View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

E. J. OURSO COLLEGE OF BUSINESS
Department of Economics

DEPARTMENT OF ECONOMICS WORKING PAPER SERIES

Kernel Weighted Smoothed Maximum Score Estimation for Applied Work

J. M. Krief
Louisiana State University

Working Paper 2011-07
http://bus.lsu.edu/McMillin/Working_Papers/pap11_07.pdf

Department of Economics
Louisiana State University
Baton Rouge, LA 70803-6306
http://www.bus.lsu.edu/economics/

Kernel Weighted Smoothed Maximum Score Estimation for Applied Work

Jerome M. Krief *

November 5, 2010

Abstract

The endogenous binary response model frequently arises in economic applications when a covariate is correlated with the error term in the latent equation due to data limitations. Applied workers generally address endogeneity using the principle of Maximum Likelihood (ML) which imposes stringent parametric assumptions. These ML estimators are inconsistent if the posited parametrization is incorrect which can translate in practice into aberrant results contradicting economic theory. Semiparametric estimators have been developed imposing weaker distributional assumptions. Some semiparametric techniques permit inferences from data but restrict heteroscedasticity which may furnish deceptive results. Other semiparametric techniques can accommodate almost any heteroscedasticity but forbid inferences. This article summarizes two new estimation techniques which allow for inferences under general heteroscedasticity conditions. Some Monte Carlo experiments are conducted highlighting the robust advantage of these estimators. Finally, these estimation techniques are applied to assess the effect of education on maternal pregnancy smoking using the 1988 National Health Interview Survey.

Key words: Smoothed maximum score, Endogenous binary choice model, Control function.

JEL codes: C14,C31,C35.

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, (ε, V) are unobservable errors, Π is a $q \times 1$ unknown parameter and β 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 ε 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

*Louisiana State University, Department of Economics, 2125 CEBA Bldg., Baton Rouge, LA 70803, phone: (225)388-3806, e-mail:jkrief1@lsu.edu. I am indebted to Dr. Joel Horowitz for having encouraged me to pursue this topic. I also thank the participants of the 2010 *Netherlands Econometric Study Group* for their criticisms and comments. Additionally, I am grateful to Dr. Kaj Gittings and Dr. Bulent Unel for their suggestions in preparing this non-technical summary. Lastly, I thank Chris Raschke, Paul Darby and Dr. Carter Hill for their aid in handling data.

article are easily generalizable when A is a vector and Π' 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 ε 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 ε 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 ε 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 ε 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 ε 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 ε 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, ε 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 β 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 β 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 β 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 ε 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 ε to remain unaffected by the covariates. For instance, Newey (1985) provided a consistent asymptotically normally distributed two stage maximum score estimator for β under the requirement that (V, ε) be symmetrically distributed around the origin, conditional on S. Also, Hong and Tamer (2003) proposed a consistent minimum distance estimator for β 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 article is to remedy this inferential problem, offering a consistent estimator of β 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 ϕ 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 ϕ 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 β (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 β_n satisfies $\beta_n - \beta = O_p(n^{-\frac{1}{2}+\kappa})$ for some $\kappa \in (0, 1/8)$ where κ 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. The proofs can be found in a technical appendix provided in the back of this paper.

3. Estimation Strategy

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

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

Loosely speaking, (1) imposes that once V has been fixed at $\overline{v}$, the center of the distribution of ε does not vary with the exogenous variables. The equality in (1) will be met for instance when (Z, W) and (ε, 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:

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

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 Π_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 β is only identifiable up to a positive scale since $d(\eta U) = d(U)$ for any scalar $\eta > 0$. The identification of β 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 β_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 β_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 (λ, δ, π) 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 β up to scale. Clearly, this estimation technique excludes binary choice models where the endogenous variable is discrete .

4.2 Estimation Procedure and Asymptotic Properties for the KWSMSE

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 Π. Under the mild assumptions for M-estimators root n consistency will be attained. The simplest estimator for Π 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 δ. 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 θ of the following objective:

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

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,

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

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

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,ℓ). 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 θ_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,ℓ) 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 ℓ, $F_{X,\ell,\overline{v}}[-\beta_1\ell+\phi(\overline{v})]$, $f_{X,\ell}(\overline{v})$ 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:

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

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

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

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 λ 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 φ is a kernel and ξ a deterministic sequence. Introduce $f(.,.)$ the joint density of (ℓ, 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 (ℓ, V) are twice differentiable on some open neighborhood of $(0,\bar{v})$, φ is a strictly positive kernel of order 2, ξ_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 η 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: $Med(\varepsilon|\dot{X}, \bar{v}) = Med(\varepsilon|\bar{v})$,

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

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

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

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 with 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 $\bar{v}$. 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 ℓ 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 Θ 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 ρ 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 ρ. A differentiable function on $[0,1]$ can be approximated arbitrarily well by some element of $Pol(\rho)$ with ρ 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 ρ. 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 ρ 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 ρ 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 Λ_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 μ_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 β the following:

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

and

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

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).

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.

[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 $||g||_{[0,1]} \equiv \sqrt{\int_{[0,1]}|g(t)|^2 dt}$.

5.2.1 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 Ψ 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 Ψ 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.

(g) The exact selection of the bandwidths for the SASMSE is not covered here owing to the fact that only a generic case for any basis $\{f_j\}_{j\geq 1}$ is treated. However, in application one needs to select an appropriate basis for smooth functions and pick three bandwidth sequences h, h_q and h_* meeting the assumptions explained in this summary plus a few others. The reader may find the exact detail for bandwidths selection in the Corollary *Bandwidths Admissibility For Power Series or Trigonometric Series* located in the technical Appendix.

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 ϱ , $V \sim \mathcal{N}(0,1)$, and Π 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 Π 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 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

[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 $n^{-1/6}$.

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 (γ_1, γ_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 β. 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 β 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 $n \geq 500$.

[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.

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	

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

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.

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 ϵ 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 $w' \equiv (1, mothereduc, white, fathereduc)$, π 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 ϵ, 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.

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

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

As exhibited in Table 4 the estimate $\hat{\pi}$ of π 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.

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

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

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

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 ρ does not affect the testing conclusion. The estimate for *mothereduc* is 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]This Hausman's type of test of exogeneity proposed in Rivers and Vuong (1988) does not require the joint normality assumption of ϵ, 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 φ and $\xi = \hat{\sigma}_l \hat{\sigma}_v n^{-\omega}$ with ω the midpoint of $(sup\{1/10; a(1 + \eta)\}, 1/5)$ where a and η are the bandwidths parameters selected to compute the KWSMS.

Conclusion

This article 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 article has explained how the objective function of the SMSE can be weighted by means of a kernel taking the reduced form's residuals 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 dataset 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.*

Appendix

1 Identification

The identification of β (up to a positive scale) is ensured under the followings:

Assumption 1:
$\tilde{W}$ has one component which is not measurable[10] in Z and whose associated slope coefficient is non null.

Assumption 2:
There exists a partition of $\dot{X}' = (C, \tilde{X}')$ where dim C =1 and such that its corresponding slope coefficient, noted β_1, is strictly positive.

Assumption 3:
(a) There exists some given $\overline{v} \in \mathbb{R}$ and some $\phi(\bar{v}) \in \mathbb{R}$ such that:

$$P[\varepsilon \leq \phi(\bar{v})|Z = z, W = w, V = \bar{v}] = \tfrac{1}{2} \text{ a.e.in } z,w.$$

(b) The distribution function of $\varepsilon|\dot{X} = \dot{x}, V = \overline{v}$ has everywhere positive density with respect to the Lebesgue measure a.e.in $\dot{x}$.

Assumption 4:
(a) The distribution function of $C|\tilde{X} = \tilde{x}$ has everywhere positive density with respect to the Lebesgue measure a.e.in $\tilde{x}$.

(b) The distribution function of $V|\dot{X} = \dot{x}$ is absolutely continuous with respect to the Lebesgue measure a.e.in $\dot{x}$ and its density evaluated at $\overline{v}$ exists a.e.in $\dot{x}$. Furthermore, there exists some real number $M_v < \infty$ such that $0 < f(\overline{v}|\dot{x}) < M_v$ a.e.in $\dot{x}$.

Assumption 5:
$E[XX']$ is positive definite where $X' \equiv (1, \tilde{X}')$.

Comments: Assumption 1 is a rank condition requiring at least one excluded instrument which is not a function of Z having an impact on the endogenous variable (see Lee 2007 and Newey, Powell and Vella 1999). Consider for instance the simple case where Z is a scalar variable and $W = (Z, Z^2)$. Even though Z^2 is not part of $\dot{X}$ assumption 1 fails. More generally, adding functions of the exogenous variables including in (i) to the reduced form equation (ii) is not a viable strategy in the context of our estimation problem. Assumptions 2 demands one variable whose marginal impact on the latent index $\dot{X}'\beta$ is positive. As pointing out earlier merely β_1 non null suffices because our parameter of interest is estimated up to the constant $\frac{1}{|\beta_1|}$ and all of our results can be generalized by adding $\frac{\beta_1}{|\beta_1|} \in \{-1, 1\}$ as an additional unknown parameter. Assumption 3(a) is a classic control function condition except that only a local restriction at some $\overline{v}$ is imposed. Assumption 3(b), introduced similarly to Manski's 1985 assumption 2b, prevents the binary outcome Y from being perfectly predictable by $(\dot{X},\bar{v})$ with some strictly positive probability.[11] Assumption 4 contains classic slack conditions permitting LMDR-identification (see Manski 1985, lemma 2) in the context of our control function approach. This is a prerequisite to identification which requires the existence of a significant (in the sense of having a coefficient non null) variable in $\dot{X}$ that must be fully supported. The additional presence of V in the controlled model imposes that $V|\dot{X}$ be supported on some neighborhood (albeit small) of $\overline{v}$. Finally, assumption 5 prevents identification of an intercept in $\dot{X}$.

Now write $\phi(\overline{v}) \equiv Med(\varepsilon|V = \overline{v})$ and $\theta_0' \equiv \frac{1}{\beta_1}(\phi(\overline{v}), \tilde{\beta}')$ where $\tilde{\beta}$ denotes the slope coefficient associated to $\tilde{X}$.

Proposition 1 (Identification)

Under assumptions 1 through 5,

$$\theta_0 \equiv Argmax_{\theta \in \mathbb{R}^K} E[d(\ell + X'(\theta - \theta_0))g_{X,\ell}(\overline{v})],$$

where $\ell \equiv C + X'\theta_0$, $g_{X,\ell}(\overline{v}) \equiv (1 - 2F_{X,\ell,\overline{v}}[-\beta_1\ell + \phi(\overline{v})])f_{X,\ell}(\overline{v})$, $F_{X,\ell,\overline{v}}[.]$ indicates the cumulative distribution function of $\varepsilon|X, \ell, V = \overline{v}$ and $f_{X,\ell}(\overline{v})$ indicates the density of $V|X, \ell$ evaluated at $\overline{v}$.

[10] A random variable is said to be measurable in Z if it has the form $f(Z)$ for some Borel function f. The function is Borel if for any real number a the preset $f^{-1}(a, \infty)$ is a Borel set. Most functions of Z encountered in applied work are measurable in Z such as powers of Z, intercept, the indicator involving the level of Z and the conditional mean $E[T|Z]$ provided $E|T| < \infty$.

[11] Assumption 3(b) is equivalent to $P[Y = 1|\dot{X} = \dot{x}, V = \bar{v}] \in (0, 1)$ a.e. in $\dot{x}$.

2 Asymptotic Properties of the KWSMS Estimator

Let $\{Y_i, \dot{X}_i\}_{i=1}^n$ be a sequence of observations and let $\hat{\Pi}$ be some given estimator from a first stage estimation inducing $\hat{V}_i \equiv A_i - \hat{\Pi}' W_i$ for $i = 1...n$. Also, let h_q and h be two strictly positive bandwidths sequences, $D(.)$ some given function from the real line into itself and $k(.)$ a kernel. For any $\theta \in \mathbb{R}^K$ define the following objective:

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

Sufficient conditions for weak consistency are given next.

Assumption 6:
$\{Y_i, \dot{X}_i, W_i\}_{i=1}^n$ is an iid sequence from $(Y, \dot{X}, W)$ satisfying $Y = d(\dot{X}'\beta + \varepsilon)$.

Assumption 7:
The support of W is a bounded subset of $\mathbb{R}^q$ with $q \geq 1$.

Assumption 8:
θ_0 is an interior point of $\Theta \subset \mathbb{R}^K$ compact.

Assumption 9:(Define $F_{x,l,v}[.]$ the cumulative distribution function of $\varepsilon|X = x, \ell = l, V = v$ and $f_{x,l}(.)$ the density of $V|X = x, \ell = l$ whenever this later exists. Also define $\Psi(\tilde{x})$ the essential supremum of the density of $C|\tilde{X} = \tilde{x}$ whenever this later exists i.e. $\Psi(\tilde{x}) \equiv \{infM \in \mathbb{R} : f_{\tilde{x}}(c) \leq M, \mu - a.e.c\}$ where μ indicates the Lebesgue measure.)

(a) The function $v \mapsto F_{x,l,v}[-\beta_1 l + \phi(\overline{v})]$ and $v \mapsto f_{x,l}(v)$ belong to $C_\infty^m(\bar{v}, M_1)$ for some $M_1 < \infty$ and some $m \geq 2$ a.e.in x, l.

(b) The density of $C|\tilde{X} = \tilde{x}$ is essentially bounded a.e.in $\tilde{x}$ and the function $\tilde{x} \mapsto \Psi(\tilde{x})$ is bounded on its domain.

Assumption 10:
There exists a given $\hat{\Pi}$ such that $\sqrt{n}(\hat{\Pi} - \Pi) = O_p(1)$.

Assumption 11:
(a)$D : \mathbb{R} \longrightarrow \mathbb{R}$. (b)$D$ is bounded. (c) $lim_{t \to -\infty} D(t) = 0$ and $lim_{t \to \infty} D(t) = 1$. (d) D is differentiable everywhere and its derivative noted K satisfies $||K||_{sup} < \infty$.

Assumption 12:
(a) k belongs to $\mathcal{K}_m$. (b) $\int |k(t)|^2 dt < \infty$. (c) k is differentiable everywhere with $||k^{(1)}||_{sup} < \infty$. (d) $\int |t^j k(t)| dt < \infty$ for $j = 1, 2, ..., m-1$ and for any $\sigma > 0$ and any deterministic sequence $c_n = o(1)$,

$$lim\ c_n^{j-m} \int_{|t| > \sigma/c_n} |t^j k(t)| dt < \infty \text{ as } n \to \infty \text{ for } j = 0, 1, ..., m-1.$$

Assumption 13:
The deterministic sequences of strictly positive real numbers $\{h_q\}_n$ and $\{h\}_n$ satisfy $lim\ h$=$lim\ h_q = 0$ and $lim\ nh_q^4$=$lim\ \frac{nh^2 h_q^2}{log(n)} = \infty$ as $n \to \infty$.

Comments: Assumption 7 is imposed for simplicity. Merely, the first moments of W must exist. The bounded support, introduced for deriving the subsequent asymptotic results, may also be dropped if one is willing to assume extra regularity conditions for the distribution of C conditional on $\tilde{X}$ and W. Assumption 8 is technical identically to assumption 4 in Horowitz (1992) because proposition 1 covers $\mathbb{R}^K$ while consistency is easier to establish for a compact set. Assumption 9(a) will be met for instance when both $F_{\varepsilon|\dot{x},v}$ and $f_{\dot{x}}(v)$ as functions of v are twice continuously differentiable on some open neighborhood of the chosen $\bar{v}$ with some bound on the first and second derivatives (a.e.in $\dot{x}$). Assumption 9(b) is technical but is needed to get a uniform convergence for the empirical moment $\widetilde{S_n}$. Assumption 10 is verified under the mild assumptions for M estimators. Assumption 11 introduces the building block for smoothing the indicator function. As explained in the introduction, an easy manner to construct such a function is by integrating a kernel but for consistency purposes this is not needed. Assumption 12 is for the most part a typical condition which demands to select the order of the kernel $k(.)$ to match the smoothness of the function it will convolute with.

Proposition 2 (KWSMS Consistency)

Under the assumptions of proposition 1 and assumptions 6 through 13,

$$\widetilde{\theta_n} \equiv Argmax_\Theta \widetilde{S_n}(\theta) \text{ is (weakly) consistent for } \theta_0.$$

To derive a normal limiting distribution for the estimator introduce the following conditions:

Assumption 14:(Define $g_{x,l}(\overline{v}) \equiv (1 - 2F_{x,l,\overline{v}}[-\beta_1 l + \phi(\overline{v})]) f_{x,l}(\overline{v})$ where $F_{x,l,\overline{v}}[.]$ indicates the cumulative distribution function of $\varepsilon|X = x, \ell = l, V = \overline{v}$ and $f_x(.)$ the density of $\ell|X = x$ whenever this later exists.). The function $l \mapsto g_{x,l}(\overline{v})$ and $l \mapsto f_x(l)$ belong to $C_\infty^r(M_2)$ for some $M_2 < \infty$ and some $r \geq 2$ a.e.in x.

Assumption 15:

(a)$E||X||^4 < \infty$.

(b)$E[XX'T_X^{(1)}(0)]$ *is positive definite where* $T_X(l) \equiv g_{X,l}(\overline{v})f_X(l)$ *and* $T_X^{(1)}(u) \equiv \frac{\partial T_X}{\partial l}|_{l=u}$.

Assumption 16:

(a)K belongs to $\mathcal{K}_r$ *and is symmetrical.*

(b)$\int |K(t)|^{2+\delta}dt < \infty$ *and* $\int |k(t)|^{2+\delta}dt < \infty$ *for some* $\delta > 0$.

(c)$\int |t||K(t)|^2dt < \infty$,$\int |t||k(t)|^2dt < \infty$ *and* $\int |tK(t)|dt < \infty$.

(d)For any $\sigma > 0$ *and any deterministic sequence* $c_n = o(1)$,

$$lim\ c_n^{-1} \int_{|t|>\sigma/c_n} |K^{(1)}(t)|dt = 0 \text{ as } n \to \infty,$$

and

$$lim\ c_n^{-1} \int_{|t|>\sigma/c_n} |k(t)|^2dt < \infty \text{ as } n \to \infty,$$

(e)K is twice differentiable everywhere, $||K^{(j)}||_{sup} < \infty$ *for* $j = 1, 2$, *and* $\int |K^{(1)}(t)|^2dt < \infty$.

Assumption 17:

$lim\ nh_q^{2m+1}h$=$lim\ nh^{2r+1}h_q$=$lim\ \frac{h}{h_q^3}$=$lim\ \frac{h_q^m}{h} = 0$ *as* $n \to \infty$, *and*

$lim\ \frac{nh_q^4h^4}{log(n)} = \infty$ *as* $n \to \infty$.

Comments: Assumption 14 is the key condition needed to derive the asymptotic result for the KWSMS estimator using the classic Taylor's expansion. The stringency in terms of the domain of smoothness may be construed as demanding. However, this is imposed for simplifying the proofs, a smoothness in a neighborhood of the origin would suffice (see Horowitz 1992, assumption 8 and assumption 9) using a lengthier argument. Assumption 15(b), is needed for deriving an asymptotic theory for the KWSMS estimator similarly to the SMSE (see Horowitz 1992, assumption 11). In fact, under assumption 14 the positive definiteness of such matrix would be implied automatically under the identification conditions if assumption 3(a) is strengthened to $F_{\varepsilon|Z,W,\bar{v}} \equiv F_{\varepsilon|\bar{v}}$ a.s..[12] However, assumption 3(a) does not forbid some degree of heteroscedasticity for ε in which case assumption 15(b) is not ensured by the identification assumptions. Assumption 15(a) is needed for A is necessarily continuously distributed (by assumption 4) and the support of X is not assumed bounded. The existence of the fourth moment permits some control to show the convergence of certain expected values notably the collapse of the limiting bias. Assumption 16(a) is a reflection of assumption 14 since various convolutions involving $K(.)$ need to converge in some senses. Assumptions 16(b) and 16(c) are stability conditions for obtaining asymptotic Normality and are satisfied by many kernels, a clear example of which being polynomials compactly supported kernels which are smooth at boundary points. Finally, assumptions 16(d) and 16(e) are needed for the Hessian to converge in probability to some finite quantity and is related to assumptions 7 of Horowitz (1992), which demands the first two derivatives of $K(.)$ to be well behaved. Finally, assumption 17 dictates the bandwidths' rate which must be selected for the asymptotic to be met with lim $nh_q^{2m+1}h$=lim $nh^{2r+1}h_q$=0 collapsing the asymptotic bias while lim $\frac{h}{h_q^3}$=0 allows the usage of the estimated nuisance $V(A, W)$ via $\hat{\Pi}$ to be asymptotically irrelevant.

Proposition 3 (KWSMS Asymptotic Normality)
Under the assumptions of proposition 2 and assumptions 14 through 17,

$$\sqrt{nhh_q}(\widetilde{\theta_n} - \theta_0) \to_d \mathcal{N}(0, H^{-1}\Sigma H^{-1}),$$

where

$$H \equiv E[XX'T_X^{(1)}(0)],\ \Sigma \equiv \int |k|^2 \int |K|^2 E[XX'\mu_X(0)] \text{ and } \mu_X(\ell) \equiv f_{X,\ell}(\overline{v})f_X(\ell).$$

Comments: So far it is implicitly assumed that both assumptions 13 and 17 are met. However, this imposes some smoothness conditions beyond those assumed in assumptions 9. When $h \propto n^{-a}$ and $h_q \propto n^{-a_q}$ for some strictly positive constants a and a_q, the bandwidths requirement put forth in proposition 3 will hold as long as $a \in (Max\{\frac{1}{1+\eta+2\eta m}; \frac{1}{1+\eta+2r}\}, \frac{1}{4+4\eta})$ and $a_q = \eta a$ for some $\eta \in (\frac{3}{2m-3}, \frac{1}{3})$.[13] Thus, the asymptotic conclusion needs a strengthening to $m \geq 7$ in assumption 9. Under

[12]In that case $E[XX'T_X^{(1)}(0)] = 2\beta_1 E[XX'f_{\bar{v}}[\phi(\bar{v})]\mu_X(0)]$ where $f_{\bar{v}}(.)$ is the density of $\varepsilon|V = \bar{v}$. This matrix is positive definite by assumptions 2,3(b),4(b)and 5.

[13]It is clear that Assumptions 13 and 17 both hold as long as lim $nh_q^{2m+1}h$=lim $nh^{2r+1}h_q$=0 , lim $\frac{h}{h_q^3}$=lim $\frac{h_q^m}{h}$=0 and lim $\frac{nh_q^4h^4}{log(n)} = \infty$. Solving these implied inequalities directly yields the bandwidths spectrum given above.

this last condition and $r \geq 2$, one can therefore obtain a rate on convergence in probability for the KWSMS estimator at least $n^{-3/8}$. Yet, this rate improves when $\lambda = Min\{m, r\}$ augments eventually reaching the parametric rate if λ approaches infinity.

As stressed in the introduction, one of the important practical advantage of the KWSMS estimator for the endogenous binary choice model is its ability to conduct inferences from a large sample of observations. The next proposition offers the consistent estimators for the covariance of the above limiting distribution.

Proposition 4 (KWSMS Inferential Feasibility)
Let

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

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* $\gamma_2 \in (0, 1]$. *Under the assumptions of proposition 3,*

$$\widetilde{H_n} \longrightarrow_p H.$$

Furthermore, if $\int |K(t)|^4 dt < \infty$ *and* $\int |k(t)|^4 dt < \infty$,

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

Comments: The rational behind $\widetilde{\Sigma_n}$ not using the bandwidths on which the KWSMS estimator is based upon is to avoid having to add additional bandwidths constraints on the already substantial list.

3 Accelerating Convergence with a Score Approximation Smoothed Maximum Score Estimator

5.2. **Asymptotic Results**

Assumption S1:

Assumptions 3, 4(b), 9 and 14 hold for all $\bar{v} \in [0, 1]$ *as well as other assumptions of proposition 3.*

Comments: This ensures that the conclusion of proposition 2 and 3 holds using any fixed value of v chosen in $[0, 1]$. The choice of $[0, 1]$ is purely symbolic and can be replaced by any compact set of $\mathbb{R}$ for which the above assumptions hold by means of an appropriate normalization.

Assumption S2:

There exists a sample size N such that for each v in $[0, 1]$ *the sequence* $\{E|\widetilde{\theta}(v) - \theta(v)|^2\}_{n \geq N}$ *is monotone.*

Comments: This is a dominance condition which ensures a uniform rate of convergence (in the outer probability sense) for the KWSMS estimator $\widetilde{\theta}(v)$ over $[0, 1]$. Under assumption S1 it is known that for each v, the sequence of mean squared errors converges to 0. This however requires no oscillations if the sample size is large enough.

Assumption S3:

(a) $\phi(.)$ *is p times continuously differentiable on* $[0, 1]$ *for some* $p \geq 1$. *(b) There exists some finite constant C and some* $\gamma \in (0, 1]$ *such that* $|\phi^{(p)}(v_1) - \phi^{(p)}(v_2)| \leq C|v_1 - v_2|^{\gamma}$ *for all* $(v_1, v_2) \in [0, 1] \times [0, 1]$.

Comments: Condition (a) is explicit with the additional slightly stronger requirement in (b) that the p^{th} derivative of $Med(\varepsilon|V = v)$ be Hölder continuous. Then the nuisance function $\phi(.)$ can be approximated (up to scale) arbitrary well by many linear Sieves methods.

Assumption S4:

$\rho(n)$ *is a given sequence of natural numbers such that* $\rho(n)/n < 1$ *for all n and* $\rho(n) \rightarrow \infty$ *as* $n \rightarrow \infty$.

Comments: Let $||f||_{sup}$ for a real valued function $f : [0,1] \to \mathbb{R}$ denotes the sup norm on $[0,1]$. Under assumption S3 and assumption S4 there exists a known basis of functions $\{f_j\}_{j\geq 1}$ such that its linear span $E_\rho = \{f : [0,1] \to \mathbb{R}, f = \sum_{j=1}^{\rho} a_j f_j, a_j \in \mathbb{R}\}$ can approximate the control function $\phi(.)$ arbitrary well in the sense that $inf_{E_{\rho(n)}}||f - \phi||_{sup} \to 0$ as $n \to \infty$ (see Chen 2007). That is, defining $p_n(.)' = (f_1(.), ..., f_{\rho(n)}(.))$ there exists $B_n' = (b_{0,1}, ..., b_{0,\rho(n)})$ such that $B_n' p_n$ provides a good approximation of the unknown control function on $[0,1]$ for n large enough.

Let Λ_n be the $n \times \rho(n)$ matrix whose i^{th} row is $p_n(i/n)'$. Also, under assumption S3 one can introduce $||p_n||_{sup} \equiv sup_{v\in[0,1]}||p_n(v)||$ and given a naturel number ρ use $L[\rho] \equiv \sum_{j=1}^{\rho} ||f_j^{(1)}||_{sup}$.

Assumption S5:

For n large enough the largest eigenvalue of $\Lambda_n'\Lambda_n/n$ is bounded from above and its smallest eigenvalue is bounded away from 0.

Comments: This can be viewed as a dominance condition which permits the discrepancy between b_n and B_n to be imposed only by the "mistakes" committed by the various KWSMS estimators on the first stage and on the approximation error from truncating the basis up to the first $\rho(n)^{th}$ terms.

Assumption S6:

The distribution function of $C|\tilde{X} = \tilde{x}, V = v$ has everywhere positive density with respect to the Lebesgue measure a.e in $\tilde{x}$,v.

Comments: Let $L \equiv C + \tilde{X}'\frac{\tilde{\beta}}{\beta_1} + \frac{\phi(V)}{\beta_1}$. This assumption permits the existence of the density of $L|\tilde{X} = \tilde{x}, V = v$ (a.e.$\tilde{x}$,v) which is needed to derive an asymptotic. Define $F_{\tilde{x},l,v}[.]$ the cumulative distribution function of $\varepsilon|X = \tilde{x}, L = l, V = v$ and $f_{\tilde{x},v}(.)$ the density of $L|\tilde{X} = \tilde{x}, V = v$. Also, use 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.

Assumption S7:

The function $l \mapsto F_{\tilde{x},l,v}[-\beta_1 l + \phi(v)]$ and $l \mapsto f_{\tilde{x},v}(l)$ belong $C^s_\infty(0, M)$ for some $M < \infty$ and some $s \geq 4$ a.e.in $\tilde{x}, v$.

Comments: Under this the classic asymptotic is permitted via non parametric convolution arguments to show consistency and normality. Also, assumption S7 along with assumption S1 ensures the existence of $Q \equiv 2E[\tau(V)\tilde{X}\tilde{X}'F^{(1)}_{\tilde{X},0,V}[\phi(V)]f_{\tilde{X},V}(0)]$.

Assumption S8:

Q is negative definite.

Assumption S9:

(a) $K(.)$ belongs to $\mathcal{K}_s$.

(b) $K(.)$ is twice differentiable everywhere and $||K^{(j)}||_{sup} < \infty$, for $j = 1, 2$.

(c) $\int |K(t)|^4 dt < \infty$ and $\int |K^{(1)}(t)|^2 dt < \infty$.

(d) $\int |t^j K(t)|dt < \infty$ for $j = 1, 2, ..., s-1$.

(e) For any $\sigma > 0$ and any deterministic sequence $c_n = o(1)$,

$$lim\ c_n^{-1} \int_{|t|>\sigma/c_n} |K^{(1)}(t)|dt = 0 \text{ as } n \to \infty,$$

and

$$lim\ c_n^{j-s} \int_{|t|>\sigma/c_n} |t^j K(t)|dt < \infty \text{ as } n \to \infty \text{ for } j = 0, 1, ..., s-1.$$

Assumption S10:

$h_* \to 0$ and $\frac{nh_*^8}{log(n)} \to \infty$ as $n \to \infty$.

Assumption S11: (Using $L_n \equiv L[\rho(n)]$)

(a)$nhh_q h_*^6 \to \infty$ as $n \to \infty$.

(b) $L_n = o(\sqrt{n}h_*^3)$.

(c)$||p_n||_{sup} = O(n^{(1-\gamma)/2} h_*^3 hh_q)$ for some strictly positive γ.

(d) $inf_{E_{\rho(n)}}||f - \phi||_{sup}||p_n||_{sup} = o(h_*^3)$.

Proposition 5 (SASMS consistency)

Under assumptions S1 though S11,

$$\bar{\beta} \text{ is (weakly) consistent for } \tilde{\beta}_0 \equiv \frac{\tilde{\beta}}{\beta_1}.$$

Comments: To make the SASMS estimator more appealing than the KWSMS estimator one needs to show its asymptotic normality and construct consistent estimators for its asymptotic covariance. In order to derive the asymptotic normality a few more assumptions are needed. Introduce the followings:

$$\Xi \equiv (\int |K(t)|^2 dt) E[\tau(V)\tilde{X}\tilde{X}' f_{\tilde{X},V}(0)],$$

and,

$$\bar{G} \equiv \frac{1}{nh_*} \sum_{i=1}^{n} \tau(V_i)\alpha_i \tilde{X}_i K(\frac{L_i}{h_*})$$

where $L_i \equiv \frac{1}{\beta_1} Med(U|\dot{X}_i, V_i)$.

Assumption S12:

$h_*/hh_q \to \infty$ *as* $n \to \infty$.

Assumption S13:

$\sqrt{nh_*}(G_n[\tilde{\beta}(v)] - \bar{G}) = o_p(1)$.

Assumption S14:

$nh_*^{2s+1} \to 0$ *as* $n \to \infty$.

Comments: Assumption S12 permits an estimator asymptotically centered. Assumption S13 can be ensured by a stochastic equicontinuity assumption whose sufficient conditions are provided in Andrews (1994). Finally, assumption S14 enables the researcher to collapse the asymptotic bias. Define the following:

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

The key result of section 6 is now provided next.
Proposition 6

Under assumptions S1 though S14,

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

Furthermore,

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

Comments: Proposition 6 implies that the SASMS estimator achieves a faster rate of convergence in probability than the KWSMS estimator while still allowing for hypothesis testing. 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 $\lim \frac{hh_q}{h_*} = 0$ as $n \to \infty$ by assumption S12. It turns out that this is not the most efficient estimator (in the asymptotic sense) under the assumptions of proposition 6. It is not very difficult to show that a more efficient CAN estimator is given by:

$$\bar{\beta_E} \equiv \tilde{\beta}(v) + \hat{\Xi}^{-1} G_n[\tilde{\beta}(v)],$$

which yields,

$$\sqrt{nh_*}(\bar{\beta_E} - \tilde{\beta_0}) \to_d \mathcal{N}(0, \Xi^{-1}).$$

This will be subsequently referred to as the "efficient" SASMS estimator.[14]

It is important to bear in mind that the SASMS estimator (respectively the "efficient" SASMS estimator) exists only with probability approaching one as $n \to \infty$ since the matrix $H_n[\tilde{\beta}(v)]$ defined in section 5.1 (respectively $\hat{\Xi}$ as defined on page 27) has an inverse only with probability approaching one. In finite sample these estimators may thus exhibit a large variance because of the instability of the matrix in question which may be near singular with a strictly positive probability. When the kernel of assumption S9 has the form $K(t) = p(t)1[|t| \leq 1]$ for some finite degree polynomial p (see Muller 1984), one way to mitigate this variability for the SASMS estimator is to compute $H_n[\tilde{\beta}(v)]$ replacing $K^{(1)}(t)$ with $K_c^{(1)}(t) = p^{(1)}(t)1[|t| \leq 1 + c_n]$ where c_n is a deterministic sequence of positive real numbers satisfying $\frac{c_n}{h_*} \to 0$ as $n \to \infty$.[15]

The selection of the bandwidths is not covered in proposition 6 owing to the fact that only a generic case for any basis $\{f_j\}_{j \geq 1}$ is treated. However, in application one needs to select an appropriate basis for smooth functions and pick three bandwidths sequences h, h_q and h_* meeting the assumptions of proposition 6. The next proposition establishes for the power series basis and trigonometric cosine basis how the bandwidths and sieves's sequence $\rho(n)$ may be selected up to a scale. The symbol $[\kappa]$ for a real number κ will refer to the least lower integer of κ.

Corollary (Bandwidths Admissibility For Power series and Trigonometric cosinus)

Suppose that assumption S1 holds with $r > m/3$, assumption S7 holds for some $s \geq 5$ and assumption S3 holds for some $p > 4$. Also, suppose that others assumptions of proposition 6 hold but assumptions S4,S10,S11,S12,S14. When $p_n(v)' = (f_1(v), ..., f_{\rho(n)}(v))$ is chosen from Power series or Trigonometric cosinus then the assumptions of proposition 6 are satisfied under the followings:

(a)$h \propto n^{-a}$ and $h_q \propto n^{-\lambda a}$, for some $a \in (\frac{1}{1+\lambda+2\lambda m}, \frac{1}{10(1+\lambda)})$ and some $\lambda \in (\frac{3}{2m-3}, min\{\frac{9}{2m-9}, 1/3\})$.

(b) $h_* \propto n^{-a*}$, for some $a^* \in (max\{a', \frac{1}{2s+1}\}, min\{\frac{1-4a'}{6}, \frac{p'}{6p'+12}\})$ where $a' = a(1+\lambda)$ and $p' = p - 1$.

(c) $\rho(n) = C_0[n^\nu]$, for some $\nu \in (\frac{3a*}{p-1}, \frac{1-6a*}{4})$ and some $C_0 \in (0, \frac{n}{[n^\nu]})$.

Comments: This corollary is based upon the fact that with power series or trigonometric series on has $||p_n||_{sup} = O(\rho(n))$ and $L_n = O(\rho(n)^2)$ while $inf_{E_{\rho(n)}} ||f - \phi||_{sup} = O(1/\rho(n)^p)$ (see Chen 2007). Some lengthy algebra can show that (a),(b) and (c) are sufficient for the conditions of proposition 6 to hold. However, those are not necessary and assumptions S4,S10,S11,S12,S14 may hold under different set of conditions which can be found by the researcher on a case to case basis.

Proofs

This section provides the proofs of the propositions. Some notations will be used. $||X||$ denotes the Euclidean norm of a vector $X \in \mathbb{R}^p$ where $p \in \mathbb{N}$ and $|||M||| = \sqrt{traceMM'}$ for a real valued Matrix M. For $r > 0$ and $z \in \mathbb{R}^p$ where $p \in \mathbb{N}$ define $B(z, r) = \{x \in \mathbb{R}^p : ||x - z|| < r\}$. The least upper integer of a real number t is noted $int[t]$.

For a given multivariate real value function twice differentiable say $F(\theta)$ the symbol $\nabla F(\theta)$ denotes its gradient and $HF(\theta)$ its hessian evaluated at θ. Also the sequences of real value functions $D_n(t) = D(t/h)$, $K_n(t) = \frac{1}{h}K(t/h)$ and $k_n(t) = \frac{1}{h_q}k(t/h_q)$ are used. However, the notations $k_n(V) = \frac{1}{h_q}K(\frac{V-\overline{v}}{h_q})$ and $k_n(\hat{V}) = \frac{1}{h_q}K(\frac{\hat{V}-\overline{v}}{h_q})$ are employed which should be kept in mind.

Moreover, the objectives,

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

and

$S_n(\theta) = \frac{1}{nh_q}\sum_{i=1}^n \alpha_i D(\frac{C_i + X_i'\theta}{h})k(\frac{V_i - \overline{v}}{h_q})$ are used.

For an arbitrary real number v use:

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

[14]Indeed, this efficient SASMS estimator requires milder assumptions than those imposed in propositions 5-6. Clearly, assumption S8 is not needed but also assumptions S9(b), S9(c),S9(e) can be shown to be stronger than required for deriving consistency and asymptotic normality.

[15]This "regularized" version for the SASMS estimator has the same limiting distribution because $K^{(1)}$ and $K_c^{(1)}$ differ only when $1 \leq |t| \leq 1 + c_n$ under assumption 9, see Lemma 13-14 and proof of proposition 5.

and

$$S_n(\theta, v) = \frac{1}{nh_q}\sum_{i=1}^n \alpha_i D(\frac{C_i + X_i'\theta}{h})k(\frac{V_i - v}{h_q}).$$

The gradient of $\widetilde{S_n}(\theta, v)$ with respect to θ is noted $\nabla\widetilde{S_n}(\theta, v)$ and its Hessian $H\widetilde{S_n}(\theta, v)$. Similarly, $\nabla S_n(\theta, v)$ and $HS_n(\theta, v)$ are used. Write $\theta_0(v)' = \frac{1}{\beta_1}(\phi(v), \tilde{\beta})$ whenever $\phi(v)$ exists. The notation $\lambda_{Min}[A]$and $\lambda_{Max}[A]$ for a symmetric matrix A will refer to the smallest (respectively largest) eigenvalue of A. Define $P^\star$ the outer probability measure i.e. $P^\star(E) = inf\{\sum P(E_i)|E \subseteq \cup E_i, \{E_i\} \subseteq \Im\}$. Given a sequence of random variables X_n (not necessarily $\Im$-measurable) define $plim^\star X_n = 0$ if for any $\delta > 0$ there exists a natural number N such $n \geq N$ implies $P^\star[|X_n| > \delta] < \delta$. When unspecified the term lim is to be understood with respect to $n \to \infty$. Finally, the complement of a set E will be noted E'.

Lemma 1: Under assumptions 2-4,6,9,11-13 and 15
(i) plim $||S_n - ES_n||_{sup\Theta} = 0$. (ii) lim $||ES_n - S||_{sup\Theta}$=0.

proof (i): Let $g_n(\theta) = S_n(\theta) - ES_n(\theta)$. We have:

$$g_n(\theta) = \frac{1}{n}\sum_{i=1}^n \alpha_i D_n(\ell_i + X'\Delta)k_n(V_i) - E[\alpha_i D_n(\ell_i + X'\Delta)k_n(V_i)]$$

Notice that $|\alpha_i D_n(\ell_i + X'\Delta)k_n(V_i) - E[\alpha_i D_n(\ell_i + X'\Delta)k_n(V_i)]| \leq ||D_n||_{sup}||k||_{sup}\frac{1}{h_q}$ where $||D_n||_{sup}||k||_{sup}$ is a constant by assumption 11 and 12. Also using a change of variable provides :

$$E_{X,\ell}|k_n(V_i)|^2 = \frac{1}{h_q}\int |k(t)|^2 f_{X,\ell}(\overline{v} + th_q)dt = O(\frac{1}{h_q})a.s.$$

due to assumption 12 and 9. Using this last finding and the fact the $D(.)$ is a bounded function provides:

$$\text{Var } [\alpha_i D_n(\ell_i + X'\Delta)k_n(V_i)] \leq E|D_n(\ell_i + X'\Delta)|^2|k_n(V_i)|^2 = O(\frac{1}{h_q}).$$

It follows by the Bennett's inequality (1962) that given $\delta > 0$ arbitrary there exists a strictly positive constant $C(\delta)$ such that:

$$P[|g_n(\theta)| > \delta] \leq 2e^{-nh_q C(\delta)} \tag{1}$$

and lim $|g_n(\theta)| = 0$ a.s. follows by the Borel-Cantelli lemma because of assumption 13. Finally, to show that the convergence is uniform consider the standard argument using non overlapping coverings (Horowitz 1992 lemma 7 or Spady and Klein 1993 lemma 1) of our compact set (assumption 8) with subsets of $\mathbb{R}^K$ such that the distance between two points in each subset is strictly less than a positive sequence r_n. Let $C_{k,n}$ for $k = 1, ..., \Gamma_n$ denotes such collection of subsets where the number of coverings Γ_n will depend on the length of the radius r_n. Let $\{\theta_{k,n}\}_{k=1...\Gamma_n}$ be some selected finite grid of points with $\theta_{k,n} \in C_{k,n}$. Noticing first that:

$$||\nabla g_n(\theta)||_{sup\Theta} \leq c_1 \frac{1}{hh_q}(\frac{1}{n}\sum ||X_i|| + c_2)$$

(where c_1 and c_2 are constants by assumption 12 and 13 by $E[||X||]$ existence i.e.assumption 15) and that any θ in some $C_{k,n}$ implies $||\theta - \theta_{k,n}|| < r_n$ yields:

$$||g_n(\theta)||_{sup\Theta} \leq r_n c_1 \frac{1}{hh_q}(\frac{1}{n}\sum ||X_i|| + c_2) + sup_{k=1..\Gamma_n}|g_n(\theta_{k,n})|$$

It then suffices to set the decreasing radius such that $r_n \propto \frac{log(n)}{nhh_q}$ yielding:

$$\text{lim } r_n c_1 \frac{1}{hh_q}(\frac{1}{n}\sum ||X_i|| + c_2) = 0 \text{ a.s.},$$

because $\lim \frac{1}{n}\sum ||X_i|| = E[||X||]$a.s. by Kolmogorov's strong law of large numbers due to our iid assumption and $r_n \frac{1}{hh_q} = o(1)$ by assumption 13. Finally, plim $sup_{k=1..\Gamma_n}|g_n(\theta_{k,n})|$=0 follows since for $\gamma > 0$ arbitrary and using (1) one can bound $P[sup_{k=1..\Gamma_n}|g_n(\theta_{k,n})| > \gamma]$ owing to:

$$P[\bigcup_{k=1..\Gamma_n} |g_n(\theta_{k,n})| > \gamma] \leq \sum_{k=1}^{\Gamma_n} P[|g_n(\theta_{k,n})| > \gamma] \leq 2\Gamma_n e^{-nh_q C(\gamma)}$$

where $\lim \Gamma_n e^{-nh_q C(\gamma)} = 0$ because $\Gamma_n \propto int[(1/r_n)^K]$. Hence, plim$||g_n(\theta)||_{sup\Theta} = 0$ is established.

proof(ii): step1: Our iid assumption and iterated expectation provide:

$$ES_n(\theta) = E[D_n(\ell + X'\Delta)k_n(V)E_{X,\ell,V}(\alpha)]$$

where

$$E_{X,\ell,V}(\alpha) = 1 - 2F_{X,\ell,V}(-\beta_1\ell + \phi(\overline{v}))$$

$F_{X,\ell,V}(.)$ indicating the distribution function of $\varepsilon|X,\ell,V$. Iterating again gives:

$$ES_n(\theta) = E[D_n(\ell + X'\Delta)E_{X,\ell}\{k_n(V)E_{X,\ell,V}(\alpha)\}]$$

where,

$$E_{X,\ell}\{k_n(V)E_{X,\ell,V}(\alpha)\}] = \int g_{X,\ell}(v)k_n(v)dv, \text{ and}$$

$$g_{X,\ell}(v) = [1 - 2F_{X,\ell,v}(-\beta_1\ell + \phi(\overline{v})][f_{X,\ell}(v)].$$

Using a change of variable with $t = \frac{v-\overline{v}}{h_q}$ and assumptions 2 and 4 further provides:

$$E_{X,\ell}\{k_n(V)E_{X,\ell,V}(\alpha)\} = \int g_{X,\ell}(\overline{v} + th_q)k(t)dt \text{ a.s}$$

Also by assumption 9(a), there exists $\sigma > 0$ and a natural number $m \geq 2$ such that on $I_n = \{|t| < \sigma/h_q\}$:

$$g_{x,\ell}(\overline{v} + th_q) = g_{x,\ell}(\overline{v}) + \sum_{j=1}^{m-1} \frac{1}{j!} g_{x,\ell}^{(j)}(\overline{v})(th_q)^j + \frac{1}{m!} g_{x,\ell}^{(m)}(\xi(x,\ell))(th_q)^m \text{ a.e.in } x, \ell,$$

for some $\xi(x,\ell)$ meeting $|\xi(x,\ell) - \overline{v}| < \sigma$ where for $|v - \overline{v}| < \sigma$,

$$g_{x,\ell}^{(j)}(v) = \sum_{k=0}^{j} \frac{1}{k!(j-k)!} j![1 - 2F_{x,\ell,v}(-\beta_1\ell + \phi(\overline{v})]^{(k)}[f_{x,\ell}(v)]^{(j-k)}$$

with

$$[1 - 2F_{x,\ell,v}(-\beta_1\ell + \phi(\overline{v})]^{(j)} = \frac{\partial^j}{\partial^j v} 1 - 2F_{x,\ell,v}(-\beta_1\ell + \phi(\overline{v})) \text{ and } [f_{x,\ell}(v)]^{(j)} = \frac{\partial^j}{\partial^j v} f_{x,\ell}(v) \text{ for } j = 1...m.$$

Simplifying and using assumption 12 offers:

$$\int g_{x,\ell}(\overline{v} + th_q)k(t)dt =$$
$$g_{x,\ell}(\overline{v}) - g_{x,\ell}(\overline{v})\int_{I'_n} k(t)dt - \sum_{j=1}^{m-1} \frac{g_{x,\ell}^{(j)}(\overline{v})}{j!} h_q^j \int_{I'_n} t^j k(t)dt + \frac{h_q^m}{m!}\int_{I_n} g_{x,\ell}^{(m)}(\xi(x,\ell))t^m k(t)dt + \int_{I'_n} g_{x,\ell}(\overline{v} + th_q)k(t)dt \text{ a.e.in } x, \ell.$$

Furthermore, $|g_{x,\ell}(v)| < M_1^*$ for all v, $|g_{x,\ell}^{(j)}(\overline{v})| < M_1^*$ for $j = 1, ..., m-1$ and $|g_{x,\ell}^{(m)}(\xi(x,\ell))| < M_1^*$ a.e in x, ℓ for some finite constant M_1^* by assumption 9(a). It follows that:

$$h_q^{-m}|\int g_{x,\ell}(\overline{v} + th_q)k(t)dt - g_{x,\ell}(\overline{v})| \leq M_1^*\beth_n \text{ a.e in } x, \ell, \tag{1'}$$

where $\beth_n = 2h^{-m}\int_{I'_n}|k(t)|dt + \sum_{j=1}^{m-1}\frac{h_q^{j-m}}{j!}\int_{I'_n}|t^j k(t)|dt + \frac{1}{m!}\int|t^m k(t)|dt$ is a bounded sequence by assumption 12(d).

Consequently,

$$E_{X,\ell}\{k_n(V)E_{X,\ell,V}(\alpha)\}] = g_{X,\ell}(\overline{v}) + R_n \text{ a.s}$$

where $g_{X,\ell}(\overline{v})$ is given as in proposition 1 due to $E_{X,\ell,V=\overline{v}}(\alpha) = 1 - 2F_{X,\ell,\overline{v}}[-\beta_1\ell + \phi(\overline{v})]$ and $|R_n| = O(h_q^m)$ a.s. by our finding in (1')and assumptions 12(d). Furthermore, $|g_{X,\ell}(\overline{v})|$ is bounded almost surely by some real number $M_v < \infty$ (under assumption 4) yielding:

$$|ES_n(\theta) - S(\theta)| \leq M_v E[|D_n(\ell + X'\Delta) - d(\ell + X'\Delta)|] + O(h_q^m).$$

Step 2: Subsequently, it is straightforward to establish lim $D_n = d$ a.e. by assumption 11 (where the convergence may not hold at the origin). It follows (by Horowitz 1992, lemma 4) that given $\varepsilon > 0$ arbitrary there exists some Borel set $\mathfrak{B}$ of Lebesgue measure strictly less than ε where $\lim ||D_n - d||_{sup\mathfrak{B}'} = 0$ holds. Consequently:

$$|ES_n(\theta) - S(\theta)| \leq M_v(||D||_{sup} + 1)P[\ell + X'\Delta \in \mathfrak{B}] + M_v||D_n - d||_{sup\mathfrak{B}'} + O(h_q^m).$$

step 3: Finally, the cumulative distribution function of $\ell|X = x$ is absolutely continuous with respect to the Lebesgue measure a.e.in x by assumption 4(a) with furthermore a density whose essential supremum is bounded by some constant M a.e.in x by assumption 9(b)implying:

$$P[\ell + X'\Delta \in \mathfrak{B}] < M\varepsilon \text{ uniformly over } \Theta,$$

where we used $P[\ell + X'\Delta \in \mathfrak{B}] = EP_X[\ell \in \mathfrak{B} - X'\Delta]$ and the invariance of the Lebesgue measure to translation. Hence:

$$|ES_n(\theta) - S(\theta)| \leq M_v(||D||_{sup} + 1)M\varepsilon + M_v||D_n - d||_{sup\mathfrak{B}'} + O(h_q^m),$$

with $O(h_q^m) = o(1)$ by assumption 13. It follows that for any $\delta > 0$ one can pick $\mathfrak{B}$ to have measure $\varepsilon < \frac{\delta}{3M(||D||_{sup}+1)M_v}$ so there exists a sample size $N(\delta)$ such that $n \geq N(\delta)$ implies $|ES_n(\theta) - S(\theta)| < \delta$ uniformly over Θ concluding (ii). QED

Lemma 2: Let G be some function in $C^2_\infty(M)$ for some finite real number M, K(.) satisfying assumption 16 and h_n some strictly positive sequence converging to 0 as n approaches infinity. Then we have:

$$\lim ||\mu_n(x) - G^{(1)}(x)||_{sup} = 0 \text{ where } \mu_n(x) = \frac{1}{h}\int -K^{(1)}(t)G(x + th)dt$$

proof: Define $E_n = \{t \in \mathbb{R} : |t| \leq \frac{1}{h}\}$ and use the indicator function $1_E(t) = 1$ if t belongs to a real Borel set E. Given an arbitrary real number x we have:

$$\mu_n(x) = I_{n1}(x) + I_{n2}(x)$$

where

$$I_{n1}(x) = \frac{1}{h}\int -K^{(1)}(t)G(x + th)1_{E'_n}(t)dt$$

and

$$I_{n2}(x) = \frac{1}{h}\int -K^{(1)}(t)G(x + th)1_{E_n}(t)dt$$

The first part is easy as:

$$|I_{n1}(x)| \leq ||G||_{sup}\frac{1}{h}\int |K^{(1)}(t)|1_{E'_n}(t)dt$$

This results in lim $|I_{n1}(x)| = 0$ uniformly in x by the tail property of the Kernel $K(.)$ (assumption 16(d)). Furthermore, integrating by part over E_n yields:

$$I_{n2}(x) = I_{n3}(x) + I_{n4}(x)$$

where

$$I_{n3}(x) = -\frac{1}{h}\{K(t)G(x+th)|_{t\in E_n}\} \text{ and } I_{n4}(x) = \int K(t)G^{(1)}(x+th)1_{E_n}(t)dt$$

Moreover, $|I_{n3}(x)| \le 2||G||_{sup}\frac{1}{h}|K(\frac{1}{h})|$ because the Kernel is symmetric by assumption. As $|t||K(t)|$ tends to 0 as t tends to infinity by assumption 16 we obtain lim $|I_{n3}(x)| = 0$ uniformly in x. Consequently we have :

$$\mu_n(x) - G^{(1)}(x) = \int[G^{(1)}(x+th)1_{E_n}(t) - G^{(1)}(x)]K(t)dt + e_n(x)$$

where the function $e_n(x) = I_{n1}(x) + I_{n3}(x)$ meets $||e_n||_{sup} = o(1)$ by our previous findings. Simplifying gives:

$$\begin{aligned}|\mu_n(x) - G^{(1)}(x)| &\le \int |G^{(1)}(x)||K(t)|1_{E_n'}dt + \int |G^{(1)}(x+th) - G^{(1)}(x)|K(t)|dt + |e_n(x)| \\ &\le ||G^{(1)}||_{sup}\int |K(t)|1_{E_n'}dt + hL\int |t||K(t)dt + ||e_n||_{sup}\end{aligned}$$

where L is a constant as the derivative of $G(.)$ is Liptchitz due to G belonging to $C^2_\infty(M)$. Finally, using lim $\int |K(t)|1_{E_n'}dt$=0 (by the Lebesgue's Dominated Convergence Theorem) and $\int |t||K(t)|dt < \infty$ by assumption 16 finishes the proof. QED

Lemma 3: Under assumptions 1-4,6,8,9,12-17

$$\textit{plim } HS_n(\theta) = -E[XX'T_X^{(1)}(-X'\Delta)] \textit{ where } \Delta = \theta - \theta_0 \textit{ uniformly over } \Theta.$$

proof: We will first use $HS_n^*(\theta) = \frac{1}{n}\sum_{i=1}^n \alpha_i X_i X_i' 1_{\{|X_iX_i'|\le a_n\}}K_n^{(1)}(\ell_i + X_i'\Delta)k_n(V_i)$ (where $|.|$ here is to be understood component wise) where $a_n \propto h^{-2}log(n)$. We will start showing the uniform consistency of HS_n^* since it is easier to establish. Then, we will have left to show plim $HS_n(\theta) - HS_n^*(\theta) = 0$ uniformly over Θ. The notation $H(\theta) = -E[XX'T_X^{(1)}(-X'\Delta)]$ is adopted.

step1: Let 's show plim $HS_n^*(\theta) - EHS_n^*(\theta) = 0$ uniformly over Θ. We have:
$HS_n^*(\theta) - EHS_n^*(\theta) =$

$$\frac{1}{n}\sum_{i=1}^n \alpha_i X_i X_i' 1_{\{|X_iX_i'|\le a_n\}}K_n^{(1)}(\ell_i + X_i'\Delta)k_n(V_i) - E[\alpha_i X_i X_i' 1_{\{|X_iX_i'|\le a_n\}}K_n^{(1)}(\ell_i + X_i'\Delta)k_n(V_i)].$$

By assumption 12 and 16 we get :

$$|\alpha_i X_i X_i' 1_{\{|X_iX_i'|\le a_n\}}K_n^{(1)}(\ell_i + X_i'\Delta)k_n(V_i) - E[\alpha_i X_i X_i' 1_{\{|X_iX_i'|\le a_n\}}K_n^{(1)}(\ell_i + X_i'\Delta)k_n(V_i)]| = O(\frac{a_n}{h^2 h_q}) \tag{2}$$

Also, by assumption 12 and 9 and a change of variable it is rapid to find $E_{X,\ell}[|k_n(V)|^2] \le \frac{C_1}{h_q}$ a.s. for some finite constant C_1 and similarly by Assumption 14 and 16 that $E_X[|K_n^{(1)}(\ell_i + X_i'\Delta)|^2] \le \frac{C_2}{h^3}$ a.s. for some finite constant C_2. Hence, by iterated expectations first with respect to X, ℓ and then with respect to X we obtain the inequality:

$$E[|X_iX_i'|^2 1_{\{|X_iX_i'|\le a_n\}}|K_n^{(1)}(\ell_i + X_i'\Delta)|^2|k_n(V_i)|^2] \le \frac{C_1C_2}{h^3h_q}E[|X_iX_i'|^2]$$

and because $E[|X_iX_i'|^2]$ exists by assumption 15:

$$Var[\alpha_i X_i X_i' 1_{\{|X_iX_i'|\le a_n\}}K_n^{(1)}(\ell_i + X_i'\Delta)k_n(V_i)] = O(\frac{1}{h^3h_q}) \tag{3}$$

Combining (2) and (3) suffices for applying again the Bennett's inequality implying that for any $\delta > 0$ arbitrary real number there exist two strictly positive constants ν_1 and ν_2 such that:

$$P[|HS_n^*(\theta) - EHS_n^*(\theta)| > \delta] \leq 2e^{\frac{-nh^3 h_q \delta^2}{\nu_1 + \nu_2 a_n h}}$$

and plim $|HS_n^*(\theta) - EHS_n^*(\theta)| = 0$ follows since lim $\frac{nh^4 h_q}{log(n)} = \infty$ by assumption 17. Subsequently, we have the bounding:

$$\frac{\partial}{\partial\theta}|\alpha_i X_i X_i' 1_{\{|X_i X_i'| \leq a_n\}} K_n^{(1)}(\ell_i + X_i'\Delta) k_n(V_i)| \leq O(\frac{a_n}{h^3 h_q})$$

due to $||K^{(2)}||_{sup}||k||_{sup}$ being a finite constant by assumptions 12 and 16. Hence, choosing a non overlapping covering with balls whose side length r_n satisfies $r_n \frac{a_n}{h^3 h_q} = o(1)$ will provide plim $|HS_n^*(\theta) - EHS_n^*(\theta)| = 0$ uniformly over Θ by a similar argument as that used for lemma 1.

step2: Let 's now show lim $EHS_n^*(\theta) - H(\theta) = 0$ uniformly over Θ.
By assumption 6 we obtain:

$$E[HS_n^*(\theta)] = E[XX'1_{\{|XX'| \leq a_n\}} K_n^{(1)}(\ell + X'\Delta) E_{X,\ell}\{k_n(V) E_{X,\ell,V}(\alpha)\}]$$

Invoking assumptions 9 and 12 and employing the same approach as in lemma 1(ii) provides:

$$E_{X,\ell}\{k_n(V) E_{X,\ell,V}(\alpha)\} = g_{X,\ell}(\overline{v}) + R_n a.s. \text{ with } R_n = O(h_q^m)\text{a.s.}$$

it follows that:

$$E[HS_n^*(\theta)] = A_{1,n}(\theta) + A_{2,n}(\theta)$$

where
$A_{1,n}(\theta) = E[XX'1_{\{|XX'| \leq a_n\}} K_n^{(1)}(\ell + X'\Delta) g_{X,\ell}(\overline{v})]$
and
$A_{2,n}(\theta) = E[XX'1_{\{|XX'| \leq a_n\}} K_n^{(1)}(\ell + X'\Delta) R_n]$

First, by assumption 15 and 16 we can use the fact that (where $f_X(.)$ indicates the density of $\ell|X$):

$$E_X\{|K_n^{(1)}(\ell + X'\Delta)| = \int \frac{1}{h}|K^{(1)}(t)| f_X(th - X'\Delta)dt \leq M_2 \int \frac{1}{h}|K^{(1)}(t)|dt \text{ a.s. for some finite constant } M_2,$$

and the existence of $E|XX'|$ (i.e.assumption 15) to derive:

$$|E[XX'1_{\{|XX'| \leq a_n\}} K_n^{(1)}(\ell + X'\Delta) R_n]| = O(\frac{h_q^m}{h}).$$

This proves lim $A_{2,n}(\theta) = 0$ uniformly over Θ since $O(\frac{h_q^m}{h}) = o(1)$ by assumption 17.

Secondly, using $\mu_n(X, \Delta) = E_X\{K_n^{(1)}(\ell + X'\Delta) g_{X,\ell}(\overline{v})\}$ and some simplifications furnishes:

$$A_{1,n}(\theta) - H(\theta) = E[XX'1_{\{|XX'| \leq a_n\}}\{\mu_n(X, \Delta) + T_X^{(1)}(-X'\Delta)\}] + E[XX'1_{\{|XX'| > a_n\}} T_X^{(1)}(-X'\Delta)]$$

But notice that $E_X\{K_n^{(1)}(\ell + X'\Delta) g_{X,\ell}(\overline{v})\} = \frac{1}{h}\int T_X(th - X'\Delta) K^{(1)}(t)dt$ where $T_x(\ell) = g_{x,\ell}(\overline{v}) f_x(\ell)$. Under assumption 14 and assumption 16 the conditions of lemma 2 holds (a.e. in x) yielding:

$$|\frac{1}{h}\int T_X(th - X'\Delta)K^{(1)}dt + T_X^{(1)}(-X'\Delta)| \leq M_2^2 b_n + 2M_2^2 c_n + 4hM_2^2 \text{ a.s for some finite constant } M_2,$$

where b_n and c_n are deterministic sequences vanishing to 0 as n approaches infinity. This last finding along with $E|XX'|$ existence establishes that:

$$\lim\ E[XX'1_{\{|XX'|\leq a_n\}}\{\mu_n(X,\Delta) + T_X^{(1)}(-X'\Delta)\}] = 0 \text{ uniformly over } \Theta.$$

Finally, $|T_X^{(1)}(-X'\Delta)|$ is almost surely bounded by a finite constant (independently of θ) by assumption 14, $E|XX'|$ exists and the sequence a_n meets lim $a_n = \infty$. Thus, the Dominated Convergence Theorem directly yields:

$$\lim\ E[XX'1_{\{|XX'|>a_n\}}T_X^{(1)}(-X'\Delta)] = 0$$

Hence, lim $A_{1,n}(\theta) = H(\theta)$ uniformly over Θ is established and thus lim $EHS_n^*(\theta) - H(\theta) = 0$ uniformly over Θ.

step 3: Using basic inequalities we find:

$$||HS_n(\theta) - HS_n^*(\theta)||_{sup\Theta} \leq \frac{1}{h^2}||K^{(1)}||_{sup}\frac{1}{n}\sum_{i=1}^n |X_iX_i'|1_{\{|X_iX_i'|>a_n\}}|k_n(V_i)|$$

so that:

$$E[||HS_n(\theta) - HS_n^*(\theta)||_{sup\Theta}] \leq \frac{1}{h^2}||K^{(1)}||_{sup}E[|X_iX_i'|1_{\{|X_iX_i'|>a_n\}}|k_n(V_i)|]$$

where $E_{X,\ell}[|k_n(V_i)|] \leq M_1\int|k(t)|dt$ a.s. which is rapid to show by change of variable in the integral along with assumption 9. Consequently:

$$E[||HS_n(\theta) - HS_n^*(\theta)||_{sup\Theta}] \leq M_1(\int|k(t)|dt)\frac{1}{h^2}E[|X_iX_i'|1_{\{|X_iX_i'|>a_n\}}]$$

But $E|X_iX_i'|^2 < \infty$ from assumption 7 so by the Cauchy-Schwartz's inequality we can assert:

$$E[|X_iX_i'|1_{\{|X_iX_i'|>a_n\}}] \leq \{E|X_iX_i'|^2\}^{1/2}\{P[|X_iX_i'| > a_n]\}^{1/2}$$

and by the Tchebychev's inequality:

$$P[|X_iX_i'| > a_n] \leq \frac{E|X_iX_i'|^2}{a_n^2}.$$

Since $\int|k(t)|dt < \infty$ (i.e.assumption 12) and lim $a_nh^2 = \infty$ we have established:

$$\lim\ E[||HS_n(\theta) - HS_n^*(\theta)||_{sup\Theta}] = 0$$

and lemma 3 follows by a triangular inequality using step 1 and step 2. QED

Lemma 4:Under assumptions 9,12,14-16

$$E[\nabla S_n(\theta_0)] = O(h_q^m) + O(h^r).$$

proof: Under the iid sampling (assumptions 6)we obtain:

$$E[\nabla S_n(\theta_0)] = E[\alpha X K_n(\ell)k_n(V)] = E[XK_n(\ell)k_n(V)E_{X,\ell,V}(\alpha)]$$

where

$$E_{X,\ell,V}(\alpha) = 1 - 2F_{X,\ell,V}[-\beta_1\ell + \phi(\overline{v})]$$

and similarly to lemma 1 using assumption 9 and 12 permits to show :

$$E_{X,\ell}\{k_n(V)E_{X,\ell,V}(\alpha)\}] = g_{X,\ell}(\overline{v}) + R_n \text{ a.s}$$

where $|R_n| = O(h_q^m)$ a.s. which henceforth returns:

$$E[\nabla S_n(\theta_0)] = B_{1,n} + B_{2,n}$$

where

$$B_{1,n} = E[XK_n(\ell)g_{X,\ell}(\overline{v})]$$

and

$$B_{2,n} = E[XK_n(\ell)R_n]$$

First notice that $|E[XK_n(\ell)R_n]| \leq O(h_q^m)E[|X||K_n(\ell)|]$ and that $E[|X||K_n(\ell)|] = E[|X|E_X\{|K_n(\ell)|\}]$ is bounded due to:

$$E_X\{|K_n(\ell)|\} = \int f_X(th)|K(t)|dt \leq M_2 \int |K(t)|dt \text{ a.s.},$$

for some finite constant M_2 ($f_X(.)$ indicating the density of $\ell|X$) by assumptions 14 and 16. Hence, $B_{2,n} = O(h_q^m)$ is established. Secondly, we can rewrite $B_{1,n}$ by iterating with respect X yielding:

$$B_{1,n} = E[X\rho_n(X)]$$

where $\rho_n(X) = E_X\{K_n(\ell)g_{X,\ell}(\overline{v})\} = \int T_X(\ell)K_n(\ell)d\ell$ with $T_X(\ell) = g_{X,\ell}(\overline{v})f_X(\ell)$. Since by assumptions 14, $T_x(\ell)$, as a function of ℓ, is $r \geq 2$ times continuously differentiable everywhere with bounded j^{th} derivatives for $j = 1...r$ (a.e.in x) we can use the same approach as in lemma 1 but this time with a change of variable $t = \frac{\ell}{h}$ "Taylorizing" $T_x(th)$ around 0 at order $r-1$ and invoking assumption 15 to find:

$$\rho_n(X) = T_X(0) + R'_n \text{ a.s.},$$

where $T_X(0) = 0$ a.s. since $F_{X,0,\bar{v}}[\phi(\bar{v})] = 1/2$ a.s. by assumption 3. Also $R'_n = O(h^r)$a.s. is straightforward to establish using the existence of some constant M such that $|T_x^{(r)}(\ell)| < M$ a.e.in x (from assumptions 14) and the same bounding principle as given in equation(1) of lemma 1. Because $E|X|$ exists by assumption 15 we have also $B_{1,n} = O(h^r)$ which concludes lemma 4. QED

Lemma 5: Under assumptions 9,11,12,14-17

$$\sqrt{nhh_q}\nabla S_n(\theta_0) \rightarrow_d \mathcal{N}(0, \Sigma)$$

proof: It will be convenient to note $s_{i,n} = \sqrt{hh_q}\alpha_i X_i K_n(\ell_i)k_n(V_i)$ and $u_{i,n} = E[s_{i,n}]$ for i=1...n. The structure of the proof is as follows. First, we will show that $\sqrt{nhh_q}(\nabla S_n(\theta_0) - E[\nabla S_n(\theta_0)]) \to_d \mathcal{N}(0, \Sigma)$. Then, we will prove $\lim \sqrt{nhh_q} E[\nabla S_n(\theta_0] = 0$. We have thus:

$$\sqrt{nhh_q}(\nabla S_n(\theta_0) - E[\nabla S_n(\theta_0)]) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_{i,n} - u_{i,n}$$

step1: We will first show some preliminary results. Let $\delta > 0$ be some arbitrary constant. Notice that under assumptions 9,11,12,16 and using a change of variable in the integral as in lemma 1 we have:

$$E_{X,\ell}|\frac{1}{\sqrt{h_q}} k(\frac{V - \overline{v}}{h_q})|^{2+\delta} = \frac{1}{h_q^{\delta/2}} \int |k(t)|^{2+\delta} f_{X,\ell}(\overline{v} + th_q)dt \leq \frac{M_1}{h_q^{\delta/2}} \int |k(t)|^{2+\delta}dt \text{ a.s. for some constant } M_1 \tag{4}$$

Similarly but using assumptions 14 15 and 16 returns:

$$E_X|\frac{1}{\sqrt{h}} K(\frac{\ell}{h})|^{2+\delta}| \leq \frac{M_2}{h_q^{\delta/2}} \int |K(t)|^{2+\delta}dt \text{ a.s. for some constant } M_2 \tag{5}$$

Letting $L_n = \sum_{i=1}^{n} |\frac{s_{i,n} - u_{i,n}}{\sqrt{n}}|^{2+\delta}$ we obtain under assumption 6:

$$E[L_n] \leq n^{-\delta/2} E|s_{i,n} - u_{i,n}|^{2+\delta} \leq 2^{1+\delta} n^{-\delta/2} E|s_{i,n}|^{2+\delta}$$

where $E|s_{i,n}|^{2+\delta} = E[|X|^{2+\delta}|\frac{1}{\sqrt{h}} K(\frac{\ell}{h})|^{2+\delta}|\frac{1}{\sqrt{h_q}} k(\frac{V - \overline{v}}{h_q})|^{2+\delta}]$. Using (4) and (5) along with assumptions 15(a) and 16(b) ensures that there exists some $\delta > 0$ meeting:

$$E|s_{i,n}|^{2+\delta} = O(\frac{1}{h_q^{\delta/2} h^{\delta/2}})$$

and consequently $\lim E[L_n] = 0$ for some some $\delta > 0$ holds because of our choice for the bandwidths meeting $\lim nh_q h = \infty$ by assumption 17.

step 2: Additionally,

$$E[s_{i,n} s'_{i,n}] = E[XX' \frac{1}{h} |K(\frac{\ell}{h})|^2 P_n(X, \ell)]$$

where

$$P_n(X, \ell) = E_{X,\ell}\{\frac{1}{h_q} |k(\frac{V - \overline{v}}{h})|^2\} = \int \frac{1}{h_q} |k(\frac{v - \overline{v}}{h_q})|^2 f_{X,\ell}(v)dv.$$

Moreover, a change of variable $t = \frac{v - \overline{v}}{h_q}$ and a similar reasoning used to derive $(1')$ of lemma 1 invoking assumption $9(a)$, $12(b)$ and $16(c) - (d)$ provides:

$$P_n(X, \ell) = \int f_{X,\ell}(\overline{v} + th_q)|k(t)|^2 dt = f_{X,\ell}(\overline{v}) \int |k(t)|^2 dt + R_n, \text{ where } R_n = O(h_q) \text{ a.s.}$$

Thus, we obtain:

$$E[s_{i,n} s'_{i,n}] = \int |k|^2 E[XX' \frac{1}{h} |K(\frac{\ell}{h})|^2 f_{X,\ell}(\overline{v})] + E[XX' \frac{1}{h} |K(\frac{\ell}{h})|^2 R_n]$$

Moreover, it is easy to show from assumption 14 and a change of variable that $|E[XX'\frac{1}{h}|K(\frac{\ell}{h})|^2]|$ is bounded by $M_2\int|K|^2E|XX'|<\infty$ from assumptions 15(a)and 16(b). As a result we find :

$$E[s_{i,n}s'_{i,n}]=\int|k|^2E[XX'\frac{1}{h}|K(\frac{\ell}{h})|^2f_{X,\ell}(\overline{v})]+o(1)$$

Lastly, assumption 9 and assumption 16 yield:

$$E_X[\frac{1}{h}|K(\frac{\ell}{h})|^2f_{X,\ell}(\overline{v})]=\int\mu_X(\ell)\frac{1}{h}|K(\frac{\ell}{h})|d\ell$$

where $\mu_x(\ell)=f_{x,\ell}(\overline{v})f_x(\ell)$ is continuous and meets $|\mu_x(.)|<C$ for some finite constant C (a.e. in x) by assumptions 14. Thus, changing the variable into $t=\frac{\ell}{h}$ provides:

$$E_X[\frac{1}{h}|K(\frac{\ell}{h})|^2f_{X,\ell}(\overline{v})]=\int\mu_X(th)|K(t)|^2dt$$

and two consecutive applications of the Dominated Convergence Theorem furnishes:

$$\lim\ E_X[\frac{1}{h}|K(\frac{\ell}{h})|^2f_{X,\ell}(\overline{v})]=\mu_X(0)\int|K(t)|^2dt \text{ a.s.,}$$

and

$$\lim\ E[XX'E_X\{\frac{1}{h}|K(\frac{\ell}{h})|^2f_{X,\ell}(\overline{v})\}]=\int|K|^2E[XX'\mu_X(0)].$$

This subsequently offers:

$$\lim\ E[s_{i,n}s'_{i,n}]=\int|k|^2\int|K|^2E[XX'\mu_X(0)]$$

Notice also that $\lim\ E[(s_{i,n}-u_{i,n})(s_{i,n}-u_{i,n})']$=$\lim\ E[s_{i,n}s'_{i,n}]$ due to $u_{i,n}=\sqrt{hh_q}E[\nabla S_n(\theta_0)]=o(1)$ by lemma 4 and assumptions 14. Hence, using the conclusion of step 1 and step 2 permits to apply the Lyapunov's Central Limit Theorem (Chung p 208) to affirm:

$$\sqrt{nhh_q}(\nabla S_n(\theta_0)-E[\nabla S_n(\theta_0)])\rightarrow_d\mathcal{N}(0,\Sigma)$$

Finally, $\sqrt{nhh_q}E[\nabla S_n(\theta_0])=O(\sqrt{nhh_q}h_q^m)+O(\sqrt{nhh_q}h^r)$ by lemma 4 and $\sqrt{nhh_q}E[\nabla S_n(\theta_0])=o(1)$ follows by assumptions 17. QED

Lemma 6: Under assumptions 6,7 and 10-14

$$plim||\widetilde{S_n}-S_n||_{sup\Theta}=0$$

proof: Using the fact that $D(.)$ is bounded by assumption 11 first let us find :

$$||\widetilde{S_n}-S_n||_{sup\Theta}\leq||D||_{sup}\frac{1}{n}\sum_{i=1}^n|k_n(\hat{V}_i)-k_n(V_i)|$$

and $||k^{(1)}||_{sup}$ is finite by assumption 12(iii) so the mean value theorem further provides:

$$||\widetilde{S_n}-S_n||_{sup\Theta}\leq||D||_{sup}||k^{(1)}||_{sup}\frac{1}{nh_q^2}\sum_{i=1}^n|\hat{V}_i-V_i|$$

finally, using $|\hat{V}_i-V_i|=|W_i'(\hat{\Pi}-\Pi)|$ and noting $C=||D||_{sup}||k^{(1)}||_{sup}$ yields:

$$||\widetilde{S_n} - S_n||_{sup\Theta} \leq C||\hat{\Pi} - \Pi||h_q^{-2}\frac{1}{n}\sum_{i=1}^n ||W_i||$$

where $\frac{1}{n}\sum_{i=1}^n ||W_i|| = O_p(1)$ by assumption 7 and $||\hat{\Pi} - \Pi||h_q^{-2} = O_p(h_q^{-2}n^{-1/2}) = o_p(1)$ by assumption 10 and 13 which shows the claim.

Lemma 7: under assumptions 6,7,10,12,16 and 17

$$plim\ ||H\widetilde{S_n} - HS_n||_{sup\Theta} = 0$$

proof: Since $||K^{(1)}||_{sup}$ is finite by assumption 16 we have:

$$||H\widetilde{S_n} - HS_n||_{sup\Theta} \leq ||K^{(1)}||_{sup}\frac{1}{nh^2}\sum_{i=1}^n |X_iX_i'||k_n(\hat{V}_i) - k_n(V_i)|$$

where $|k_n(\hat{V}_i) - k_n(V_i)| \leq \frac{1}{h_q^2}||k^{(1)}||_{sup}|W_i'(\Pi - \hat{\Pi})|$ by assumption 12. Noting $C = ||K^{(1)}||_{sup}||k^{(1)}||_{sup}$ and simplifying further yields:

$$||H\widetilde{S_n} - HS_n||_{sup\Theta} \leq C\frac{1}{nh^2h_q^2}||\Pi - \hat{\Pi}||\sum_{i=1}^n |X_iX_i'|||W_i||$$

where $\frac{1}{n}\sum_{i=1}^n |X_iX_i'|||W_i|| = O_p(1)$ by assumption 6-7 and 15 and $\frac{1}{h^2h_q^2}||\Pi - \hat{\Pi}|| = O_p(\frac{1}{h^2h_q^2n^{1/2}})$ by assumption 10. Consequently $||H\widetilde{S_n} - HS_n||_{sup\Theta} = o_p(1)$ by assumption 17. QED

Lemma 8: Under assumptions 6,7,10,11,13,14,16 and 17

$$plim\ \sqrt{nhh_q}||\nabla\widetilde{S_n}(\theta_0) - \nabla S_n(\theta_0)|| = 0$$

proof: Using assumption 11 and $|k_n(\hat{V}_i) - k_n(V_i)| \leq \frac{1}{h_q^2}||k^{(1)}||_{sup}|W_i'(\hat{\Pi} - \Pi)|$ easily shows that for some constant C:

$$\sqrt{nhh_q}||\nabla\widetilde{S_n}(\theta_0) - \nabla S_n(\theta_0)|| \leq C\sqrt{nhh_q}\frac{||\Pi - \hat{\Pi}||}{h_q^2}T_n$$

where $T_n = \frac{1}{nh}\sum_{i=1}^n ||X_i|||K(\frac{\ell_i}{h})|$. Now assumptions 13-14-16-17 and a double application of the Dominated Convergence Theorem easily yields $\lim E[T_n] = \int |K|E[||X||f_X(0)]$ (where $f_X(0)$ is the density of $\ell|X$ evaluated at 0). Also, under the iid sampling (assumptions 6), $Var(T_n) \leq \frac{1}{nh^2}E[||X||^2|K(\frac{\ell}{h})|^2]$ and the classic change of variable subsequently offers:

$$Var(T_n) \leq \frac{1}{nh}E[||X||^2\int |K(t)|^2 f_X(th)dt]$$

where again the by Dominated Convergence Theorem applied twice establishes that $E[||X||^2\int |K(t)|^2 f_X(th)dt]$ is bounded for:

$$\lim\ E[||X||^2\int |K(t)|^2 f_X(th)dt] = E[||X||^2 \lim\ \{\int |K(t)|^2 f_X(th)dt\}] = \int |K(t)|^2 E[||X||^2 f_X(0)]$$

Since $\lim\ nh = \infty$ by assumption 17 we conclude that T_n is bounded in probability. Therefore we have:

$$\sqrt{nhh_q}||\nabla\widetilde{S_n}(\theta_0) - \nabla S_n(\theta_0)|| = O_p(\frac{\sqrt{hh_q}}{h_q^2})$$

and the choice of bandwidths from assumption 17 finalizes the proof. QED

Lemma 9: Under assumptions S1 and S2
Let $\theta_n(v)$ in the line segment between $\tilde{\theta}(v)$ and $\theta_0(v)$ for any v $\in [0,1]$. Then there exists $H_0(v)$ negative definite such that:

$$plim^{\star} H\widetilde{S_n}(\theta_n(v), v) \equiv H_0(v) \text{ uniformly over } [0,1]$$

proof: Under assumption S1 we know (from lemma 3) that for all v $\in [0,1]$ and almost every x there exists a bounded function $\Psi_x(.,v)$ such that plim $H\widetilde{S_n}(\theta_n(v), v) \equiv E[XX'\Psi_X(\theta_0(v), v)]$ is negative definite. Let introduce $H(\theta, v) \equiv E[XX'\Psi_X(\theta, v)]$ for any θ and let $\theta_n(v)$ in the line segment between $\tilde{\theta}(v)$ and $\theta_0(v)$. Using 2 consecutive triangular inequalities yields:

$$|H\widetilde{S_n}(\theta_n(v), v) - H(\theta_0(v), v)| \leq |H\widetilde{S_n}(\theta_n(v), v) - HS_n(\theta_n(v), v)| + |HS_n(\theta_n(v), v) - E[HS_n(\theta_n(v), v)]| + |E[HS_n(\theta_n(v), v)] - H(\theta_n(v), v)| + |H(\theta_n(v), v) - H(\theta_0(v), v)|$$

By lemma 7 we obtain $plim^{\star}|H\widetilde{S_n}(\theta_n(v), v) - HS_n(\theta_n(v), v)| = 0$ uniformly over [0,1]. Also, invoking assumption S1 and a similar approach as in lemma 2 (or lemma 3) results in:

$$sup_{(\theta,v)\in\Theta\times[0,1]}|HS_n(\theta, v) - E[HS_n(\theta, v)]| = o_p(1)$$

and

$$\lim sup_{(\theta,v)\in\Theta\times[0,1]}|E[HS_n(\theta, v)] - H(\theta, v)| = 0$$

It therefore follows that,

$plim^{\star}|HS_n(\theta_n(v), v) - E[HS_n(\theta_n(v), v)]| + |E[HS_n(\theta_n(v), v)] - H(\theta_n(v), v)| = 0$ uniformly over $[0,1]$.

Finally, under S1, $sup_{l,v\in[0,1]}|\partial\Psi_x(l, v)\partial l|$ exists and is bounded by some constant constant M (a.e in x). It follows by the mean value theorem along with the Cauchy-Schwartz inequality that:

$$|H(\theta_n(v), v) - H(\theta_0(v), v)| \leq ME[|XX'|^2||X||^2]^{1/2}E[||\theta_n(v) - \theta_0(v)||^2]^{1/2}$$

Since $\theta_n(v)$ in the line segment between $\tilde{\theta}(v)$ and $\theta_0(v)$ we have plim $||\theta_n(v) - \theta_0(v)|| = 0$ under assumption S1 implying lim $E[||\theta_n(v) - \theta_0(v)||^2]^{1/2}$=0 by dominated convergence since both $\theta_n(v)$ and $\theta_0(v)$ lie in a compact set by assumption S1. It follows under assumption S2 that lim $E[||\theta_n(v) - \theta_0(v)||^2]^{1/2}$=0 uniformly over [0,1] by Dini's Theorem establishing $plim^{\star}$ $|H(\theta_n(v), v) - H(\theta_0(v), v)| = 0$ uniformly over [0,1]. QED

Lemma 10: Under assumptions S1, S2 and S3

$$plim^{\star} n^{\frac{1-\gamma}{2}} hh_q sup_{v\in[0,1]}||\Delta_n(v)|| = 0 \text{ for all } \gamma > 0 \text{ where } \Delta_n(v) \equiv \tilde{\theta}(v) - \theta_0(v).$$

proof: We use $\widetilde{g}(v) \equiv \nabla\widetilde{S_n}(\theta_0(v), v)$ as well as $\bar{g}(v) \equiv \nabla S_n[\theta_0(v), v]$. Since $[0,1]$ is compact we can invoke assumption S1 and assumption S3 to show in a similar fashion as in lemma 1-3 that:

$$n^{\frac{1-\gamma}{2}} hh_q sup_{v\in[0,1]}||\bar{g}(v) - E\bar{g}(v)|| = o_p(1) \text{ for all } \gamma > 0.$$

Also, by assumption S1 we have Assumption 9 and 14 holding uniformly for an arbitrary $\bar{v} \in [0,1]$. Thus, by lemma 4 we obtain:

$$sup_{v\in[0,1]}||E\bar{g}(v)|| = O(h_q^m + h^r)$$

Hence, the bandwidths conditions of proposition 3 (i.e. assumption 17) shows that:

$$n^{\frac{1-\gamma}{2}} hh_q sup_{v\in[0,1]}||\bar{g}(v)|| = o_p(1) \text{ for all } \gamma > 0 \tag{6}$$

Additionally, lemma 8 provides:

$$n^{\frac{1-\gamma}{2}} hh_q sup_{v\in[0,1]}||\bar{g}(v) - \widetilde{g}(v)|| = o_p(1) \tag{7}$$

Now with wpa.1 as $n \to \infty$, the mean value theorem gives:

$$-H\widetilde{S_n}(\overline{\theta(v)}, v).\Delta_n(v) \equiv \bar{g}(v) + E_n(v)$$

where plim $\overline{\theta(v)} = \theta_0(v)$ for all v in $[0, 1]$ due to assumption S1. Using the triangular inequality furnishes:

$$|| - H\widetilde{S_n}(\overline{\theta(v)}, v).\Delta_n(v)|| \leq ||\bar{g}(v)|| + ||E_n(v)||$$

since $||\Delta_n(v)||.\ |\lambda_{Min}[-H\widetilde{S_n}(\overline{\theta(v)}, v)]| \leq || - H\widetilde{S_n}(\overline{\theta(v)}, v).\Delta_n(v)||$ by the spectral decomposition of $-H\widetilde{S_n}(\overline{\theta(v)}, v)$ we further obtain:

$$Min_{v\in[0,1]}|\lambda_{Min}[-H\widetilde{S_n}(\overline{\theta(v)}, v)].|sup_{v\in[0,1]}||\Delta_n(v)||$$
$$\leq sup_{v\in[0,1]}||\bar{g}(v)|| + sup_{v\in[0,1]}||E_n(v)||$$

where $plim^\star Min_{v\in[0,1]}|\lambda_{Min}[-H\widetilde{S_n}(\overline{\theta(v)}, v)]|$ is some finite strictly positive constant by lemma 9. This last fact along with (6) and (7) combined yield the result. QED

Lemma 11: Under assumptions S1 through S5
(a) For n large enough there exists $B_n \in \mathbb{R}^{\rho(n)}$ and a given $p_n(.)' = (f_1(.), ..., f_{\rho(n)}(.))$ such that:

$$||b_n - B_n|| = O(||\Delta_n||_{sup}) + O(||R_n||_{sup})$$

where $\Delta_n(v) \equiv \tilde{\theta}(v) - \theta_0(v)$ and $||R_n||_{sup} = inf_{E_{\rho(n)}}||f - \phi||_{sup} \to 0$ as $n \to \infty$.

(b) $||b_n' p_n - \tilde{\phi_0}||_{sup} = O(||p_n||_{sup}||b_n - B_n||) + O(||R_n||_{sup})$

proof(a): From assumption S3 and assumption S4 there exists(see Chen 2007,Timan 1963)$B_n \in \mathbb{R}^{\rho(n)}$ and a basis of function $p_n(.)'$ such that:

$$inf_{E_{\rho(n)}}||f - \phi||_{sup} = ||B_n' p_n - \tilde{\phi_0}||_{sup}$$

where $||B_n' p_n - \tilde{\phi_0}||_{sup} = o(1)$. Let define $\tilde{\phi_{0,n}} \in \mathbb{R}^{\rho(n)}$ the vector whose i^{th} element is $\tilde{\phi_0}(i/n)$ and $\delta_n = \tilde{\phi_n} - \tilde{\phi_{0,n}}$. From (1) we have $\tilde{\phi_{0,n}} = \Lambda_n B_n + r_n$ where $||r_n||/n = O(||R_n||_{sup})$. It is also easy to show that:

$$b_n - B_n \equiv (\Lambda_n'\Lambda_n/n)^{-1}(\Lambda_n'\delta_n/n) + (\Lambda_n'\Lambda_n/n)^{-1}(\Lambda_n' r_n/n)$$

and consequently,

$$||b_n - B_n|| \leq ||(\Lambda_n'\Lambda_n/n)^{-1}(\Lambda_n'\delta_n/n)|| + ||(\Lambda_n'\Lambda_n/n)^{-1}(\Lambda_n' r_n/n)||$$

Now use assumption S5 which permits to use the spectral decomposition of $\Lambda_n'\Lambda_n/n$ yielding:

$$||(\Lambda_n'\Lambda_n/n)^{-1}(\Lambda_n'\delta_n/n)||^2 \leq \frac{\lambda_{max}[\Lambda_n'\Lambda_n/n]}{\lambda_{min}[\Lambda_n'\Lambda_n/n]}||\delta_n||^2/n \text{ for n large}$$

and likewise

$$||(\Lambda_n'\Lambda_n/n)^{-1}(\Lambda_n'r_n/n)||^2 \leq \frac{\lambda_{max}[\Lambda_n'\Lambda_n/n]}{\lambda_{min}[\Lambda_n'\Lambda_n/n]}||r_n||^2/n \text{ for n large}$$

But $||\delta_n||^2/n = O(||\Delta_n||^2_{sup})$, $||r_n||^2/n = O(||R_n||^2_{sup})$ and $\frac{\lambda_{max}[\Lambda_n'\Lambda_n/n]}{\lambda_{min}[\Lambda_n'\Lambda_n/n]}$ is bounded by assumption S5 for n large. QED

proof(b): use the decomposition $\tilde{\phi_0} = B_n'p_n + e_n$ where $e_n(.)$ meets $||e_n||_{sup} = inf_{E_{\rho(n)}}||f-\phi||_{sup}$. Then from $|b_n'p_n(v) - \tilde{\phi_0}(v)| \leq |(b_n - B_n)'p_n(v)| + |e_n(v)|$ use the Cauchy-Schwartz inequality $|(b_n - B_n)'p_n| \leq ||b_n - B_n||.||p_n(v)||$ and take the supremum over $[0,1]$ on both sides. QED

Lemma 12: Under assumptions S1 through S5

$$sup_{i=1...n}\tau(V_i)|b_n'p_n(\hat{V}_i) - \tilde{\phi_0}(V_i)| = O_p(1)O(||\hat{\Pi} - \Pi||.L_n) + O(||p_n||_{sup}||R_n||_{sup})$$

proof: For all $i = 1...n$ we have:

$$\tau(V_i)|b_n'p_n(\hat{V}_i) - \tilde{\phi_0}(V_i)| \leq \tau(V_i)|b_n'p_n(\hat{V}_i) - b_n'p_n(V_i)| + \tau(V_i)|b_n'p_n(V_i) - \tilde{\phi_0}(V_i)|$$

where

$$\tau(V_i)|b_n'p_n(V_i) - \tilde{\phi_0}(V_i)| \leq ||b_n'p_n - \tilde{\phi_0}||_{sup} \text{ (1)}$$

and

$$\tau(V_i)|b_n'p_n(\hat{V}_i) - b_n'p_n(V_i)| \leq ||b_n||.||p_n(\hat{V}_i) - p_n(V_i)|| \text{ (2)}$$

Notice that $||b_n||$ is bounded in probability by lemma 11. Also the mean value theorem for each function f_j comprising p_n relying on $\hat{V}_i - V_i = W_i'(\hat{\Pi} - \Pi)$ and $||W_i|| < C$ a.s. for some constant C by assumption S1 establishes $||p_n(\hat{V}_i) - p_n(V_i)|| \leq CL_n||\hat{\Pi} - \Pi||$. Using this last finding into (2) and the results of lemma 11 into (1) shows the claim. QED

Lemma 13: Under assumptions S6,S7 and S9
(a) $E[\nabla S_*(\tilde{\beta_0}, \tilde{\phi_0})] = O(h_*^s)$ for some natural number $s \geq 2$
(b) Var $[\nabla S_*(\tilde{\beta_0}, \tilde{\phi_0})] = O(\frac{1}{nh_*^2})$

proof(a):Iterating first with respect to $\tilde{X}, L, V$ and then with respect to $\tilde{X}, V$ yields:

$$E[\nabla S_*(\tilde{\beta_0}, \tilde{\phi_0})] = E[\tau(V)\tilde{X}\mu_n(\tilde{X}, V)]$$

where $\mu_n(\tilde{X}, V) = E_{\tilde{X},V}\{(1 - 2F_{\tilde{X},L,V}[-\beta_1 L + \phi(V)])K_n(L)\}$ a.s. and using S6 along with a change of variable yields:

$$\mu_n(\tilde{x}, v) = \int p_{\tilde{x},v}(th_*)K(t)dt \text{ a.e. } \tilde{x}, v$$

where $p_{\tilde{x},v}(l) = 1 - 2F_{\tilde{x}l,v}[-\beta_1 l + \phi(v)]f_{\tilde{x},v}(l)$. Now considers the expression for $\mu_n(x, v)$. Under assumption S7 there exists $\sigma > 0$ such that on $E_n = \{|t| < \sigma/h_*\}$:

$$p_{\tilde{x},v}(th_*) = p_{\tilde{x},v}(0) + \sum_{j=1}^{s-1}\frac{p_{\tilde{x},v}^{(j)}(0)}{j!}(th_*)^j + \frac{p_{\tilde{x},v}^{(s)}(\xi(\tilde{x},v))}{s!}(th_*)^s$$

for some $s \geq 2$ and $|\xi(x, v)| < \sigma$ a.e. $\tilde{x}, v$. Since $p_{\tilde{x},v}(0) = 0$ a.e $\tilde{x}, v$ by assumption S1 we can further obtain:

$$\mu_n(\tilde{x},v)=\mu_{n,1}(\tilde{x},v)+\mu_{n,2}(\tilde{x},v) \text{ a.e.}\tilde{x},v.$$

where

$\mu_{n,1}(\tilde{x},v)=\sum_{j=1}^{s-1}\frac{p_{x,v}^{(j)}(0)}{j!}h_*^j\int_{E_n}t^jK(t)dt+\frac{h_*^s}{s!}\int_{E_n}p_{x,v}^{(s)}(\xi(x,v))t^sK(t)dt$ a.e. $\tilde{x},v.$

and

$\mu_{n,2}(\tilde{x},v)=\int_{E_n'}K(t)p_{\tilde{x},v}(th_*)dt$ a.e.$\tilde{x},v.$

notice that for $j=1,2,...,s-1$:

$$\int_{E_n}t^jK(t)dt=-\int_{E_n'}t^jK(t)dt \text{ by assumption 9(a)}$$

and that there exists a finite M for which:

$$|p_{x,v}^{(j)}(0)|<M \text{ for } j=1,2,...,s-1,\ |p_{\tilde{x},v}(.)|<M \text{ and } |p_{x,v}^{(s)}(\xi(x,v))|<M \text{ a.e. } \tilde{x},v \text{ by assumption S7.}$$

Thus we obtain the following bounding:

$|\mu_{n,1}(\tilde{X},V)|+|\mu_{n,2}((\tilde{X},V)|\leq M(\sum_{j=1}^{s-1}\frac{h_*^j}{j!}\int_{E_n'}|t^jK(t)|dt+\frac{h_*^s}{s!}\int|t^sK(t)|dt+\int_{E_n'}|K(t)|dt)$ a.s.

and subsequently:

$$h_*^{-s}|E[\nabla S_*(\tilde{\beta_0},\tilde{\phi_0})]|\leq h_*^{-s}E[|\tilde{X}\mu_n(\tilde{X},V)|]\leq M\beth_nE|\tilde{X}|,$$

where $\beth_n=\sum_{j=1}^{s-1}\frac{h_*^{j-s}}{j!}\int_{E_n'}|t^jK(t)|dt+\frac{1}{s!}\int|t^sK(t)|dt+h_*^{-s}\int_{E_n'}|K(t)|dt$. But $\beth_n$ is a bounded sequence by assumption S1,S9(a)and S9(e) while $E|\tilde{X}|$ exists by assumption S1. QED

proof(b): This is immediate under the iid sampling assumption since $\frac{1}{n}E[|\tilde{X}|^2|K_n(L)|^2]\leq\frac{||K||_{sup}^2}{nh_*^2}E|\tilde{X}|^2$ where $E|\tilde{X}|^2$ exists by assumption S1 and $||K||_{sup}$ exists by S9(b). QED

Lemma 14: Under assumptions S6,S7 and S9
(a) lim $E[HS_*(\tilde{\beta_0},\tilde{\phi_0})]=Q$ as n $\to\infty$
(b) Var $[HS_*(\tilde{\beta_0},\tilde{\phi_0})]=O(\frac{1}{nh_*^4})$

proof(a): By the same approach as in lemma 12 we get:

$$E[HS_*(\tilde{\beta_0},\tilde{\phi_0})]=E[\tau(V)\tilde{X}\tilde{X}'A_n(\tilde{X},V)]$$

where

$$A_n(\tilde{X},V)=E_{\tilde{X},V}\{(1-2F_{\tilde{X},L,V}[-\beta_1L+\phi(V)])K_n^{(1)}(L)\} \text{ a.s.,}$$

and invoking assumptions S7,S9(e) and a similar argument as in lemma 3 one can easily derive:

$$\lim A_n(\tilde{X},V)=-p_{\tilde{X},V}^{(1)}(0) \text{ a.s.,}$$

where $p_{\tilde{X},V}(.)$ is as defined in lemma 12. The claim follows by Dominated convergence since $E|\tilde{X}\tilde{X}'|$ exists by assumption S1. QED

proof(b): This is immediate using the same bounding principle as in proof(b) of lemma 12 invoking instead the existence of both $||K^{(1)}||_{sup}$ (by assumption S9(b)) and $E[\tilde{X}\tilde{X}'\tilde{X}\tilde{X}']$ (by assumption S1). QED

Proposition 1

proof: under assumption 4(b), $g_{X,\ell}(\overline{v})$ is well defined and $S(\theta)$ exists uniformly over $\mathbb{R}^K$. For any $\theta \in \mathbb{R}^K$ such that $||\Delta|| > 0$ where $\Delta = \theta - \theta_0$ we have:

$$S(\theta_0) - S(\theta) = E[1[|X'\Delta| > 0](d(\ell) - d(\ell + X'\Delta))g_{X,\ell}(\overline{v})]$$

Using iterated expectation yields:

$$S(\theta_0) - S(\theta) = E[1[|X'\Delta| > 0]E_X\{(d(\ell) - d(\ell + X'\Delta))g_{X,\ell}(\overline{v})\}].$$

Using $Med(\varepsilon|X, \ell, \bar{v}) = \phi(\bar{v})$ a.s. by assumption 3 subsequently offers :

$(d(\ell) - d(\ell + x'\Delta))g_{x,\ell}(\overline{v}) = |(1 - 2F_{x,\ell,\overline{v}}[-\beta_1\ell + \phi(\overline{v})]|f_{x,\ell}(\overline{v}) > 0$ a.e.in x whenever $|d(\ell) - d(\ell + x'\Delta)| > 0$,

because of assumption 2, assumption 3(b) and assumption 4(b). Lastly, $f_x(\ell) > 0$ a.e.in x (by assumption 4(a)) implies by Manski's 1985 lemma 2 that:

$$P[|d(\ell) - d(\ell + x'\Delta)| > 0] > 0 \text{ provided } |x'\Delta| > 0.$$

Thus, the random variable $E_X\{(d(\ell) - d(\ell + X'\Delta))g_{X,\ell}(\overline{v})\} > 0$ a.s. on the event $|X'\Delta| > 0$ which has a strictly positive probability by assumption 5 and $S(\theta_0) - S(\theta) > 0$ follows. QED

Proposition 2

proof: By a triangular inequality $||\widetilde{S_n} - S||_{sup\Theta} \leq ||\widetilde{S_n} - S_n||_{sup\Theta} + ||S_n - S||_{sup\Theta}$ where $||\widetilde{S_n} - S_n||_{sup\Theta} = o_p(1)$ by lemma 6 and $||S_n - S||_{sup\Theta} = o_p(1)$ by lemma 1. Hence, $plim||\widetilde{S_n} - S||_{sup\Theta} = 0$ with in addition $S(.)$ continuous everywhere under the assumptions of proposition 2 (see Manski's 1985 lemma 5) and admitting a unique global maximizer at θ_0 by proposition 1. Invoking assumption 8 concludes the proof of Proposition 2 by Theorem 4.1.1 of Amemiya (1985). QED

Proposition 3

proof: By assumption 8 and proposition 2, the estimator $\widetilde{\theta_n}$ is an interior point of Θ with probability approaching 1 as $n \to \infty$. Since $\widetilde{S_n}$ is twice differentiability everywhere (by assumption 16(e))and attains a maximum over Θ at $\widetilde{\theta_n}$ one can use a mean value expansion yielding:

$$0 = \nabla\widetilde{S_n}(\theta_0) + H\widetilde{S_n}(\bar{\theta})(\widetilde{\theta_n} - \theta_0) \text{ wpa.1},$$

for some $\bar{\theta}$ in the line segment joining $\widetilde{\theta_n}$ and θ_0 which may vary from row to row. Also, combining lemma 3 and lemma 7 furnishes:

$$H\widetilde{S_n}(\bar{\theta}) = H(\bar{\theta}) + o_p(1),$$

where $H(\theta) = -E[XX'T_X^{(1)}(-X'(\theta - \theta_0))]$ is continuous at θ_0 by assumption 14 and 16(a). Hence, proposition 2 implies $plim - H\widetilde{S_n}(\bar{\theta}) = H$. Moreover, $-H\widetilde{S_n}(\bar{\theta})^{-1}$ exists wpa.1 by assumption 16(b) and $\sqrt{nhh_q}\nabla\widetilde{S_n}(\theta_0) = O_p(1)$ by lemma 8 and lemma 5 yielding:

$$\sqrt{nhh_q}(\widetilde{\theta_n} - \theta_0) = H^{-1}\sqrt{nhh_q}\nabla\widetilde{S_n}(\theta_0) + o_p(1),$$

where lemma 8 further yields:

$$\sqrt{nhh_q}(\widetilde{\theta_n} - \theta_0) = H^{-1}\sqrt{nhh_q}\nabla S_n(\theta_0) + o_p(1),$$

and proposition 2 follows from lemma 5. QED

Proposition 4
proof: The first part of the proposition is straightforward by simply combining lemma 3 and lemma 7. For the second part, introducing some notation is convenient. Given a bandwidth pair $\sigma* \equiv (h* = h^{\gamma_1}, h_{q*} = h_q^{\gamma_2})$ define:

$$\Sigma_n(\sigma*) = \frac{1}{nh_*h_{q*}} \sum_{i=1}^n X_i X_i' |K(\frac{C_i + X_i'\theta_0}{h_*})|^2 |k(\frac{V_i - \overline{v}}{h_{q*}})|^2$$

Hence we have:

$$\widetilde{\Sigma_n} = \widetilde{\Sigma_n} - \Sigma_n(\sigma*) + \Sigma_n(\sigma*)$$

Using the same approach as in lemma 5, it is rapid to show lim $E[\Sigma_n(\sigma*)] = \Sigma$ as long as both $h*$ and h_{q*} converge to 0 as n approaches infinity. Furthermore, using a similar bounding method as in lemma 7 one has Var $[\Sigma_n(\sigma*)] \leq \frac{M_1 M_2}{nh_*h_{q*}} \int |K|^4 \int |k|^4$ where M_1 and M_2 are finite constants. Hence, if both $\int |K|^4$ and $\int |k|^4$ exist, one needs the additional condition that lim $nh_*h_{q*} = \infty$ to ensure plim $\Sigma_n(\sigma*) = \Sigma$. Under the assumption of proposition 4 this condition holds for h and h_q by assumption 13 and lim $nh^4 h_q^4 = \infty$ by assumption 17 so a fortiori for h^{γ_1} and $h_q^{\gamma_2}$. Secondly, we have:

$$\widetilde{\Sigma_n} - \Sigma_n(\sigma*) = \frac{1}{nh_*h_{q*}} \sum_{i=1}^n X_i X_i' [|\hat{K}_i|^2 |\hat{k}_i|^2 - |K_i|^2 |k_i|^2]$$

where

$$\hat{K}_i = K(\frac{C_i + X_i'\widetilde{\theta_0}}{h_*})$$

and

$$\hat{k}_i = k(\frac{\hat{V}_i - \overline{v}}{h_{q*}})$$

while K_i,k_i are their counterparts when both θ_0 and Π are used instead. Doing some simplifications with a triangular inequality and using the fact that $k(.)$ and $K(.)$ are bounded functions yields:

$$|\widetilde{\Sigma_n} - \Sigma_n(\sigma*)| \leq R_{1,n} + R_{2,n},$$

where

$$R_{1,n} = 2(||k||_{sup})(||K||_{sup})^2 \frac{1}{nh_*h_{q*}} \sum_{i=1}^n |X_i X_i'||\hat{k}_i - k_i|,$$

and

$$R_{2,n} = 2(||K||_{sup})(||k||_{sup})^2 \frac{1}{nh_*h_{q*}} \sum_{i=1}^n |X_i X_i'||\hat{K}_i - K_i|.$$

Finally, by assumption 17 the mean value theorem gives:

$$|\hat{k}_i - k_i| \leq \frac{1}{h_{q*}} ||k^{(1)}||_{sup} |\hat{V}_i - V_i|,$$

and

$$|\hat{K}_i - K_i| \leq \frac{1}{h_*} ||K^{(1)}||_{sup} |X_i'\hat{\Delta}|,$$

where $\hat{\Delta} = O_p(\frac{1}{\sqrt{nhh_q}})$ by proposition 3. Hence, there exists two finite constants ζ_1 and ζ_2 such that:

$$R_{1,n} \leq \zeta_1 \frac{1}{nh_*h_{q*}^2} ||\hat{\Pi} - \Pi|| \sum_{i=1}^n |X_i X_i'|||W_i||$$

and

$$R_{2,n} \leq \zeta_2 \frac{1}{nh_*^2 h_{q*}} ||\hat{\Delta}|| \sum_{i=1}^n |X_i X_i'|||X_i||$$

But under the assumption of proposition 3 we have $\frac{1}{n} \sum_{i=1}^n |X_i X_i'|||W_i|| = O_p(1)$ and $||\hat{\Pi} - \Pi|| = O_p(n^{-1/2})$ leading to:

$$\frac{1}{nh_* h_{q*}^2}||\hat{\Pi}-\Pi||\sum_{i=1}^n |X_i X_i'|||W_i|| = O_p(\frac{1}{h_* h_{q*}^2 n^{1/2}}) = o_p(1)$$

because lim $nh^4 h_q^4 = \infty$ by assumption 17, a fortiori lim $nh^{2\gamma_1} h_q^{4\gamma_2} = \infty$ when $\gamma_1 \in (0, 3/4]$ and $\gamma_2 \in (0,1]$. Additionally, $\frac{1}{n}\sum_{i=1}^n |X_i X_i'|||X_i|| = O_p(1)$ and $||\hat{\Delta}|| = O_p(\frac{1}{\sqrt{nhh_q}})$ yielding:

$$\frac{1}{nh_*^2 h_{q*}}||\hat{\Delta}||\sum_{i=1}^n |X_i X_i'|||X_i|| = O_p(\frac{1}{h_*^2 h_{q*}})O_p(\frac{1}{\sqrt{nhh_q}}) = o_p(1)$$

because assumption 17 implies lim $nh^{4\gamma_1+1} h_q^{2\gamma_2+1} = \infty$ whenever $\gamma_1 \in (0, 3/4]$ and $\gamma_2 \in (0,1]$. We conclude that $\widetilde{\Sigma_n} - \Sigma_n(\sigma*) = o_p(1)$. QED

Proposition 5
proof: For any function $f(.)$ defined on $[0,1]$ and parameter β let us introduce the followings:

$$\nabla S_*(\beta, f) \equiv \frac{1}{nh_*}\sum_{i=1}^n \tau(V_i)\alpha_i \tilde{X}_i K(\frac{C_i + \tilde{X}_i{}'\beta + f(V_i)}{h_*}),$$

and

$$HS_*(\beta, f) \equiv \frac{1}{nh_*^2}\sum_{i=1}^n \tau(V_i)\alpha_i \tilde{X}_i \tilde{X}_i{}' K^{(1)}(\frac{C_i + \tilde{X}_i{}'\beta + f(V_i)}{h_*}).$$

It is not too difficult using assumption S9(b)to establish (componentwise):

$$|HS_*(\tilde{\beta}_0, \tilde{\phi_0}) - H_n[\tilde{\beta}(v)]| \leq R_{1,n} + R_{2,n} + R_{3,n}$$

where,

$R_{1,n} \equiv ||K^{(2)}||_{sup} h_*^{-3}||\tilde{\beta}(v) - \tilde{\beta}_0||\frac{1}{n}\sum_{i=1}^n |\tilde{X}_i \tilde{X}_i{}'|.||\tilde{X}_i||,$

$R_{2,n} \equiv ||K^{(2)}||_{sup} h_*^{-3} sup_{i=1..n}\tau(V_i)|b_n' p_n(\hat{V}_i) - \tilde{\phi_0}(V_i)|\frac{1}{n}\sum_{i=1}^n |\tilde{X}_i \tilde{X}_i{}'|,$
and
$R_{3,n} \equiv ||K^{(1)}||_{sup}\frac{1}{nh_*^2}\sum_{i=1}^n |\tau(\hat{V}_i) - \tau(V_i)||\tilde{X}_i \tilde{X}_i{}'|.$

First, $\frac{1}{n}\sum_{i=1}^n |\tilde{X}_i \tilde{X}_i{}'|.||\tilde{X}_i|| = O_p(1)$ and $\sum_{i=1}^n |\tilde{X}_i \tilde{X}_i{}'| = O_p(1)$ by assumption S1. Also $h_*^{-3}||\tilde{\beta}(v) - \tilde{\beta}_0|| = O_p(\frac{1}{\sqrt{nhh_q h_*^3}})$ by assumption S1. Consequently $R_{1,n} = o_p(1)$ by assumption S11(a).

Secondly, $\frac{1}{n}\sum_{i=1}^n |\tilde{X}_i \tilde{X}_i{}'| = O_p(1)$ by assumption S1 and $sup_{i=1..n}\tau(V_i)|b_n' p_n(\hat{V}_i) - \tilde{\phi_0}(V_i)| = O_p(1)O(||\hat{\Pi} - \Pi||L_n) + O(||p_n||_{sup}||R_n||_{sup})$ by lemma 12. It follows that $R_{2,n} = o_p(1)$ by assumption S11(b), S11(c) and S11(d).

Lastly, writing $||K^1|| \equiv ||K^{(1)}||_{sup}$ yields:

$$R_{3,n} \equiv M_{1,n} + M_{2,n},$$

where,

$M_{1,n} = ||K^1||\frac{1}{nh_*^2}\sum_{i=1}^n |\tau(\hat{V}_i) - \tau(V_i)||\tilde{X}_i \tilde{X}_i{}'|1\{|\tilde{X}_i \tilde{X}_i{}'| < a_n\},$
and
$M_{2,n} \equiv ||K^1||\frac{1}{nh_*^2}\sum_{i=1}^n |\tau(\hat{V}_i) - \tau(V_i)||\tilde{X}_i \tilde{X}_i{}'|1\{|\tilde{X}_i \tilde{X}_i{}'| \geq a_n\},$

for any positive deterministic sequence a_n. It is rapid to establish $M_{1,n} = O_p(\frac{a_n}{\sqrt{n}h_*^2})$ by Newey et al.(1999) lemma A3. Also, a Cauchy Schwartz's inequality followed by a Tchebychev's inequality as in step 3 of lemma 3 gives $EM_{2,n} = O(\frac{1}{a_n h_*^2})$. So pick $a_n \propto \frac{log(n)^{1/2}}{h_*^2}$ and $R_{3,n} = o_p(1)$ follows by assumption S10.

Hence, $HS_*(\tilde{\beta}_0, \tilde{\phi_0}) \equiv H_n[\tilde{\beta}(v)] + o_p(1)$ is established and a fortiori $\nabla S_*(\tilde{\beta}_0, \tilde{\phi_0}) \equiv G_n[\tilde{\beta}(v)] + o_p(1)$. Lastly, invoking lemma 13 and 14 along with assumption 10 yields:

$$\text{plim } \nabla S_*(\tilde{\beta_0}, \tilde{\phi_0}) = 0 \text{ and plim } HS_*(\beta_0, \phi_0) = Q$$

The conclusion of proposition 5 arises since plim $\tilde{\beta}(v) = \tilde{\beta_0}$ by assumption S1 and Q^{-1} exists by assumption S8. QED

Proposition 6
proof: Since $\sqrt{nh_*}(\tilde{\beta}(v) - \tilde{\beta_0}) \equiv o_p(1)$ by assumption S1 and assumption S12, we obtain:

$$\sqrt{nh_*}(\bar{\beta} - \tilde{\beta_0}) \equiv -H_n[\tilde{\beta}(v)]^{-1}\sqrt{nh_*}G_n[\tilde{\beta}(v)] + o_p(1) \tag{8}$$

Also, by assumption S13 we get:

$$\sqrt{nh_*}G_n[\tilde{\beta}(v)] - \sqrt{nh_*}\nabla S_*(\tilde{\beta_0}, \tilde{\phi_0}) = o_p(1) \tag{9}$$

Furthermore, one can use the analogue of lemma 5 invoking this time assumption S6,S7,S9 S10 to allow the usage of the Lyapunov's Central Limit Theorem yielding:

$$\sqrt{nh_*}\{\nabla S_*(\tilde{\beta_0}, \tilde{\phi_0}) - E[\nabla S_*(\tilde{\beta_0}, \tilde{\phi_0})]\} \to_d N(0, \Xi) \tag{10}$$

Since $\sqrt{nh_*}E[\nabla S_*(\tilde{\beta_0}, \tilde{\phi_0})] = O(\sqrt{nh_*}h_*^s)$ by lemma 13 we conclude using assumption S14 that:

$$\sqrt{nh_*}E[\nabla S_*(\tilde{\beta_0}, \tilde{\phi_0})] = o(1) \tag{11}$$

Now use plim $H_n[\tilde{\beta}(v)]^{-1} = Q^{-1}$ under the assumptions of proposition 5 and the claim directly follows combining (8),(9),(10)and(11). QED

Section C
Assume that the assumptions of proposition 3 hold. 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 $\hat{\ell}_i \equiv C_i + X_i'\tilde{\theta}_n$. Here $\bar{v}$ is the value chosen to compute the KWSMS estimator. Suppose that there exists a partition of $\tilde{W} = (W_1, W_2')$ where W_1 is a scalar variable and W_2 is non empty. Let $\mu \bigotimes \mu$ indicates the product measure on $\mathbb{R}^2$ where μ is the Lebesgue measure. Define the following statistic:

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

where φ is a kernel and ξ a deterministic sequence. Also, define $M(l,v) \equiv E[\alpha(Y)|\ell = l, V = v]$ and $f(.,.)$ the joint density of (ℓ, V) with respect to $\mu \bigotimes \mu$ whenever this density exists. Suppose that the following assumptions hold:

C1. *$\partial M(l,v)\backslash\partial l$ and $\partial M(l,v)\backslash\partial v$ exist and are continuous in some open neighborhood of $(0,\bar{v})$. Also, $\partial^2 M(l,v)\backslash\partial^2 l$, $\partial^2 M(l,v)\backslash\partial^2 v$ and $\partial^2 M(l,v)\backslash\partial l\partial v$ exist in some open neighborhood of $(0,\bar{v})$.*

C2. *φ is a strictly positive kernel belonging to $\mathcal{K}_2$ and meets the same conditions as K in assumption 16.*

C3. *ξ_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 η are the bandwidths parameters selected to compute the KWSMS estimator as defined on page 28.*

C4. *The cdf of (ℓ, V) is absolutely continuous with respect to $\mu \bigotimes \mu$, its density at (l,v)=$(0,\bar{v})$ exists and is strictly positive. Also, there exists some open neighborhood of $(0,\bar{v})$ where $\partial f(l,v)\backslash\partial l$, $\partial f(l,v)\backslash\partial v$, $\partial^2 f(l,v)\backslash\partial^2 l$, $\partial^2 f(l,v)\backslash\partial^2 v$ and $\partial^2 f(l,v)\backslash\partial l\partial v$ exist and are continuous with $|\partial^2 f(l,v)\backslash\partial^2 l| < M$, $|\partial^2 f(l,v)\backslash\partial^2 v| < M$ and $|\partial^2 f(l,v)\backslash\partial l\partial v| < M$ for some $M < \infty$.*

C5. *The (cdf of) $C|\tilde{x}, v, w$ is absolutely continuous with respect to the Lebesgue measure a.e in $\tilde{x}, v, w$ and $W_1|\dot{x}, w_2$ is absolutely continuous with respect to the Lebesgue measure a.e in $\dot{x}, w_2$.*

C6. *(Define $F[.|x,l,v,w]$ the cdf of $\varepsilon|x,l,v,w$. Also, write $f(.|x,v,w)$ the density of $\ell|x,v,w$ and $f(.|x,l,w_2)$ the density of $V|x,l,w_2$ whenever those densities exist.)*

(i) *As functions of l:*

(ia)$f(l|x,v,w)$ *and* $F[-l|x,l,v,w]$ *belong to* $C^2_\infty(M)$ *for some* $M<\infty$ *a.e in x,v,w.*

(ib)$f(l|x,w)$ *and* $F[-l|x,l,-\lambda'w+\bar{v},w]$ *belong to* $C^2_\infty(M)$ *for some* $M<\infty$ *a.e in x,w for all* λ *parameter having the dimension of* W*. Furthermore,* $f(-\lambda'w+\bar{v}|x,l,w_2)$ *belongs to* $C^2_\infty(M)$ *for some* $M<\infty$ *a.e in x,*w_2 *for all* λ *parameter having the dimension of* W*.*

(ii) *As functions of* v:

(iia) $F[-l|x,l,v,w]$ *belongs to* $C^2_\infty(M)$ *for some* $M<\infty$ *a.e in x,l,w. Also,* $f(v|x,l,w_2)$ *belongs to* $C^2_\infty(M)$ *for some* $M<\infty$ *a.e in x,l,*w_2

(iib) $F[\lambda'x|x,-\lambda'x,v,w]$ *and* $f(-\lambda'x|x,v,w)$ *belong to* $C^2_\infty(M)$ *for some* $M<\infty$ *a.e in x,w for all* λ *parameter having the dimension of X. Also,* $f(v|x,w_2)$ *belongs to* $C^2_\infty(M)$ *for some* $M<\infty$ *a.e in x,*w_2 *for all* λ *parameter having the dimension of X.*

then under H_o*:* $Med(\varepsilon|\dot{X},\bar{v})=Med(\varepsilon|\bar{v})$ *a.s.,*

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

and,

$$(n\xi^2)^{-1}\sum\varphi(\tfrac{\hat{\ell}_i}{\xi})\varphi(\tfrac{\hat{V}_i-\bar{v}}{\xi})\to_p f(0,\bar{v}).$$

proof: The structure of this proof is analogous to that provided in Horowtiz (1993), proposition 2. The only difference deals with the number of variables conditioning Y and the presence of an additional nuisance term Π from the reduced form. The test is based upon the fact that under H_o: $Med(\varepsilon|\dot{X},\bar{v})=Med(\varepsilon|\bar{v})$ *a.s* one must have $M(0,\bar{v})=0$. The proof for the consistent estimator of $f(0,\bar{v})$ is omitted since it stems directly from what is to follow.

For any $\Delta'\equiv(\Delta_1',\Delta_2')$ where Δ_1 is $K\times 1$ and Δ_2 is $d\times 1$ introduce the following:

$$\bar{M}(\triangle)=\frac{(n\xi^2)^{-1}\sum\varphi(\frac{\ell_i+X_i'\Delta_1}{\xi})\varphi(\frac{\upsilon_i+W_i'\Delta_2}{\xi})\alpha(Y_i)}{(n\xi^2)^{-1}\sum\varphi(\frac{\ell_i+X_i'\Delta_1}{\xi})\varphi(\frac{\upsilon_i+W_i'\Delta_2}{\xi})},$$

where $\upsilon_i\equiv V_i-\bar{v}$. The key is to notice that $T_n=\bar{M}(\hat{\Delta})$ where $\hat{\Delta}'=((\tilde{\theta}-\theta_0(\bar{v}))',(\Pi-\hat{\Pi})')$. Applying Theorem 3.5-3.6 of Pagan and Ullah (1999) using assumptions C1 through C4 yields:

$$\sqrt{n\xi^2}\bar{M}(0)-M(0,\bar{v})\to_d\mathcal{N}(0,f(0,\bar{v})^{-1}(\textstyle\int|\varphi|^2)^2).$$

Also, using a Taylor's expansion furnishes:

$$\bar{M}(\hat{\Delta})=\bar{M}(0)+\tfrac{\partial\bar{M}}{\partial\Delta}|'_{\ddot{\Delta}}\hat{\Delta}$$

where plim $\ddot{\Delta}=0$ by the assumptions of proposition 3. Writing $\bar{a}$ the numerator of $\bar{M}$ and $\bar{b}$ its denominator gives:

$$\tfrac{\partial\bar{M}}{\partial\Delta}=\bar{b}^{-2}(\tfrac{\partial\bar{a}}{\partial\Delta}\bar{b}-\bar{a}\tfrac{\partial\bar{b}}{\partial\Delta}).$$

Under C2 and C6 one can apply lemma 2 as in lemma 3 to derive plim $\frac{\partial\bar{a}}{\partial\Delta}$ =lim E$\frac{\partial\bar{a}}{\partial\Delta}<\infty$ and plim $\frac{\partial\bar{b}}{\partial\Delta}$ =lim E$\frac{\partial\bar{b}}{\partial\Delta}<\infty$ uniformly over a compact set of $\mathbb{R}^{K+q}$ which contains 0. Likewise, by the same token as in lemma 4 using a classic convolution argument invoking C1-C4 returns plim $\bar{a}$=lim E$\bar{a}<\infty$ and plim $\bar{b}$ =lim E$\bar{b}<\infty$ uniformly over a compact set of $\mathbb{R}^{K+q}$ which contains 0. This establishes $\frac{\partial\bar{M}}{\partial\Delta}|_{\ddot{\Delta}}=O_p(1)$ and $\sqrt{n\xi^2}\bar{M}(\hat{\Delta})-\bar{M}(0)=o_p(1)$ follows because $\sqrt{n\xi^2}\hat{\triangle}=o_p(1)$ by proposition 3 and C3 (i.e. $\xi^2/hh_q=o(1)$). QED