2010

# A kernel weighted smoothed maximum score estimator for the endogenous binary choice model

Jerome Krief
*Louisiana State University and Agricultural and Mechanical College*

# A KERNEL WEIGHTED SMOOTHED MAXIMUM SCORE ESTIMATOR FOR THE ENDOGENOUS BINARY CHOICE MODEL

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Economics

by
Jerome Krief
Mairise, Paris IX Dauphine, 2000
M.S., Louisiana State University, 2007
December 2010

**Acknowledgements**

I am indebted to three professors who have influenced me as a researcher. First, I would like to thank Dr. R.Carter Hill for having taught me the art of econometrics. Also, I would like to thank Dr. Joel Horowitz for having encouraged me to pursue this dissertation topic. Finally, I would like to thank Dr. Leonard Richardson for having provided me with solid foundations in Real Analysis.

Also, I thank the participants of the 2010 *Netherlands Econometric Study Group* for their comments. Moreover, I thank Dr. R. Kaj Gittings and Dr. Bulent Unel for suggestions. Lastly, I am grateful to Chris Raschke and Paul Darby for their aid in handling the dataset.

**Table of contents**

**Abstract**

This paper considers a local control function approach for the binary response model under endogeneity. The objective function of the Smoothed Maximum Score estimator (SMSE)(Horowitz 1992) is modified by weighting the observations with a kernel. Under some mild regularity conditions similar in nature to those of the SMSE, the consistency of this Kernel Weighted Smoothed Maximum Score estimator is established. Under some reasonable smoothness conditions the estimator's asymptotic normality is derived with a convergence rate in probability of at least $n^{-3/8}$ which can be rendered arbitrarily close to $n^{-1/2}$ as the regularity conditions improve. Additionally, the covariance matrix of the limiting distribution can be estimated consistently from data, permitting convenient inferences. Under stronger regularity conditions, an alternative C.A.N. estimator using a two stage procedure via sieves is shown to achieve a faster rate of convergence in probability. Some Monte Carlo experiments are conducted highlighting the robust advantage of these estimators. Finally, these estimation techniques are used to assess the determinants of maternal pregnancy smoking using the 1988 National Health Interview Survey.

## 1. Introduction

This paper considers the endogenous binary choice model of the form:

$$\text{(i) } U = \dot{X}'\beta + \varepsilon,$$

$$\text{(ii) } A = \Pi' W + V,$$

$$\text{(iii) } Y = d(U) \text{ with } d(.) \equiv \mathbf{1}[. \geq 0],$$

where $Y$ is the observable response variable, $\dot{X}' \equiv (Z', A)$ is a $1 \times K$ observable vector, $W$ a $q \times 1$ observable vector, $(\varepsilon, V)$ are unobservable errors, $\Pi$ is a $q \times 1$ unknown parameter and $\beta$ a $K \times 1$ parameter of interest. Write $\tilde{W}$ as the components of $W$ which are excluded from $\dot{X}$. Here the vector $S \equiv (Z', \tilde{W}')$ contains exogenous instruments while $A$ is the endogenous variable due to the correlation between $\varepsilon$ and $V$. For simplicity assume that $\dot{X}$ contains no intercept since it is not identifiable under the estimation technique which is to be discussed soon. Under appropriate identification restrictions the results put forth in this dissertation are easily generalizable when $A$ is a vector and $\Pi'$ a matrix. Importantly, the proposed estimator allows for powers of the endogenous variable.

In the economics literature the latent variable $U$ usually represents the agent's willingness to pay, or the difference in utility between two mutually exclusive alternatives. This model may have an omitted variable interpretation when $A$ is correlated with $\varepsilon$ through some unobservable factors. The model also has an *errors in the variables* interpretation when $A$ represents a misreported variable. Here are some (simplified) examples taken from the economics literature where the above endogenous binary model applies:

**Example 1**: **Labor force participation of men without college education**, Powell and Blundell (2004).

Let $Y = 1$ if a man without a college education works. Equation $(i)$ applies, with $Z$ containing the years of education of the men and $A = log(spouseinc)$ where *spouseinc* is the income of his spouse. According to economic theory the spouse's income is endogenously given by a Mincer's equation. The authors use $(ii)$ with $W' = (spouseduc, log(benef))$ where *spouseduc* is the years of education of the spouse and *benef* is the monetary amount of welfare entitlement combining child benefit, unemployment benefit and other allowances. Here $\varepsilon$ contains unobservable factors which drive the man's labor force decision such as his family background, while $V$ includes unobservable variables driving the spouse's income such as her family background. It is expected that the slope coefficient of $log(spouseinc)$ is negative since a higher extra source of income gives less incentive to search for a job. However, given that married individuals tend to share some common attributes $\varepsilon$ and $V$ are positively correlated. Using a probit (or logit) regression of $Y$ on $Z, A$ will yield misleading estimate, in effect underestimating the importance of the spouse's income as a work disincentive.

**Example 2**: **Stock option and earnings manipulation**, Burns and Kedia (2004).

Let $Y = 1$ if a firm restates its earnings. Equation $(i)$ applies with $Z$ containing a firm's financial characteristics such as its debt, liquidity and spending on research and development while $A = log(delta * shares)$ where *delta* is the delta of the option on the firms' stock (i.e. the derivative of the option value with respect to its stock price in the Black and Scholes Option Pricing Model) and *shares* indicate the number of shares granted to the managers. Thus $delta * shares$ measures the potential gain in stock option value for a small increase in stock price. The number of shares granted is partly determined by the labor market characteristics for the industry in which the firm operates. Hence, the authors use $(ii)$ with $W' = (Labor', Z')$ where $Labor'$ is a vector of labor market characteristics. Here $\varepsilon$ contains unobservable factors which promote earning restatements while $V$ include unobservable variables driving the stock option value. It is expected that the slope coefficient of $log(delta * shares)$ is positive. However, there are unobservable attributes for a firm such that the CEO's risk aversion, growth potential which affect both restatement and the stock value, therefore inducing a correlation between $\varepsilon$ and $V$. Using a probit (or logit) regression of $Y$ on

$Z, A$ will yield misleading estimates, in effect overestimating the effect of stock option as an incentive for earnings'manipulation.

**Example 3**: **Foreign direct investment and spill-over on exports**, Aitken et al, (1997).

Let $Y = 1$ if a domestic firm exports goods. Equation $(i)$ applies with $Z$ containing the cost attributes of the firm such as its labor cost, capital cost and transportation cost while $A = log(FDI)$ where $FDI$ is the amount of foreign direct investment in the region where the firm operates. Since the level of $FDI$ received by a region is to a larger extent the product of a cost benefit analysis from foreign firms, the authors use $(ii)$ with $W' = (foreignwage, foreignlaborVA, foreignlaboroutput, Z')$ where $foreignwage$ indicates the foreign real wage for the industry in which the firm operates, $foreignlaborVA$ measures the foreign labor share of value added and $foreignlaboroutput$ the foreign labor share of output. Here $\varepsilon$ contains unobservable factors influencing the decision of whether to export while $V$ includes unobservable characteristics of the region which are relevant for foreign firms. It is expected that the slope coefficient of $log(FDI)$ is positive since a larger amount of $FDI$ in a region may facilitate exports notably via better infrastructure. However, $\varepsilon$ and $V$ share common variables rendering both exports and FDI more appealing such as the quality of the regional labor force. Using a probit (or logit) regression of $Y$ on $Z, A$ will yield misleading estimates, in effect overestimating the effect of FDI on exports.

## 2. Literature, Motivation and Summary of Contribution

In principle when either $(\varepsilon, V)|S$ or $\varepsilon|S, V$ has a distribution function known up to some finite dimensional parameter, one may estimate $\beta$ consistently via maximum likelihood (ML). A vast literature assumes this is the case with a normal homoscedastic distribution posited for $(\varepsilon, V)|S$ such as in Heckman (1978), Amemiya (1978), Lee (1981) and Newey (1987) or for $\varepsilon|S, V$ as in Smith and Blundell (1986) and Rivers and Vuong (1988). If the parametrization of the distribution in question is incorrect, those estimators will be inconsistent. As a result, new semi-parametric estimators have been proposed, relaxing this parametric requirement. For instance, the quasi-ML estimator developed in Rothe (2009) is consistent for $\beta$ whenever the distribution function of $\varepsilon|\dot{X}, V$ depends only on $\dot{X}'\beta$ and $V$. Also, the two stage least square estimator proposed in Lewbel (2000) is consistent for $\beta$ provided there exists a special regressor in $\dot{X}$ meeting a certain conditional independence restriction. Even though these semi-parametric estimators offer a robust advantage, they present some limitations in terms of either the permitted form of heteroscedasticity (Rothe 2009) or which variables affect the conditional variance of both $\varepsilon$ and $V$ (Lewbel 2000). This is due to the very nature of their distributional oriented assumptions.

Estimators that are robust to unknown heteroscedasticity are based instead on some conditional median restrictions which loosely speaking only require the center of the distribution of $\varepsilon$ to remain unaffected by the covariates. For instance, Newey (1985) provided a consistent asymptotically normally distributed two stage maximum score estimator for $\beta$ under the requirement that $(V, \varepsilon)$ be symmetrically distributed around the origin, conditional on $S$. Also, Hong and Tamer (2003) proposed a consistent minimum distance estimator for $\beta$ under the less restrictive condition that $Med(\varepsilon|S) = 0$. However, in Newey (1985) a consistent estimator for the asymptotic covariance is not provided (see Newey 1985, page 228) while Hong and Tamer's estimator has an unknown limiting distribution.

The main motivation behind this dissertation is to remedy this inferential problem, offering a consistent estimator of $\beta$ under a weak median restriction which also allows for testing. The main estimator presented in this article, named the Kernel Weighted Smoothed Maximum Score (KWSMS) estimator, meets these objectives. The KWSMS estimator is constructed by imposing a restriction on $Med(\varepsilon|S, V)$ which must not vary with the instrument $S$. This ensures the existence of some random variable $\phi$ and unobservable term $e$ such that $Y = d(\dot{X}'\beta + \phi + e)$ where now $e$ satisfies the classic median restriction introduced for maximum score estimation (Manski 1985). Then, a smoothed maximum score estimation (Horowitz 1992) is performed as if $\phi$ were a constant, correcting this approximation by means of a kernel. Doing so facilitates the asymptotic analysis using the framework laid out in Horowitz (1992). An interesting additional contribution of this article is in fact to offer a robust estimation procedure for a semi-linear random utility model.

Not surprisingly, this estimation approach imposes stronger assumptions than those required from the SMSE albeit similar in essence. The KWSMS estimator's consistency for $\beta$ (up to a positive scale) requires that one element of $\dot{X}$ be fully supported and that the endogenous variable be continuous. Additionally, if certain cumulative distribution functions involving the random variables $V$ and $\dot{X}'\beta$ are sufficiently differentiable then the KWSMS estimator is asymptotically normally distributed provided the fourth moments of $\dot{X}$ exist. Finally, the KWSMS estimator say $\beta_n$ satisfies $\beta_n - \beta = O_p(n^{-\frac{1}{2}+\kappa})$ for some $\kappa \in (0, 1/8)$ where $\kappa$ becomes arbitrarily small under adequate regularity conditions. Hence, the parametric rate is potentially achievable.

This paper relates to the previous literature using the control function approach which has already been employed to handle endogeneity in the context of binary choice models (Blundell and Powell 2004), triangular equation models (Newey, Powell and Vella 1999) and quantile regression models (Lee 2007). Also, the technique used to derive the asymptotic results is similar to that of the SMSE using nonparametric convolution based arguments. Finally, its local nature can be thought as a smoothed version of the local quantile regression estimator (Chaudhuri 1991, Lee 2003) in the context of the random utility model.

As explained in Section 4, a KWSMS estimator in effect uses only observations of $V$ close to a given value. This local nature suggests that the rate of convergence can be accelerated by using all the observations of $V$ instead. Thus, in this paper a second stage estimation is offered with a Score Approximation Smoothed Maximum Score (SASMS) estimator which uses the information content from various KWSMS estimators retrieved in a first stage estimation. Under stronger regularity conditions the SASMS estimator is still consistent and asymptotically normally distributed while achieving a faster rate of convergence in probability.

The rest of the paper is organized as follows. Section 3 provides a review of the control function approach in the context of this binary choice model. Section 4 describes the KWSMS estimator and summarizes its asymptotic properties. Section 5 describes the SASMS estimator and summarizes its asymptotic properties. Section 6 contains some Monte Carlo simulations to illustrate the finite sample qualities of the suggested estimators. Finally, Section 7 applies these estimation techniques using data from the 1988 National Health Interview Survey to determine the factors influencing maternal pregnancy smoking.

## 3. Estimation Strategy

The key condition introduced in this paper is that there exists some $\overline{v}$ in the support of $V$ satisfying:

$$Med(\varepsilon|Z, W, V = \overline{v}) = Med(\varepsilon|V = \overline{v}) \quad (1)$$

Loosely speaking, (1) imposes that once $V$ has been fixed at $\overline{v}$, the center of the distribution of $\varepsilon$ does not vary with the exogenous variables. The equality in (1) will be met for instance when $(Z, W)$ and $(\varepsilon, V)$ are statistically independent or under a conditional independence restriction of the form $\varepsilon|Z, W, V \sim \varepsilon|V$, but those are not necessary. This key median assumption, which can be tested from data as explained in Section 4.3, is neither stronger nor weaker than that assumed in Hong and Tamer (2003) because each restriction can imply the other under certain conditions. This median restriction can accommodate heteroscedasticity in $V$ of virtually any form in the error term.

Now suppose that (1) holds for an arbitrary $\overline{v}$. As will be explained shortly, this is stronger than required for the KWSMS estimator but is needed for the SASMS estimator (at least over a range of values for $v$). Invoking this last condition and the fact $(\dot{X}, V)$ is one to one with $(Z, \Pi' W, V)$ yields:

$$Med(\varepsilon|\dot{X}, V) = Med(\varepsilon|V),$$

and noting $\phi(V) = Med(\varepsilon|V)$ thus provides:

$$Med(U|\dot{X}, V) = \dot{X}'\beta + \phi(V), \quad (2)$$

showing that the restriction in (1) treats endogeneity as an omitted variable problem. The conditional median in (2) becomes the starting point for consistent estimation since by the quantile invariance property to monotonic transformations (Powell 1986) one derives :

$$Med(Y|\dot{X}, V) = d(\dot{X}'\beta + \phi(V))$$

This conditional median restriction on the response variable $Y$ is, up to the nuisance parameter $\phi(.)$, identical to the restriction for maximum score estimation proposed in Manski (1985). *A priori*, the control function $\phi(.)$ has an unknown form. However, when $V$ is fixed at some given $v$, the nuisance $\phi(.)$ becomes a constant and the lack of knowledge on $\phi(.)$ is no longer a problem. This fixing is the foundation of the estimation procedure elaborated in this article. This principle is analogous to that used in the literature for unspecified quantile regression (Chaudhuri 1991) or semi-linear quantile regression (Lee 2003).

## 4. Description of the KWSMS Estimator

### 4.1 Identification

Define $\Pi_{\tilde{w}}$ and $\Pi_z$ from $\Pi' W = \Pi'_{\tilde{w}} \tilde{W} + \Pi'_z Z$ where $\tilde{W}$ contains exogenous variables excluded from $Z$. The parameter of interest $\beta$ is only identifiable up to a positive scale since $d(\eta U) = d(U)$ for any scalar $\eta > 0$. The identification of $\beta$ up to a positive scale requires three main conditions. The parameter $\Pi_{\tilde{w}}$ must be non-null, that is, $W$ contains some variable excluded from $Z$ having an effect on the endogenous variable. Also, one element of $\dot{X}$ conditional on its remaining elements needs to admit a distribution function absolutely continuous with respect to Lebesgue measure. Let $(C, \widetilde{X}')$ be a partition of $\dot{X}'$ such that the scalar variable $C$ satisfies this property, with an associated slope coefficient noted $\beta_1$. Finally, identification up to scale requires $V|\dot{X}$ to admit a Lebesgue density. These combined with [1] and some mild conditions suffice for identification up to the scaling factor $1/|\beta_1|$ whenever $\beta_1 \neq 0$. From now on assume without loss of generality that $\beta_1$ is known to be strictly positive.

It is useful to illustrate the relevance of those conditions using a simple example of the form $U = Z\lambda + \delta A + \varepsilon$ with $A = \pi W + V$ where $(Z, W, V)$ are three scalar variables and $(\lambda, \delta, \pi)$ real parameters. For simplicity further assume that $Z$ is independent with $(V, W)$. Since here $\dot{X}' = (Z, A)$ one condition for identification as explained above is that the variable $V|Z, A$ is continuous. Suppose first that $W$ is some function of $Z$. Then $V$ becomes a deterministic function $(A, Z)$ and $V|Z, A$ is a single atom thus not continuously distributed. Evidently, even if $W$ is not a function of $Z$ the same problem arises if $\pi = 0$. More generally, this illustrates the importance of having one component in $W$ which is not only excluded from $Z$ but also not a function of $Z$ and which has an impact on the endogenous variable. Suppose now that this the case. Since $V|Z, A \equiv V|Z, \pi W + V$ and $Z$ is independent with $(V, W)$ the required continuity thus deals here with the distribution of $V|\pi W + V = a$ which admits a Lebesgue density as soon as $V|W$ does[1]. Thus, by construction the variable $A$ must be continuous for being able to identify $\beta$ up to scale. Clearly, this estimation technique excludes binary choice models where the endogenous variable is discrete .

### 4.2 Estimation Procedure and Asymptotic Properties

Let $\{Y_i, \dot{X}_i\}_{i=1}^n$ be a random sample from $(Y, \dot{X})$. Also, let $\{\hat{V}_i\}_{i=1}^n$ be residuals with $\hat{V}_i \equiv A_i - \hat{\Pi}' W_i$ where $\hat{\Pi}$ is a given root $n$ consistent estimator of $\Pi$. Under the mild assumptions for M-estimators root $n$ consistency will be attained. The simplest estimator for $\Pi$ when $W$ is exogenous is probably the OLS if $V$ and $W$ are uncorrelated. There are two cases worth mentioning which do not *a priori* meet the model for equation (ii) but which allow the results to be still valid. The first case is when $A = \Pi(W, \delta) + V$ where $\Pi(., \delta)$ is a parametric function for some unknown $\delta$. Then If $(V, W)$ are uncorrelated, one can derive via non-linear least squares the estimator $\hat{\delta}$ (Amemiya 1985) and residuals $\hat{V}_i = A_i - \Pi(W_i, \hat{\delta})$ which conserves our results. The second case is when $A = \Pi(W) + V$ where $\Pi(.)$ is some unknown function and $W$ contains only discrete variables whose support is bounded. Then if $E[V|W] = 0$, one can estimate non parametrically $\Pi(.)$ point wise at the parametric rate (Bierens 1987) and the residuals $\hat{V}_i \equiv A_i - \hat{\Pi}(W_i)$ still satisfy the assumptions needed for the KWSMS estimator.

It is convenient at this stage to introduce some notations. For f:$\mathbb{R} \longrightarrow \mathbb{R}$ define $f^{(j)}(t)$ as its $j^{th}$ derivative at $t$ whenever this latter exists. Also, write $L^2[0, 1]$ the space of Lebesgue measurable real-valued functions from $[0, 1]$ to the real line which are square integrable with respect to Lebesgue measure.

Let $\alpha_i \equiv 2Y_i - 1$ and $X' \equiv (1, \tilde{X}')$. The KWSMS estimator, noted $\widetilde{\theta_n}$, is defined as the maximizer in $\theta$ of the following objective:

$$\widetilde{S_n}(\theta) \equiv \frac{1}{nh_q} \sum_{i=1}^{n} \alpha_i D(\frac{C_i + X_i'\theta}{h}) k(\frac{\hat{V}_i - \overline{v}}{h_q}),$$

where $(\{h_q\}_n, \{h\}_n)$ is a given pair of strictly positive bandwidth sequences vanishing to 0 as $n$ approaches infinity and $D(.)$ is some chosen bounded function from the real line into itself meeting:

[1]In that case the density is given by $f(v|a) = p_{VW}(v, \frac{a-v}{\pi}) / \int p_{VW}(v, \frac{a-v}{\pi}) dv$ where $p_{VW}$ indicates the probability density function of $(V, W)$.

$$lim_{t\to-\infty}D(t)=0,\ lim_{t\to\infty}D(t)=1,$$

and

$$D'=K \text{ everywhere with } |K(t)|<M_1 \text{ for some finite real number } M_1.$$

This function $D(.)$, whose tail behavior mimics that of a cumulative distribution function, introduces the building block for deriving an asymptotic theory. This permits us to approximate, after tuning with the bandwidth $h$, the indicator variable. Simultaneously this allows us to easily derive a limiting distribution for the estimator because the score of the objective will have a Taylor's expansion as soon as $K$ is itself differentiable. For instance, the cumulative distribution function of the standard normal distribution meets these conditions. Because of the subsequent asymptotic conditions, a natural choice for $D(.)$ is to use the antiderivative of a kernel that is compactly supported (see Müller 1984). A good example for such function (apart from the lack of differentiability for $|t|=1$) is given by:

$$D(t)=[0.5+\tfrac{105}{64}(t-\tfrac{5}{3}t^3+\tfrac{7}{5}t^5-\tfrac{3}{7}t^7)]1[|t|\leq 1]+1[t>1].$$

The function $k(.)$ is a given kernel satisfying notably,

$$\textstyle\int k(t)dt=1,\ \int t^u k(t)dt=0 \text{ for } u=1,...,m-1,$$

$$\textstyle\int |t^u k(t)|dt<\infty \text{ for } u=0,m \text{ for some } m\geq 2,\ \int |k(t)|^2 dt<\infty,$$

and

$$k \text{ is differentiable everywhere with } |k^{(1)}(t)|<M_2 \text{ for some finite real number } M_2.$$

That is, $\widetilde{S_n}$ is similar to the objective of the SMSE (had $V$ been fixed at $\overline{v}$) apart from our weighting the $i^{th}$ observation with $\frac{1}{h_q}k(\frac{\hat{V}_i-\overline{v}}{h_q})$. The above integrability conditions for $k(.)$ are met using a kernel of order $m$. For consistency purposes $m=2$ suffices. However, obtaining asymptotic normality for the KWSMS estimator requires $m\geq 7$.

### 4.2.1 Consistency

Suppose that $\phi(\overline{v})\equiv Med(\varepsilon|V=\overline{v})$ exists. Define $\tilde{\beta}$ the slope coefficient associated to $\tilde{X}$ and write $\ell\equiv C+X'\theta_0$ where $\theta_0'\equiv\frac{1}{\beta_1}(\phi(\overline{v}),\tilde{\beta}')$. Introduce $F_{X,\ell,V}[.]$ the cumulative distribution function of $\varepsilon|X,\ell,V$ and $f_{X,\ell}(.)$ the Lebesgue density of $V|X,\ell$. This last density exists by the identification conditions because $\dot{X}$ is one *to* one with $(X,\ell)$. Suppose that on some open neighborhood of $\overline{v}$ the functions $v\mapsto F_{X,\ell,v}[-\beta_1\ell+\phi(\overline{v})]$ and $v\mapsto f_{X,\ell}(v)$ are continuous. Also, assume that the bandwidth sequences are chosen to satisfy lim $nh_q^4=\infty$ and lim $\frac{nh^2h_q^2}{log(n)}=\infty$ as $n\to\infty$. Under these and some mild regularity conditions the KWSMS estimator will be consistent for $\theta_0$.

### 4.2.2 Asymptotic Normality

Define $F_{X,\ell,\overline{v}}[.]$ the distribution function of $\varepsilon|X,\ell,V=\overline{v}$ and $f_X(\ell)$ the Lebesgue density of $\ell|X$. This last density is well defined under the identification requirement that the distribution of $C|\widetilde{X}$ be absolutely continuous with respect to Lebesgue measure because of the one *to* one relationship between $(X,\ell)$ and $\dot{X}$. Also, write $\mu_X(\ell)\equiv f_{X,\ell}(\overline{v})f_X(\ell)$ and $F^{(1)}_{X,\ell,\overline{v}}[-\beta_1\ell+\phi(\overline{v})]\equiv\partial F_{X,\ell,\overline{v}}[-\beta_1\ell+\phi(\overline{v})]/\partial\ell$ whenever the derivatives exist. Suppose that both $\Sigma_0\equiv\int|k|^2\int|K|^2E[XX'\mu_X(0)]$ and $H_0\equiv 2E[XX'F^{(1)}_{X,0,\overline{v}}[\phi(\overline{v})]\mu_X(0)]$ exist with the latter matrix negative-definite.

Now assume that as functions of $v$, $F_{X,\ell,v}[-\beta_1\ell+\phi(\overline{v})]$ and $f_{X,\ell}(v)$ are $m$ times differentiable on some open neighborhood of $\overline{v}$ for some $m\geq 7$. Also, assume that as functions of $\ell$, $F_{X,\ell,\overline{v}}[-\beta_1\ell+\phi(\overline{v})]$ , $f_{X,\ell}(\overline{v})$ and $f_X(\ell)$ are $r$ times differentiable everywhere for some $r\geq 2$. Furthermore, choose $K$ to satisfy notably,

$$\int K(t)dt = 1,\ \int t^u K(t)dt = 0 \text{ for } u = 1, ..., r-1 \text{ and } \int |t^u K(t)|dt < \infty \text{ for } u = 0, r$$

$K$ is symmetrical, twice differentiable everywhere, $|K^{(j)}(t)| < B$ for $j = 1, 2$ where $B$ is some finite real number and $\int |K^{(1)}(t)|^2 dt < \infty$.

Finally, select the bandwidths $h \propto n^{-a}$ and $h_q \propto n^{-a_q}$ where $a$ and $a_q$ are chosen according to the following:

$$a \in (sup\{\tfrac{1}{1+\eta+2\eta m}; \tfrac{1}{1+\eta+2r}\}, \tfrac{1}{4+4\eta}) \text{ and } a_q = \eta a \text{ for some } \eta \in (\tfrac{3}{2m-3}, \tfrac{1}{3}).$$

These combined with some mild technical conditions permit to establish:

$$\sqrt{nhh_q}(\widetilde{\theta_n} - \theta_0) \rightarrow_d \mathcal{N}(0, \Omega),$$

where $\Omega \equiv H_0^{-1} \Sigma_0 H_0^{-1}$ can be estimated consistently from data according to the following:

Let $\widetilde{H_n} \equiv \frac{1}{nh^2 h_q} \sum_{i=1}^n \alpha_i X_i X_i' K^{(1)}(\frac{C_i + X_i'\widetilde{\theta_n}}{h}) k(\frac{\hat{V}_i - \overline{v}}{h_q})$,

and

$\widetilde{\Sigma_n} \equiv \frac{1}{nh^{\gamma_1} h_q^{\gamma_2}} \sum_{i=1}^n X_i X_i' |K(\frac{C_i + X_i'\widetilde{\theta_n}}{h^{\gamma_1}})|^2 |k(\frac{\hat{V}_i - \overline{v}}{h_q^{\gamma_2}})|^2$,

for some constant $\gamma_1 \in (0, 3/4]$ and $\gamma_2 \in (0, 1]$. Then under the previous assumptions:

$$\widetilde{H_n} \longrightarrow_p H_0,$$

and

$$\widetilde{\Sigma_n} \longrightarrow_p \Sigma_0.$$

Thus, if the data set is large, the testing of hypothesis can be based upon the asymptotic approximation:

$$\sqrt{nhh_q}(\widetilde{\theta_n} - \theta_0) \sim \mathcal{N}(0, \widetilde{H_n}^{-1} \widetilde{\Sigma_n} \widetilde{H_n}^{-1})$$

**Remarks:**

($a$) From the asymptotic result one concludes that $\sqrt{nhh_q}(\widetilde{\theta_n} - \theta_0)$ is bounded in probability. It follows by the bandwidths conditions previously enumerated in Section 4.2.2 that the KWSMS estimator satisfies at least $\widetilde{\theta_n} - \theta_0 = O_p(n^{-3/8})$. However, this rate accelerates when $\lambda \equiv Min\{m, r\}$ augments and the KWSMS estimator eventually reaches the parametric rate, i.e. $O_p(n^{-1/2})$ as $\lambda$ approaches infinity.

($b$) The KWSMS estimator has an asymptotically centered normal distribution because the bandwidths pair has been selected purposefully such that the asymptotic bias vanishes. As established in Horowitz (1992) this is not optimal from an asymptotic mean squared error perspective which requires some strictly positive finite bias. This choice is driven by two considerations. First, the construction of an asymptotically biased KWSMS estimator would impose additional regularity conditions. Secondly, the unbiased SMSE has superior bootstrapping properties than the biased SMSE (see Horowitz 2002) in terms of the accuracy of its bootstrapped critical values which suggests the analogue for the KWSMS estimator since the objective of the KWSMS estimator is just a weighted version of SMSE's objective.

($c$) The maximization of the objective function will be carried out by an iterative procedure such as the quadratic hill climbing (Goldfeld, Quandt and Trotter 1966). Additionally, the starting value for the iterative search may be better chosen as a result of some annealing procedure (Szu and Hartley 1987).

### 4.3 Testing the Key Median Restriction

If assumption (1) is violated then the KWSMSE is inconsistent. Thus, it is important to have a testing procedure which can reveal from data the plausibility of this assumption. To sketch how to perform the testing of (1) suppose that the assumptions of Section 4.2.2 hold. Let $\alpha(Y_i) \equiv 2Y_i - 1$ and write $\ell_i \equiv C_i + X_i^{'}\theta_0(\bar{v})$ where $\theta_0(v)' \equiv \frac{1}{\beta_1}(\phi(v), \tilde{\beta}^{'})$ and $\hat{\ell}_i \equiv C_i + X_i^{'}\widetilde{\theta_n}$. Here $\bar{v}$ is the value chosen to compute the KWSMS estimator. Define the following statistic:

$$T_n \equiv \frac{(n\xi^2)^{-1}\sum_{i=1}^{n}\varphi(\frac{\hat{\ell}_i}{\xi})\varphi(\frac{\hat{V}_i-\bar{v}}{\xi})\alpha(Y_i)}{(n\xi^2)^{-1}\sum_{i=1}^{n}\varphi(\frac{\hat{\ell}_i}{\xi})\varphi(\frac{\hat{V}_i-\bar{v}}{\xi})},$$

where $\varphi$ is a kernel and $\xi$ a deterministic sequence. Introduce $f(.,.)$ the joint density of $(\ell, V)$ and $M(l,v) \equiv E[\alpha(Y)|\ell = l, V = v]$. The idea behind the test is analogous to that provided in Horowitz (1993), Proposition 2. The test is based upon the fact that under $H_o$: $Med(\varepsilon|\dot{X},\bar{v}) = Med(\varepsilon|\bar{v})$ one must have $M(0,\bar{v}) = 0$. But under certain mild conditions $T_n$ is consistent for $M(0,\bar{v})$. Thus, the test consists of measuring $|T_n|$ with large values undermining the validity of our median restriction.

More formally, suppose that $M(l,v)$ and the density of $(\ell, V)$ are twice differentiable on some open neighborhood of $(0,\bar{v})$, $\varphi$ is a strictly positive kernel of order 2, $\xi_n$ is a strictly positive sequence of real numbers satisfying $\xi \propto n^{-\omega}$ for some $\omega \in (sup\{1/10; a(1+\eta)\}, 1/5)$ where $a$ and $\eta$ are the bandwidth parameters selected to compute the KWSMS estimator as defined in Section 4.2.2. These regularity conditions combined with some further smoothness conditions suffice to establish that under the null hypothesis $H_o$: $Med(\varepsilon|\dot{X},\bar{v}) = Med(\varepsilon|\bar{v})$,

$$\sqrt{n\xi^2}T_n \to_d \mathcal{N}(0, f(0,\bar{v})^{-1}||\varphi||_{L2}^4),$$

where $||\varphi||_{L2} \equiv \int(|\varphi(t)|^2 dt)^{1/2}$. Furthermore,

$$(n\xi^2)^{-1}\sum_{i=1}^{n}\varphi(\frac{\hat{\ell}_i}{\xi})\varphi(\frac{\hat{V}_i-\bar{v}}{\xi}) \to_p f(0,\bar{v})$$

Consequently, testing can be performed in practice from data using the asymptotic approximation:

$$\sqrt{n\xi^2}T_n \sim \mathcal{N}(0, \hat{f}(0,\bar{v})^{-1}||\varphi||_{L2}^4),$$

where,

$$\hat{f}(0,\bar{v}) \equiv (n\xi^2)^{-1}\sum_{i=1}^{n}\varphi(\frac{\hat{\ell}_i}{\xi})\varphi(\frac{\hat{V}_i-\bar{v}}{\xi}).$$

## 5. Accelerating Convergence: A Score Approximation Smoothed Maximum Score Estimator

As explained in the previous section, a KWSMS estimator in effect uses only observations of $V$ close to a given $\overline{v}$. One may seek to construct an alternative estimator with a faster rate of convergence by using more observations of $V$. The SASMS estimator described next can attain that target provided some stronger conditions hold, notably if $Med(\varepsilon|V = v)$ has enough derivatives. The basic intuition is that the control function smoothness compensates for the low degree of differentiability of the functions of $v$ and $\ell$ introduced in Section 4.2.2.

### 5.1. Description of the SASMS Estimator

Suppose now that [1] holds for an arbitrary $\bar{v} \in [0,1]$, which will be simply noted henceforth as $v$. The choice of $[0,1]$ is chosen here for the sake of simplicity but can be replaced by any compact set of the real line which is contained in the support of $V$ by means of an appropriate normalization. Define $e'_K = [O, I_{K-1}]$ the $K-1 \times K$ matrix where the first column is the zero vector, while $I_{K-1}$ represents the $K-1 \times K-1$ identity matrix and $e'_1$ the $1 \times K$ vector whose first entry is 1 and zero elsewhere. Let $\Theta$ be some compact set and for a given $v$ introduce the following:

$$\tilde{\theta}(v) \equiv Argmax_{\Theta} \frac{1}{nh_q} \sum_{i=1}^{n} \alpha_i D(\frac{C_i + X_i'\theta}{h})k(\frac{\hat{V}_i - v}{h_q}),$$

and

$$\tilde{\beta}(v) \equiv e'_K \tilde{\theta}(v) \text{ while } \tilde{\phi}(v) \equiv e'_1 \tilde{\theta}(v),$$

where $D(.)$, $k(.)$ and the bandwidth pair $(h, h_q)$ are as described in Section 4. Let $\{f_j\}_{j\geq 1}$ be a known basis of functions such that $\sum_{j=1}^{\rho} b_j f_j$ can approximate a smooth function of $[0,1]$ arbitrary well using some real sequence $\{b_j\}_{j\geq 1}$ and natural number $\rho$ large enough. Here are some easy examples taken from Chen (2007):

- Power series:

Let $Pol(\rho) = \{f : [0,1] \to \mathbb{R}, f(v) = \sum_{j=0}^{\rho} b_j v^j, b_j \in \mathbb{R}\}$ the space of polynomials on $[0,1]$ of degree less or equal to $\rho$. A differentiable function on $[0,1]$ can be approximated arbitrarily well by some element of $Pol(\rho)$ with $\rho$ large enough. Thus, here $f_j(v) = v^{j-1}$ for $j \geq 1$.

- Trigonometric cosine:

Let $cosPol(\rho) = \{f : [0,1] \to \mathbb{R}, f(v) = b_1 + \sum_{j=2}^{\rho} b_j \sqrt{2} cos(2\pi(j-1)v), b_1, b_j \in \mathbb{R}\}$ the space of cosinus polynomials on $[0,1]$ of degree less or equal to $\rho$. A differentiable function on $[0,1]$ (or merely a square integrable function on $[0,1]$) can be approximated arbitrarily well by some element of $cosPol(\rho)$ with $\rho$ large enough. Thus, here $f_j(v) = \sqrt{2}cos(2\pi(j-1)v)$ for $j \geq 2$ and $f_1(v) = 1$. This choice is particularly suited for the SASMS estimator because $\{f_j\}_{j\geq 1}$ forms an orthonormal basis of $L_2[0,1]$.

- Splines:

For a given a natural number $d$ define, $Spl(d+1, \rho) = \{f : [0,1] \to \mathbb{R}, f(v) = \sum_{j=0}^{d} a_j v^j + \sum_{j=1}^{\rho} b_j[(v - t_j)_+]^d, a_j, b_j \in \mathbb{R}\}$, the space of splines on $[0,1]$ of order $d+1$ where $(.)_+ = Max(., 0)$ and $(t_1, t_2, ...t_\rho)$ is a given increasing sequence of knots partitioning $[0,1]$ such that $t_1 = 0$ and $t_\rho = 1$. Here $\sum_{j=1}^{\rho} b_j[(v - t_j)_+]^d$ is a piecewise polynomial shifter which permits the adjustment of a baseline polynomial on each interval $I_j = [t_j, t_{j+1}]$. Define $|I_j| = t_{j+1} - t_j$ for $j = 1, ..., \rho - 1$. A differentiable function on $[0,1]$ can be approximated arbitrarily well by some element of $Spl(d+1, \rho)$ with $\rho$ large enough provided the mesh ratio $Max|I_j|/Min|I_j|$ stays bounded. Thus, here $f_j(v) = v^{j-1}$ if $1 \leq j \leq d+1$ and $f_j(v) = [(v - t_{j-d-1})_+]^d$ if $d+2 \leq j \leq d+1+\rho$.

Now define $p_n(.)' \equiv (f_1(.), ..., f_{\rho(n)}(.))$ where $\rho(n)$ is some chosen deterministic sequence of natural numbers satisfying $\rho(n) \to \infty$ as n $\to \infty$ but $\rho(n) < n$. Write $\Lambda_n$ the $n \times \rho(n)$ matrix whose $i^{th}$ row is $p_n(i/n)'$ and $\tilde{\phi}_n$ the $n \times 1$ vector whose $i^{th}$ entry is $\tilde{\phi}(i/n)$. That is, running a first stage estimation with n locals KWSMS estimators at $v = 1/n, 2/n, ..., 1$ (where $n$ still indicates the sample size) permits the collection of $\tilde{\phi}_n$ and to retrieve the following:

$$b_n \equiv Argmin_{b \in \mathbb{R}^{\rho(n)}} ||\tilde{\phi}_n - \Lambda_n b|| \equiv (\Lambda_n' \Lambda_n)^{-1} \Lambda_n' \tilde{\phi}_n. \tag{3}$$

This estimator $b_n$ is nothing but the OLS estimator of $b$ in the artificial regression model:

$$\tilde{\phi}(v) = b' p_n(v) + error \text{ using the fixed design } v = 1/n, 2/n, ..., 1.$$

To get some sense about the motivation behind (3) consider the case where the trigonometric cosine basis is chosen. Use the notation $< g_1, g_2 >= \int_{[0,1]} g_1(v) g_2(v) dv$ whenever $g_1$ and $g_2$ belong to $L_2[0,1]$. Recall that each local KWSMS estimators $\tilde{\phi}(v) \equiv e_1' \tilde{\theta}(v)$ for $v = 1/n, ..., 1$ estimates the (scaled) control function say $\phi(v)$ for $v = 1/n, ..., 1$. The trigonometric cosine sequence $\{f_j\}_{j \geq 1}$ constitutes an orthonormal basis of $L_2[0,1]$ which implies $< f_i, f_j >= 1$ if $i = j$ and $< f_i, f_j >= 0$ otherwise. Also, this implies that $\phi(.)$ (if square integrable on $[0,1]$) has the representation[2] $\phi = \sum_j \mu_j f_j$ where $\{\mu_j\}_{j \geq 1}$ are the Fourier coefficients meeting $\mu_j =< \phi, f_j >$. Thus, if the sample size is large enough, $\tilde{\phi}(v) \approx \phi(v)$ for $v = 1/n, ..., 1$. Also, because of our fixed design with $v = 1/n, ..., 1$ the matrix $\Lambda_n' \Lambda_n$ for $n$ large will be approximately equal to the $\rho(n)$ by $\rho(n)$ identity matrix since its $j^{th}$ diagonal element approximates $< f_j, f_j >= 1$ and its cross diagonal elements say $(i,j)$ approximates $< f_i, f_j >= 0$. Thus, what $b_n$ estimates in that case are the Fourier coefficients $\mu_j$ for $j = 1, 2, ..., \rho(n)$. As the sample size $n$ increases, $\rho(n)$ also increases allowing for the recovery of more and more Fourier coefficients and consequently a more accurate estimator for the control function.

This first stage estimation yielding (3) constitutes the essence of the SASMS estimator since for $\rho(n)$ well-chosen and under some regularity conditions, the function $b_n' p_n(.)$ is consistent for $\tilde{\phi}_0(.) = \frac{1}{\beta_1}\phi(.)$ in the sense that plim $sup_{v \in [0,1]} |b_n' p_n(v) - \tilde{\phi}_0(v)| = 0$. However, $\{V_i\}_{i=1}^n$ is not observed but only $\{\hat{V}_i\}_{i=1..n}$. Hence, a natural way to proceed is to estimate $\tilde{\phi}_0(V_i)$ with $b_n' p_n(\hat{V}_i)$ for $i = 1...n$. Let $\Psi(.)$ be some kernel (possibly different from the function $D'(.)$ used in the first stage) from the real line into itself whose derivative exists everywhere. Now define for an arbitrary $\beta$ the following:

$$G_n[\beta] \equiv \frac{1}{nh_*} \sum_{i=1}^{n} \tau(\hat{V}_i) \alpha_i \tilde{X}_i \Psi\left(\frac{C_i + \tilde{X}_i{}' \beta + b_n' p_n(\hat{V}_i)}{h_*}\right),$$

and

$$H_n[\beta] \equiv \frac{1}{nh_*^2} \sum_{i=1}^{n} \tau(\hat{V}_i) \alpha_i \tilde{X}_i \tilde{X}_i{}' \Psi^{(1)}\left(\frac{C_i + \tilde{X}_i{}' \beta + b_n' p_n(\hat{V}_i)}{h_*}\right),$$

where $\tau(.) \equiv \mathbf{1}[0 \leq . \leq 1]$ and $h_*$ is a deterministic strictly positive sequence of real numbers meeting lim $h_* = 0$ as $n \to \infty$. The SASMS estimator, noted $\bar{\beta}$, is given by:

$$\bar{\beta} \equiv \tilde{\beta}(v) - H_n[\tilde{\beta}(v)]^{-1} G_n[\tilde{\beta}(v)],$$

where $\tilde{\beta}(v)$ is the slope coefficient estimator of a KWSMS estimator using some fixed $v \in [0,1]$. The reader familiar with Horowitz (1992) would have noticed that $\bar{\beta}$ is an approximation for a feasible SMSE based upon [2] which would use $b_n' p_n(\hat{V})$ in lieu of $\phi(V)$ (up to a scale). This estimator belongs to the class of score approximation estimators (Stone 1975, Bickel 1982, Lee 2003).

[2] Strictly speaking this representation is to be understood in the sense that $lim_{N \to \infty} || \sum_{j=1}^N \mu_j f_j - \phi||_{[0,1]} = 0$ where $||g||_{[0,1]} \equiv \sqrt{\int_{[0,1]} |g(t)|^2 dt}$.

### 5.2 Asymptotic Properties

Assume that the conditions of section 4.2.2. hold for any $\bar{v} \in [0,1]$. Introduce $L_i \equiv \frac{1}{\beta_1} Med(U|\dot{X}_i, V_i)$. Define $F_{\tilde{x},l,v}[.]$ as the cumulative distribution function of $\varepsilon|X = \tilde{x}, L = l, V = v$ and $f_{\tilde{x},v}(.)$ the Lebesgue density of $L|X = \tilde{x}, V = v$. This last density exists as long as that of $C|\tilde{X} = \tilde{x}, V = v$ exists because $(L, X, V)$ is one *to* one with $(C, X, V)$. Also, adopt the convention $F^{(1)}_{\tilde{x},l,v}[-\beta_1 l + \phi(v)] \equiv \partial F_{\tilde{x},l,v}[-\beta_1 l + \phi(v)]/\partial l$ whenever this derivative exists. Suppose that $Q \equiv 2E[\tau(V)\tilde{X}\tilde{X}' F^{(1)}_{\tilde{X},0,V}[\phi(V)] f_{\tilde{X},V}(0)]$ exists and is negative-definite. The subsequent sections treat the case where the researcher selects either the power series or trigonometric cosine basis.

#### 5.2.1 Consistency

Suppose that $\phi(.)$ is $p$ times continuously differentiable on $[0,1]$ for some $p \geq 5$ and that (3) is computed with the series length $\rho(n)$ such that $\rho(n)^{p-1} h_*^3 \to \infty$ as $n \to \infty$. Also, suppose that $F_{\tilde{x},l,v}[-\beta_1 l + \phi(v)]$ and $f_{\tilde{x},v}(l)$, as functions of $l$, are $s$ times differentiable on some open neighborhood of the origin for some $s \geq 4$. Let $\Psi$ be a kernel of order $s$ and $h_*$ a deterministic sequence of real numbers satisfying $nh_*^8/log(n) \to \infty$ as $n \to \infty$. Under these the estimator $\bar{\beta}$ will be consistent for $\tilde{\beta}_0 \equiv \frac{\tilde{\beta}}{\beta_1}$ provided some mild technical conditions hold.

#### 5.2.2 Asymptotic Normality

Suppose that $\Xi \equiv \int |\Psi|^2 E[\tau(V)\tilde{X}\tilde{X}' f_{\tilde{X},V}(0)]$ exists. Also, assume that the researcher selects $h_*$ to meet $h_*/hh_q \to \infty$ as $n \to \infty$ and $nh_*^{2s+1} \to 0$ as $n \to \infty$. Some further mild conditions and a certain stochastic equicontinuity condition suffice then to establish:

$$\sqrt{nh_*}(\bar{\beta} - \tilde{\beta}_0) \to_d \mathcal{N}(0, Q^{-1}\Xi Q^{-1}).$$

Define the following matrix:

$$\hat{\Xi} \equiv \frac{1}{nh_*}\sum_{i=1}^{n} \tau(\hat{V}_i)\tilde{X}_i\tilde{X}_i{}' |\Psi(\frac{C_i + \tilde{X}_i{}'\tilde{\beta}(v) + b_n' p_n(\hat{V}_i)}{h_*})|^2.$$

Under the assumptions yielding asymptotic normality,

$$H_n[\tilde{\beta}(v)] \longrightarrow_p Q \text{ and } \hat{\Xi} \longrightarrow_p \Xi.$$

Thus inferences can be carried out in practice from data using the asymptotic approximation:

$$\sqrt{nh_*}(\bar{\beta} - \tilde{\beta}_0) \sim \mathcal{N}(0, H_n[\tilde{\beta}(v)]^{-1}\hat{\Xi} H_n[\tilde{\beta}(v)]^{-1}).$$

#### Remarks

($e$) The SASMS estimator achieves a faster rate of convergence than the KWSMS estimator. To be more specific, the SASMS estimator's rate of convergence is $(\frac{hh_q}{h_*})^{1/2}$ times that achieved on the KWSMS estimator which is faster since the bandwidths are selected to meet $\lim \frac{hh_q}{h_*} = 0$ as $n \to \infty$.

($f$) It is important to bear in mind that the SASMS estimator exists only with probability approaching one as $n \to \infty$ since the matrix $H_n[\tilde{\beta}(v)]$ has an inverse only with probability approaching one. In finite sample, the SASMS estimator may exhibit a large variance because of the instability of the inverse in question which may be singular with strictly positive probability. In practice, this poses the same problem as that induced by collinearity where a small change in data produces a substantial variation in estimates. When the kernel $\Psi$ has the form $\Psi(t) = P(t)1[|t| \leq 1]$ for some finite degree polynomial $P$ (see Müller 1984), one way to mitigate this problem is to compute $H_n[\tilde{\beta}(v)]$ by replacing $\Psi^{(1)}(t)$ with $\Psi_c^{(1)}(t) = P^{(1)}(t)1[|t| \leq 1 + c_n]$, where $c_n$ is a deterministic sequence of positive real numbers satisfying $\frac{c_n}{h_*} \to 0$ as $n \to \infty$. This regularized version for the SASMS estimator has the same limiting distribution under the assumptions yielding asymptotic normality.

## 6. Monte Carlo Simulations

This section examines the finite sample properties of the estimators put forth in this paper using Monte Carlo experiments. These estimators are used to estimate the parameter $\beta = 1$ when the data generating process obeys:

$$Y = 1 \; if \; Z + \beta A + \varepsilon \geq 0 \text{ and } Y = 0 \text{ otherwise},$$

$$A = \Pi W + V,$$

$$\varepsilon = \phi(V) + e,$$

where $(Z, W)$ is a standard bivariate Normal couple of correlation coefficient $\varrho$ , $V \sim \mathcal{N}(0,1)$, and $\Pi$ is set equal to 1. In this experiment three designs are considered satisfying the following:

**Design ST**: $\varrho = 0.5$; $\phi(V) = exp(-V^2)$; $e = (1 + Z^2 + Z^4)T$ where $T$ is Student with 3 degrees of freedom.

**Design PR**: $\varrho = 0.5$; $\phi(V) = 0.5V$; $e \sim \mathcal{N}(0,1)$.

**Design LG**: $\varrho = 0$; $\phi(V) = cos(\pi V)$; $e \sim$ *Logistic.*

In addition, two other estimators addressing endogeneity for the binary choice model are used. The first one is the limited information ML estimator[3] (LIML) proposed in Rivers and Vuong (1988) and the second is the artificial two stage least square estimator[4] (2SLS) suggested in Lewbel (2000). Design ST has a non-linear control function with an heteroscedastic error term. Design PR has a linear control function with a normally distributed (conditional on $V$) error term, which satisfies the parametric theory laid out in Rivers and Vuong (1988). Design LG has $Z$ and $W$ independent which makes $Z$ a special regressor as defined in Lewbel (2000).

In all designs the variable $e$ is normalized to have a 0.5 standard deviation. A simulation for a sample size $n = 250, 500$ and 1000 consists of 1000 replications for all estimators but the SASMS estimator. For the latter, experiments with $n = 1000$ are not performed and 500 replications are completed due to the long computational time required. The simulations are conducted in Gauss.

For the KWSMS estimator the smoothing of the indicator function is carried out using:

$$D(t) = [0.5 + \tfrac{105}{64}(t - \tfrac{5}{3}t^3 + \tfrac{7}{5}t^5 - \tfrac{3}{7}t^7)]1[|t| \leq 1] + 1[t > 1].$$

The derivative of $D(.)$ (almost everywhere) is a kernel of order $r = 4$ (Müller 1984). Also, the weighting of the objective is performed using:

$$k(t) = \tfrac{1}{48}(105 - 105t^2 + 21t^4 - t^6)\tfrac{1}{\sqrt{2\pi}}exp(-\tfrac{1}{2}t^2),$$

providing a kernel of order $m = 7$ (Pagan and Ullah 1999). The first stage estimation of the nuisance parameter $\Pi$ is conducted via least squares. The local choice $\bar{v} = 0$ is selected. The bandwidths conditions explained in (3) are only qualitative. Since the optimal bandwidths' selection is not covered in this article, a simple Silverman's like rule of thumb (see Silverman 1986) is adopted. This consists of using $h = \hat{\sigma}_l n^{-3/16}$ and $h_q = \hat{\sigma}_v n^{-3\eta/16}$ where $\eta = 1/3$, $\hat{\sigma}_v$ is to the sample standard deviation of $\{\hat{V}_i\}_{i=1..n}$ and $\hat{\sigma}_l$ is the sample standard deviation of $\{C_i + X_i'\tilde{\theta}\}_{i=1..n}$ with $\tilde{\theta}$ a KWSMS estimator retrieved in a first stage using $(h, h_q) = (n^{-3/16}, n^{-3\eta/16})$. This plug-in method is of course arbitrary in that it depends on the bandwidths

[3]Under the assumptions of Rivers and Vuong (1988) the coefficients are identified up to a different scaling factor. In our context, the LIML refers thus to the ratio between the LIML estimator of A's slope coefficient and Z's slope coefficient since this is how a researcher would estimate our coefficient of interest.

[4]One choice left to the researcher for computing this estimator is the kernel which is needed for estimating the density of $Z$ given $W$, see Lewbel (2000). The Monte Carlo experiments are performed with a normal kernel along with the bandwidths $n^{-1/6}$.

selected originally. Even though this choice for the bandwidths does not *a priori* satisfy any optimal criteria in the context of our specific problem, it has the benefit of being easy to implement while performing reasonably well compared to other choices used in preliminary experiments. The covariance matrix estimator described in Section 4.2.2 relies on $\gamma_1 = 3/8$ and $\gamma_2 = 1$. Other choices for $(\gamma_1, \gamma_2)$ meeting the restrictions of Section 4.2.2 were employed in a preliminary study but this did not materially alter the quality of the sizes.

Finally, the KWSMS estimator is computed by maximizing the objective with the quadratic hill climbing procedure (Goldfeld, Quandt and Trotter 1966). A search for the global maximum consists of selecting out of 10 iterative searches, the local maximum maximizing the objective[5] as there is no guaranty in a finite sample that the local maximum is unique.

For the SASMS estimator, the first stage uses $n$ locals KWSMS estimators which are retried as above but for the value $\bar{v}$. The pseudo least squares $b_n$ is then computed as described in (3) using the trigonometric cosine basis. The sieves' dimensionality sequence $\rho(n) \propto n^{1/11}$ meets the assumptions for the SASMSE. The optimal choice for $\rho(n)$ is beyond the scope of this paper. Here we have the advantage of knowing that the smoothness of the functions involved in all designs is very large so we simply use $\rho(n) = 2[n^{1/11}]$, which amounts to using the first three elements of the trigonometric cosine basis for our displayed simulations.

The SASMS estimator is then computed in the second stage as described in Section 5.1 using a KWSMS estimator with $v = 1/n$ and the following:

$$\Psi(t) = \tfrac{315}{2048}(15 - 140t^2 + 378t^4 - 396t^6 + 143t^8)1[|t| \leq 1],$$

which is a kernel of order 6 (Müller 1984) meeting the conditions of Section 5.2. The kernel bandwidths $h_* = \hat{\sigma}_L n^{-1/10}$ is chosen where $\hat{\sigma}_L$ refers to the sample standard deviation of $\{C_i + \tilde{X}_i{}' \tilde{\beta}(v) + b_n' p_n(\hat{V}_i)\}_{i=1\ldots n}$. **Table 1** contains loss measures enabling to assess the quality of the estimators $\hat{\beta}$ of $\beta$. The Bias refers to absolute value of the bias, i.e.$|E(\hat{\beta}) - \beta|$. The $RM$ refers to the root mean squared error, i.e. $\sqrt{E|\hat{\beta} - \beta|^2}$. **Table 2** provides the sizes of the t-test for $\beta$ relying on the asymptotic covariance estimator given in Section 4.2.2 using the asymptotic critical values for a 1 percent, 5 percent and 10 percent type I error level. As displayed on **Table 1**, the qualitative behaviors of the proposed estimators agree with the asymptotic theory developed in this paper. For all designs the bias and RM of the KWSMS estimator (hereafter noted KWSMSE) consistently shrink as $n$ increases. The same applies to the SASMS estimator (hereafter noted SASMSE). For the KWSMSE, on average across designs, a doubling of the sample size from 500 observations leads to a nearly 30 percent decrease in the loss measures (i.e. bias and RM) which is slightly faster than a 24 percent decrease hinted by asymptotic theory.[6] The SASMSE performs poorly when $n = 250$ relative to the KWSMSE expect for the PR design where a lower RM is achieved. As suggested by asymptotic theory the performance gap between the SASMSE and KWSMSE narrows for all designs if $n = 500$ where the SASMSE outperforms the KWSMSE (in terms of the RM) except for the LG design. That is, the SASMSE needs a large enough sample to reach its asymptotic regime. As explained in section 5.1 the SASMSE may not even exist in a finite sample. The regularization scheme employed for the SASMSE is one out of many possible means to solve this existence problem at the origin of the larger RM experienced for $n = 250$. Motivated by these simulations and those of **Table 2** (discussed soon) there seems to be a need to develop in future research optimal regularization criteria for the SASMSE.

With respect to the overall competitiveness of the proposed estimators, the ST design clearly favors the KWSMSE (or SASMSE provided $n$ is large enough) for every sample size. In that case, the LIML is inconsistent with a RM twice larger when $n = 1000$. As expected the PR design unambiguously supports the LIML, which shows all its efficiency power. In that instance, the KWSMSE (respectively SASMSE) exhibits a RM approximately 3 times larger for $n = 1000$ (respectively for $n = 500$). Finally, the LG design still favors the LIML (which in not too surprising owing to the fact that the logistic distribution and normal distribution have relatively close shapes). In that logistic design, the second best performing estimator when $n = 250$ is the 2SLS, which is eventually slightly outperformed by the KWSMS for $n \geq 500$.

Table 1: Losses

| n=250 | LIML | 2SLS | KWSMS | SASMS |
|---|---|---|---|---|
| | **Bias—RM** | **Bias—RM** | **Bias—RM** | **Bias—RM** |
| ST | 0.135—0.300 | 0.625—0.638 | 0.081—0.240 | 0.125—0.368 |
| PR | 0.005—0.178 | 0.666—0.676 | 0.256—0.939 | 0.296—0.786 |
| LG | 0.007—0.141 | 0.298—0.318 | 0.127—0.434 | 0.314—1.106 |
| **n=500** | | | | |
| ST | 0.132—0.236 | 0.588—0.596 | 0.044—0.146 | 0.040—0.135 |
| PR | 0.006—0.118 | 0.623—0.630 | 0.115—0.355 | 0.121—0.347 |
| LG | 0.000—0.104 | 0.256—0.270 | 0.040—0.244 | 0.119—0.380 |
| **n=1000** | | | | |
| ST | 0.133—0.184 | 0.554—0.560 | 0.034—0.098 | |
| PR | 0.000—0.082 | 0.580—0.584 | 0.075—0.255 | |
| LG | 0.001—0.070 | 0.227—0.236 | 0.028—0.168 | |

Table 2: Sizes

| n=250 | KWSMS | SASMS |
|---|---|---|
| **Nominal level** | 0.01—0.05—0.10 | 0.01—0.05—0.10 |
| ST | 0.11—0.20—0.27 | 0.03—0.07—0.09 |
| PR | 0.23—0.34—0.42 | 0.10—0.16—0.21 |
| LG | 0.26—0.38—0.45 | 0.09—0.17—0.20 |
| **n=500** | | |
| ST | 0.07—0.12—0.19 | 0.01—0.02—0.06 |
| PR | 0.17—0.26—0.33 | 0.08—0.14—0.18 |
| LG | 0.24—0.36—0.42 | 0.06—0.10—0.13 |
| **n=1000** | | |
| ST | 0.04—0.10—0.16 | |
| PR | 0.13—0.23—0.30 | |
| LG | 0.19—0.29—0.35 | |

As exhibited in **Table 2**, the sizes of the test for the KWSMSE using the asymptotic critical values are systematically above the asymptotic sizes even for a sample of 1000 observations. For instance, the size using the 5 percent critical value ranges from 10 to 29 percent across designs. Hence, one requires a much larger sample for the asymptotic critical values to provide an accurate probability coverage for the t-statistic. The same inferential problem affects the smoothed maximum score estimator (see Horowitz 1992). Even though one cannot yet affirm whether the theory of bootstrapping applies to the KWSMS, the result established in Horowitz (2002) concerning the SMSE does suggest that the critical value of a bootstrapped t-statistics will provide a more reliable coverage in finite sample for the KWSMSE. Alternatively, the SASMSE seems to offer somewhat superior testing capability in terms of sizes, which for $n = 500$ are closer to the ones promised by asymptotic theory. This is notably true for the ST design where the type I error of the null hypothesis is more accurately provided by the asymptotic critical value.

[5] The different starting values are drawn from a uniform distribution of mean $\theta_0' = (1, 1)$ and variance 5.

[6] Proposition 3 suggests that the rate of convergence on the loss is $1/\sqrt{n^{1-a-a\eta}}$ which here implies a 24 percent decrease in losses for a doubling of the sample size. This discrepancy does not undermine our theory because the moments of $\sqrt{nhh_q}(\widetilde{\theta}-\theta_0)$ need not to converge unless strong uniform integrability conditions hold, see Chung page 100-101.

## 7. Application: An Effect of Education on Maternal Pregnancy Cigarettes Smoking?

In this section the estimators described in this article are used to determine whether the mother's education impacts the propensity of smoking while pregnant. According to the Centers for Disease Control and Prevention (2004) "infants born to mothers who smoke during pregnancy weigh less, have a lower birth weight which is a key predictor to infant mortality". Finding statistical evidence as to whether the mother's education affects the smoking decision of a pregnant woman is thus important for policy making purposes notably for designing cost effective programs targeting U.S. women.

The source of the dataset is the 1988 National Health Interview Survey. This contains a cross section of 1155 pregnant women in the United Sates. The variables are defined in Table 3. Define $Y = 1$ if the pregnant woman smokes cigarettes and $Y = 0$ otherwise. The decision of whether to engage in smoking is modeled according to the following:

$$Y = \mathbf{1}[\beta_0 + \beta_1 linc + \beta_2 mothereduc + \beta_3 white + \beta_4 cigtax + \epsilon \geq 0],$$

where $\epsilon$ contains unobservable factors influencing the smoking decision process of a pregnant woman. In this application the suspected endogenous variable is the income of the household with a reduced form given by:

$$linc = w'\pi + v,$$

where $w' \equiv (1, mothereduc, white, fathereduc)$, $\pi$ is an unknown parameter while $v$ includes unobservable drivers of the family's income. These unobservable attributes comprise the household's age, the household's work experience and possibly other qualitative traits such as the household's level of self restraint. Given that some of those unobservable factors are probably redundant in $\epsilon$, estimating the parameter $\beta' \equiv (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ without taking into account this link using classic estimation techniques may lead to misleading estimates and invalid testing. As exhibited in Table 4 the estimate $\hat{\pi}$ of $\pi$ via least squares suggests that $w$ is a strong instrument in that $\hat{\pi}$ provides null p-values for the hypothesis (componentwise) $Ho : \pi = 0$. This result is comforting since a prerequisite for the estimation techniques elaborated in this article is the existence of a father's educational effect on *linc* by the identification assumption (see Section 4.1).

The KWSMSE is computed using *linc* as the fully supported variable while the kernels, bandwidths and tuning parameters are chosen as described in Section 6. As explained in Section 4.2.2, an appropriate value for $\bar{v}$ is such that the density of $V|\dot{X}$ is sufficiently differentiable on some neighborhood of $\bar{v}$. Writing $\dot{X}_n$ as the sample mean of $\dot{X}$ and $\hat{\sigma}_v$ the empirical standard deviation of $\{\hat{v}_i\}_{i=1}^n$, a practical rule of thumb consists of selecting some $\bar{v} \in (-2\hat{\sigma}_v, 2\hat{\sigma}_v)$ where the density of $V|\dot{X}_n$ is smooth. Here, $(-2\hat{\sigma}_v, 2\hat{\sigma}_v) = (-1.2, 1.2)$ and nonparametric estimators for the density in question[7] exhibit a few spikes in the range $[-0.5, 1]$. Thus, the conservative choice $\bar{v} = -0.8$ is selected. The major computational difference compared to Section 6 pertains to the maximization of the objective for the KWSMSE which is here conducted employing a simulated annealing (SAN) procedure similar to that used in Horowitz (1992). The SAN is performed with a budget of 500 iterations, providing a starting value relatively close to the global maximizer. Having such a direct optimization algorithm is important as one does not *a priori* know the region of the parameter space which should be emphasized upon because of the unknown scaling coefficient (the slope coefficient of *linc* here). Then, the Climbing Hill algorithm using this starting value converges in less than 30 steps to the global maximum. The SASMSE is computed with kernels, bandwidths as described in Section 6 and the sieves basis truncated with $\rho = 4, 8$. Since the trigonometric cosine basis is chosen, the residuals are normalized by using $F(\hat{v}_i)$ in lieu of $\hat{v}_i$ to compute the SASMSE where $F(.)$ indicates the cumulative distribution function of the standard normal random variable. Finally, the trimming term $\tau(.) \equiv \mathbf{1}[|.| \leq 2\hat{\sigma}_v]$ is used to avoid having the KWSMSE unduly influenced by boundary observations.

[7]Using either the Parzen kernel or the Epanechnikov kernel

Table 3: Variables

| obs=1153 | |
|---|---|
| **Variable** | **Meaning** |
| *linc* | log of family's income in thousands of dollars |
| *mothereduc* | mother's years of education |
| *white* | =1 if mother is white |
| *cigtax* | cigarette tax in Home State in dollars per pack |
| *fathereduc* | father's years of education |

Table 4: Reduced Form via OLS

| **obs=1153** | **coefficient** | **t-stat** |
|---|---|---|
| **Variable** | | |
| *mothereduc* | 0.071 | 7.09 |
| *white* | 0.357 | 6.90 |
| *fathereduc* | 0.060 | 6.75 |

Tables 5 and 6 show the results using these estimation techniques, the probit and the LIML. Because of the scaling chosen, $\tilde{\beta_k}$ for $k = 2, 3, 4$ in Table 5 refers to the estimate of $\frac{\beta_k}{|\beta_1|}$. This permits comparison with the parametric estimators (probit and LIML) since those latter rely on a different scaling factor. The statistic $t_k$ for $k = 2, 3, 4$ in Table 6 refers to the t-statistic for the null $Ho : \beta_k = 0$. Under their assumptions, each of the four estimation procedures conclude that $t_k$ is asymptotically distributed as a standard normal variable under $Ho$.

The probit model provides a negative estimate for *mothereduc* which is significant at conventional confidence levels. In sum, the probit model leads to the conclusion that, everything else held constant, an increase in the mother's education reduces the propensity of pregnancy smoking. The LIML yields also a negative estimate for *mothereduc* albeit smaller in absolute value, suggesting that the benefit of education in reducing pregnancy smoking is less pronounced. However, according to the LIML model, *mothereduc* is not significant at conventional confidence levels. In sum, according to the LIML model the claim that, everything else held constant, an increase in the mother's education reduces her smoking propensity is more uncertain. As shown in Rivers and Vuong (1988), a test of exogeneity for *linc* consists of testing the significance of the reduced form residual $\hat{v}$ in the probit regression of $Y$ on the variables and $\hat{v}$. Under the exogeneity hypothesis $Ho : E[\epsilon v] = 0$ the t-statistic for $\hat{v}$ is $\mathcal{N}(0, 1)$ asymptotically. The t-statistic in question is equal to 1.82, which leads to the rejection of the exogeneity hypothesis at a 10 percent significance level. Provided the parametric assumption of the Rivers and Vuong's estimation method holds[8], this last finding hints that the endogeneity of income is to be taken seriously.

The KWSMSE offers estimates whose signs are the same as those furnished by the LIML. Yet, the results are somewhat contrasting in that the estimates for *mothereduc* is 40 percent larger is magnitude, 50 percent larger for *white* and 50 percent smaller for *cigtax*. The main difference in terms of testing between the KWSMSE and the LIML concerns the prime variable of interest *mothereduc*. Unlike the LIML, the KWSMSE leads to the conclusion that *mothereduc* is significant at conventional levels of significance. The testing of the key median restriction (1) needed for the KWSMSE was conducted[9] as explained in Section 4.3 resulting in $T_n = -0.694$. Therefore, at conventional confidence levels the median restriction assumed in (1) cannot be rejected. The SASMSE provides estimates relatively close to the ones furnished by the KWSMSE. The choice of the sieves parameter $\rho$ does not affect the testing conclusion. The estimate for *mothereduc* is

[8]This Hausman's type of test of exogeneity proposed in Rivers and Vuong (1988) does not require the joint normality assumption of $\epsilon, v$ (or merely $\epsilon|v$) which is needed for the LIML. However, the validity of this test hinges on the classic probit assumption that $\epsilon|X \sim \mathcal{N}(0, 1)$ where $X$ denotes the explanatory variables.

[9]The test was performed using the density of the standard normal distribution for the kernel $\varphi$ and $\xi = \hat{\sigma}_l \hat{\sigma}_v n^{-\omega}$ with $\omega$ the midpoint of $(sup\{1/10; a(1 + \eta)\}, 1/5)$ where $a$ and $\eta$ are the bandwidths parameters selected to compute the KWSMS.

Table 5: Estimates

| obs=1153 | Probit | LIML | KWSMS | SASMS $\rho = 4$ | SASMS $\rho = 8$ |
|---|---|---|---|---|---|
| **Variable** | | | | | |
| *mothereduc* | -0.905 | -0.091 | -0.126 | -0.121 | -0.132 |
| *white* | 0.978 | 0.587 | 0.857 | 0.680 | 0.893 |
| *cigtax* | 0.065 | 0.103 | 0.053 | 0.057 | 0.052 |

Table 6: Statistics

| obs=1153 | Probit | LIML | KWSMS | SASMS $\rho = 4$ | SASMS $\rho = 8$ |
|---|---|---|---|---|---|
| **Variable** | | | | | |
| *mothereduc* | -7.06 | -1.38 | -8.07 | -20.51 | -13.77 |
| *white* | 1.43 | 2.67 | 9.65 | 4.47 | 5.50 |
| *cigtax* | 1.89 | 1.15 | 10.37 | 11.08 | 7.37 |

still negative and significant suggesting that, everything else constant, education reduces pregnancy smoking.

To conclude, data have revealed from testing that the household income is likely correlated with unobservable characteristics of a pregnant woman. Both the LIML and the new proposed estimators suggest that the benefit of education in reducing pregnancy smoking is less pronounced than hinted by a probit. The LIML estimator also hints that the mother's education is not relevant in affecting the smoking decision during pregnancy. However, both the KWSMSE and the SASMSE suggest that the mother's education does reduce the smoking propensity of a pregnant woman. In sum, not addressing the endogeneity of income leads to exaggerating the importance of education in reducing pregnancy smoking. This is probably due to the fact that there are unobservable environmental characteristics for a pregnant woman which encourage smoking and simultaneously depress income.

## 8. Conclusion

This dissertation has presented a local version of the control function approach for the binary choice model to reach consistency when one of the explanatory variables is endogenous. This dissertation has explained how the objective function of the SMSE can be weighted by means of a kernel taking the control variables' estimates as arguments in order to derive an asymptotically centered normal estimator. Finally, a consistent estimator for the asymptotic covariance matrix has been offered enabling expedient inferences for applied work whenever a large data set is available. An alternative score approximation based smoothed maximum score estimator has also been described combining many first stage estimators to obtain a faster rate of convergence. The Monte Carlo simulations hint that both of these estimators can provide new tools to estimate the coefficients of interest and conduct hypothesis testing in the binary choice model when endogeneity is present without having to impose strong distributional assumptions.

## References


Aitken. B., Hanson.G., Harrison.A., 1997. Spillovers, foreign investment, and export behavior. *Journal of international Economics, Vol 43, 103-132..*

Amemiya.T., 1978. The estimation of a simultaneous equation generalized probit model. *Econometrica, Vol 46, 1193-1205 .*

Amemiya.T., 1985. Advanced Econometrics. *Cambridge, Harvard University Press.*

Andrews.D., 1994. Asymptotic for semi parametric econometrics models via stochastic equicontinuity. *Econometrica, Vol 62, 1.*

Bickel.P., 1982. On adaptive estimation. *Annals of Statistics, Vol 10, 647.*

Blundell.R. and Powell.J., 2004. Endogeneity in Semiparametric Binary Response Models. *Review of Economic Studies, Vol 71, 655-679..*

Burns, N., and S. Kedia., 2004. The Impact of Performance-Based Compensation on Misreporting. *Journal of Financial Economics, Vol 3, 305-360..*

CDC Centers for Disease Control and Prevention, 2004. Smoking During Pregnancy–United States, 1990–2002. *Weekly, Vol 53(39); 911-915..*

Chaudhuri.P., 1991 Nonparametric estimates of regression quantiles and their local Bahadur representation. *Annals of Statistics, Vol 19, 2 .*

Chen.X., 2007. Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics, Vol 6B..*

Chung K., 2001 A course in probability theory. *Academic press, third edition..*

Goldfeld.S. Quandt.R. and Trotte.H., 1966. Maximization by quadratic hill-climbing. *Econometrica, Vol.34, No.3.*

Heckman.J., 1978. Dummy endogenous variables in a Simultaneous equation system. *Econometrica, Vol 46, 931-959.*

Hong.H. and Tamer.E., 2003. Endogeneous binary choice model with median restriction. *Economics Letters, Vol 80, 219-225.*

Horowitz.J., 1992. A smoothed maximum score estimator for the binary response model. *Econometrica, Vol 60, No.3.*

Horowitz.J., 1993. Semiparametric estimation of a work trip mode choice model. *Journal of Econometrics, Vol 58, 49-70.*

Horowitz.J., 2002. Bootstrap critical values for tests based on the smoothed maximum score estimator. *Journal of Econometrics, Vol 111, 141-167.*

Lee.L., 1981. Simultaneous equation models with discrete and censored variables., in: *C. Manski and D. Mc Fadden, eds., Structural analysis of discrete data with economic applications, MIT press, Cambridge, MA.*

Lee.S., 2003. Efficient semi parametric estimation of a partially linear quantile regression model. *Econometric Theory, Vol 19, 1−31.*

Lee.S., 2007. Endogeneity in quantile regression models: A control function approach. *Journal of Econometrics, Vol 141, No.2.*

Lewbel.A., 2000. Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrument variables. *Journal of Econometrics, Vol 97, 145-177.*

Manski.C., 1985. Semi parametric analysis of discrete response, asymptotic properties of the maximum score estimator. *Journal of Econometrics, Vol 27, 313-334.*

Müller.H., 1984. Smooth optimum kernel estimators of regression curves, densities and modes. *Annals of Statistics, Vol 12, 766-774.*

Newey.W., 1985. Semiparametric estimation of limited dependent variables models with endogenous explanatory variables. *Annales de l'Insee, No 59/60, Econométrie non lineaire asymptotique.*

Newey.W., 1987. Efficient estimation of limited dependant variable models with endogenous explanatory variables. *Journal of Econometrics, Vol 36, 230-251.*

Newey.W, Powell.J. and Vella.F., 1999. Non parametric estimation of triangular simultaneous equations models. *Econometrica, Vol 67, 565-603.*

Pagan.A. and Ullah.A., 1999. Non parametric econometrics. *Cambridge University Press.*

Pagan.A. and Vella.F., 1989. Diagnostic Tests for Models Based on Individual Data: A Survey. *Journal of Applied Econometrics, Vol 4(S), pages S29-59.*

Powell.J., 1986. Censored regression quantiles. *Journal of Econometrics, Vol 32, 143-155.*

Rivers.D. and Vuong.Q., 1988. Limited information estimation and exogeneity tests for simultaneous probit models. *Journal of Econometrics, Vol 39, 347-366.*

Rothe.C., 2009. Semiparametric estimation of binary response models with endogenous regressors. *Journal of Econometrics, Vol 153, 51-64.*

Smith.R. and Blundell.R., 1986. An exogeneity test for a simultaneous equation tobit model with an application to Labor supply. *Econometrica, Vol 54, 679-685.*

Stone.C., 1975. Adaptive maximum likelihood estimators of a location parameter. *Annals of Statistics, Vol 3, 267-284.*

Szu.H. and Hartley.R., 1987. Fast simulated annealing. *Physics letters, Vol 122, 3-4.*

**Vita**

Jerome Krief was born and raised in Paris, France. After getting his Baccalaureat degree majoring in science in 1995 with honors, he was admitted in *Mathématiques Supérieures* in preparation for the French *Grandes Ecoles*. His interest in mathematical statistics led him to subsequently apply to Paris IX Dauphine University where he obtained a *Maitrise* in 2000 with a thesis concerned about the pricing of Asian options via Monte Carlo simulations using the theory of Brownian motions. His desire to experience a new continent led him to work in the U.S. teaching high school mathematics in Louisiana in 2004-2005. His enjoying teaching led him to pursue a doctorate in economics at LSU. After his passing the qualifying exams in 2007, Jerome focused on econometrics, the field of economics which uses statistical analysis to test causality between economic variables. In the summer of 2010, Jerome was invited to present his job market paper at the 2010 Conference of The Netherlands Econometric Study Group in Leuven, Belgium. More recently, Jerome was selected to present at the 20th Annual Meetings of the Midwest Econometrics Group in St. Louis, Missouri, U.S.A. The degree of Doctor of Philosophy will be conferred on Jerome at the December 2010 Commencement.