\theta) R}{N}$.

Before proceeding to discuss estimation of the index parameters, it is useful to further discuss $\hat{\alpha}(v,\theta)$ as it and its derivative play a key role in the large sample theory. It can be shown that:

$$
\hat{\alpha}(v,\theta) \xrightarrow{p} \left[E\left(Q_i' R_i | V_i(\theta) = v\right)\right]^{-1} E\left(Q_i' Y_i | V_i(\theta) = v\right) \equiv \alpha(v;\theta) \qquad (3)
$$

Here, $\alpha(v;\theta)$ depends not only on v the point at which it is evaluated, but also on θ because the conditioning random variable $V_i(\theta)$ depends on it. At $\theta = \theta_0$, $\alpha_0(v) \equiv \alpha(v;\theta_0)$ is a vector of the true functions evaluated at the point v.

Next we estimate the semiparametric parameter θ by another IV step with

$$
\hat{\theta} = \arg\min_{\theta} \left[\sum_{i=1}^{N} \hat{Z}_i' \left[Y_i - R_i \hat{\alpha}(V_i(\theta), \theta)\right] \right]^2 \qquad (4)
$$

where

$$
\hat{Z}_i = \hat{P}_{1i} W_{1i} t_i + \ldots + \hat{P}_{Li} W_{Li} t_i + W_{L+1i} t_i. \qquad (5)
$$

The $W_i's$ are weights that can depend on V_i and X_i. We will discuss the optimal weights in the Large Sample Results Section. The structure of this instrument is suggested by Newey's (1990) optimal instrument in parametric models.

2.2 Nonparametric Model and Estimator

Return to the general model in (1), let $\mathcal{W}_i = X_i$ so that all functions depend on the vector of exogenous variables for observation i. In this case, employing notation similar to that above, our objective is to estimate $\alpha(x)$ at many values of x in a compact set As above, write the model in localized form as:

$$\begin{aligned} Y &= R\alpha(x) + \varepsilon + \Delta(x), \\ \text{where } \alpha(x) &\equiv \left\{ \begin{array}{l} M(x)_{Lx1} \\ B(x) \end{array} \right\}_{(L+1)x1} \\ \Delta(x) &\equiv \left\{ \begin{array}{l} \Delta_1(x) \\ ... \\ \Delta_N(x) \end{array} \right\}_{Nx1}, \Delta_i(x) \equiv R_i \left[\alpha(X_i) - \alpha(x)\right] \end{aligned}$$

Let $X_i \equiv [X_{1i}...X_{di} \;\; X_{Di}]$ be i.i.d., where $X_{C_i} \equiv [X_{1i}...X_{di}]$ is a vector of continuous variables, while X_{Di} is a vector of discrete variables. Denote $x = [x_1...x_d \;\; x_D]$ and define:

$$K(x, X_i) = 1\left\{X_{Di} = x_D\right\} \prod_{j=1}^{d} \frac{1}{h} k(\frac{X_{ji} - x_j}{h}),$$

where $h = O\left(N^{-\frac{1}{4+d}}\right)$, $k(\cdot)$ is a symmetric bounded density, and $\int t^2 k(t)dt$ is bounded. Let g_X be the density for X_i and x a point in a compact subset of the support for X_i with $g_X(x) \neq 0$.

Define $\hat{Q}_i$ as in (2) with the nonparametric treatment probability, $\hat{P}_l$ as the treatment instrument for treatment T. Then,the IV estimator at the point x satisfies:

$$\sum_{i=1}^{N} \hat{Q}_i' \left[Y_i - \hat{M}_1(x) T_{i1} - ...\hat{M}_l(x) T_{iL} - \hat{B}(x)\right] K(x, X_i) = 0$$

Then, the estimator satisfying the above moment conditions is given as:

$$\begin{aligned} \hat{\alpha}(x) &\equiv \left\{ \begin{array}{l} \hat{M}(x)_{Lx1} \\ \hat{B}(x) \end{array} \right\}_{(L+1)x1} \equiv \hat{\Omega}(x)^{-1} \frac{\hat{Q}' D_N(x) Y}{N} \\ \text{where } \hat{\Omega}(x)_{(L+1)x(L+1)} &\equiv \frac{\hat{Q}' D_N(x) R}{N}. \end{aligned}$$

3 Assumptions and Definitions

In Progress

4 Large Sample Results

With all results being proved in the Appendix, the first theorm below provide identification and consistency results for the nonparametric case.

Theorem 1. Nonparametric Identification and Consistency. Assume that the covariance matrix of $P_{i*}|X_i = x$ is positive definite, which requires that P_{il} depend on one or more variables excluded from X. Define constants c_p^* and c_u^* such that $0 < c_p^* < \frac{2}{4+d}$ and $0 < c_u^* < \frac{1}{2+d}$, where d is the dimension of the continuous component of X. With x in a compact set. Then with $\Omega(x) \equiv E[Q_i^{'}R_i|X_i = x]g(x)$,

$$\begin{aligned}
a) &: \sup_x \left\|\hat{\Omega}(x) - \Omega(x)\right\| = o_p(1) \\
b) &: \Omega(x) \text{ is non-singular} \\
c) &: |\hat{\alpha}(x) - \alpha(x)| = o_p\left(N^{-c_p^*}\right) \\
d) &: \sup_x |\hat{\alpha}(x) - \alpha(x)| = o_p\left(N^{-c_u^*}\right)
\end{aligned}$$

where $\|\cdot\|$ means absolute value for each element.

With our focus being on the semiparametric model, the theorems below provide identification, consistency, normality and efficiency results.

Theorem 2 Semiparametric Identification and Consistency. Assume that the covariance matrix of $P_{i*}|V_i(\theta) = v$ is positive definite. Defining $\Omega(v,\theta) \equiv E[Q_i^{'}R_i|V_i(\theta) = v]g(v)$ and $\alpha(v,\theta) \equiv \Omega(v,\theta)^{-1}E\left[Q_i'Y_i|V_i(\theta) = v\right]$, it follows that:

$$\begin{aligned}
a) &: \sup_{v,\theta} \left\|\hat{\Omega}(v,\theta) - \Omega(v,\theta)\right\| \xrightarrow{p} 0 \\
b) &: \Omega(v,\theta) \text{ is non-singular} \\
c) &: \sup_{v,\theta} |\hat{\alpha}(v,\theta) - \alpha(v,\theta)| \xrightarrow{p} 0 \\
d) &: \hat{\theta} \xrightarrow{p} \theta_0,
\end{aligned}$$

where $\|\cdot\|$ means absolute value for each element and where (b) holds for correctly specified models.[3]

Theorem 3. Normality for $\hat{\theta}$. Define

$$\begin{aligned} H_0 &\equiv -E\left(Z_j' R_j \nabla_\theta \alpha\left(V_j\left(\theta\right), \theta\right)_{\theta=\theta_0}\right).^4 \\ R_z'(V_j\left(\theta_0\right)) &\equiv \left[Z_j' - \left[E\left(Z_j R_j | V_j\right)\right]\left[E[Q_j' R_j | V_j]\right]^{-1} Q_j'\right]. \end{aligned}$$

Then

$$\sqrt{N}\left[\hat{\theta} - \theta_0\right] \xrightarrow{d} \mathrm{W}\tilde{\ } N\left(\ 0, \sigma_\varepsilon^2 H_0^{-1} E\left[R_z(V_j)' R_z(V_j)\right] H_0^{-1}\right),$$

Recalling the definition of the $\alpha - function$ in (3), the theorem below requires $\nabla\alpha(V_j(\theta);\theta)_{\theta=\theta_0}$, the derivative of the $\alpha - function$ with respect to θ and evaluated at $\theta_{0.}$ We note that $\alpha_0(V_j(\theta)) \equiv \alpha(V_j(\theta);\theta_0)$ is the true $\alpha - function$ evaluated at the point $V_j(\theta)$. In the Appendix, we show that:

$$\nabla_\theta \alpha(V_j(\theta);\theta)_{\theta=\theta_0} = \nabla_\theta \alpha_0(V_j(\theta))_{\theta=\theta_0} - E\left[Q_j' Q_j | V_j\right]^{-1} E\left(Q_j' Q_j \nabla_\theta \alpha_0(V_j(\theta))_{\theta=\theta_0} | V_j\right),$$

a result that is useful for efficiency in the next theorem.

Theorem 4. Efficiency. Defining $\varphi_j = Q_j \nabla_\theta \alpha_0(V_j(\theta))_{\theta=\theta_0}$, the j^{th} observation on the optimal instrument is given by

$$R_\varphi(V_j) = Q_j \nabla_\theta \alpha(V_j(\theta);\theta)_{\theta=\theta_0} = \varphi - Q_j E\left[Q_j' Q_j | V_j\right]^{-1} E\left(Q_j' \varphi | V_j\right),$$

where $\nabla_\theta \alpha(V_j\left(\theta\right);\theta)_{\theta=\theta_0}$ denotes the the derivative of $\alpha\left(V_j\left(\theta\right), \theta\right)$ taken with respect to θ and then evaluated at θ_0.

Therefore, the optimal instrument is (5) with weight $W_i = \nabla_\theta \alpha(V_j(\theta);\theta)_{\theta=\theta_0}$. The form of this weight is suggested by the optimal instrument for the parametric case with known α functions as derived in Newey (1990).

To give some sense as to how the estimator performs, in the next section we report several Monte-Carlo simulations. A fuller set of Monte-Carlo experiments will be included in the completed paper.

[3] We do have conditions that ensure identification in the case that the treatment model is not correctly specified, but are trying to develop more primitive conditions for this case.

5 Monte-Carlo Results

In the Monte-Carlo study, we investigated two different designs. With all exogenous variables $X_{1,}X_2$ and Z generated to be standard normal, and the error terms u and ε being correlated, the treatment and outcome models are given as:

$$\begin{aligned} Y_2 &= 1\{V_2 \geq u\} \text{ where } V_2 = X_2 + X_3; \\ Y_1 &= V_1 + V_1 * Y_2 + Y_2 + \varepsilon \text{ where } V_1 = X_1 + X_2. \end{aligned}$$

In the first design the error terms u and ε are homoskedastic and follow normal distributions. In the second design, the treatment error is heteroscedastic with $u = cX_2^2 u^*$, where c is a constant and u^* is distributed as standard normal, independent of X. The sample size in the Monte-Carlo is 2000, and we ran 100 replications. The instrument that we use for estimating the marginal effects function is given by the probability of treatment conditioned on a single index. This instrument is "correct" for the homoscedastic model, but "incorrect" in the case of heteroskedasticity.

We also require instruments for estimating the index parameters. The parameter estimator $\hat{\theta}$ solves:

$$\frac{1}{N}\sum \hat{Z}_j' \left[Y_j - [T_j\ 1]\,\hat{\alpha}\left(V_j(\theta), \theta\right)\right] = 0,$$

If the α-functions were known, then with α_0 as the vector of true functions and $X_j \equiv [X_{1j}, X_{2j}, X_{3j}]$, from Newey (1990) the optimal instrument would be:

$$Z_j' = E\left([T_j\ 1]\,[\nabla_v \alpha_0]\, X_{2j} \mid X_j\right) = [P_j\ 1]\,[\nabla_v \alpha_0]\, X_{2j}.$$

In reporting results below, we simply use $\hat{Z}_j' = \left[\hat{P}_j\ 1\right] t_j X_{2j}$ as the instrument.

The results from an illustrative Monte-Carlo experiment are as follows:

Table I: Simulation Results

	homoskedastic model			heteroskedastic model		
	estimated		*truth*	*estimated*		*truth*
	mean	*std*		*mean*	*std*	1.00
parameter	1.05	.10	1.00	1.05	.09	1.00
marginal effect overall	1.02	.24	1.00	1.00	.16	1.00
marginal effect in 1st quartile	$-.15$	.37	$-.20$	$-.16$	.29	$-.20$
marginal effect in 2nd quartile	0.67	.30	0.65	0.66	.21	0.65
marginal effect in 3rd quartile	1.35	.29	1.35	1.34	.19	1.35
marginal effect in 4th quartile	2.18	.37	2.19	2.17	.27	2.19

The Monte-Carlo results shows that our estimator performs very well in finite samples. For the parameter estimates, the bias and standard deviations are small for both models. If we had employed the optimal instrument, then one would expect better performance under the homoskedastic design. The completed paper will investigate this issue in a variety of designs.

When we look at the marginal effect estimates, it is well estimated overall. Since our model is not linear, we report the marginal effects at different points of the index distribution as well. Basically we report the marginal effect averaged over observations with V_1 in the each of the four quartiles. Our results show that the marginal effects were well estimated overall.

6 Conclusions

In conclusion, we have proposed estimators for both semiparametric and nonparametric models where the discrete treatment is endogenous. Both the outcome model and the treatment model accommodate a large variety of different model structures. The outcome model allows flexible interactions between the endogenous treatment and other exogenous variables. Further, our estimator has the desirable large sample properties and performs well even when the treatment model is misspecified. Such robustness against misspecification is achieved by using IV in estimating both marginal effect functions and index parameters. For the case when the model is correctly specified, we proved the efficiency of the estimator. Monte Carlo results show that the estimator performs well in finite samples.

References

[1] Ai, C. and X. Chen (2003): "Efficient estimation of models with conditional moment restrictions containing unknown functions," *Econometrica*, 71, 1795–1843.

[2] Ai, C. and X. Chen (2007): "Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables," *Journal of Econometrics*, vol. 141, issue 1, pages 5-43.

[3] Amemiya, T. (1974): "The Non-linear Two-stage Least Squares Estimator," *Journal of Econometrics*,2, 105-110.

[4] Amemiya, T. (1977): "The Maximum Likelihood and Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model," *Econometrica*, 45, 955-968.

[5] Blundell, R. and J.L. Powell (2001): "Endogeneity in Nonparametric and Semiparametric Regression Models," presented at the World;d Congress of the Econometric Society, 2000.

[6] Blundell, R. W. and J.L. Powell (2004): "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, 71, 655-679

[7] Cai, Z., M. Das, H. Xiong, and X. Wu (2005): "Functional Coefficient Instrumental Variable Models, " *Journal of Econometrics*, 133, 207-241.

[8] Card, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric. Problems," *Econometrica*, Vol. 69, No. 5, pp. 1127-1160.

[9] Das, M. (2005): "Instrumental Variables Estimators of Nonparametric Models with Discrete Endogenous Regressors," *Journal of Econometrics*, 124, 335-361.

[10] Fan, J. and I. Gijbels (1996): *Local Polynomial Modeling and Its Applications*, Chapman and Hall, Monographs on Statistics and Applied Probability, 66.

[11] Hausman, J.A. (1983): "Specification and Estimation of Simultaneous Equation Models," Chapter 7 in *Handbook of Econometrics*, V. 1, Elsevier, Z. Griliches and M.D. Intriligator editors, 391-448.

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

[13] Imbens, G. and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, V. 77, No. 5, 1481-1512.

[14] Lewbel, A. (2000): "Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables ", Journal of Econometrics, 97, 145–177.

[15] Newey, W. K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," Chapter 9 in *Handbook of Econometrics*, V. 4, Elsevier, R. F. Engle and D. L. Mcfadden editors, 2111-2245.

[16] Newey, W. K. (1990): "Efficient Instrumental Variable Estimation of Nonlinear Models, *Econmometrica*, JV. 58, No. 4, 809-817.

[17] Newey, W. K., J.L. Powell, and F. Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Models," *Econometrica*, 67, 565–603.

[18] Newey, W. K. and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, v. 71, No. 5, 1565-1578..

[19] Rothe, C. (2009): "Semiparametric Estimation of Binary Response Models with Endogenous Regressors," *Journal of Econometrics*, V. 153, 61-64.