

Smooth Minimum Distance Estimation and Testing with Conditional Estimating Equations: Uniform in Bandwidth Theory

Pascal Lavergne, Toulouse School of Economics (GREMAQ)

Valentin Patilea, CREST (ENSAI) and IRMAR (UEB)

This version: March 2013

Abstract

We study the influence of a bandwidth parameter in inference with conditional estimating equations. In that aim, we propose a new class of smooth minimum distance estimators and we develop a theory that focuses on uniformity in bandwidth. We establish a \sqrt{n} -asymptotic representation of our estimator as a process indexed by a bandwidth that can vary within a wide range including bandwidths independent of the sample size. We develop an efficient version of our estimator. We also study its behavior in misspecified models. We develop a procedure based on a distance metric statistic for testing restrictions on parameters as well as a bootstrap technique to account for the bandwidth's influence. Our new methods are simple to implement, apply to non-smooth problems, and perform well in our simulations.

Keywords: Semiparametric Estimation, Conditional Estimating Equations, Smoothing Methods, Asymptotic Efficiency, Hypothesis Testing, Bootstrap.

JEL classification: Primary C14; Secondary C12.

This work was completed while the first author was visiting CREST-ENSAI. This author gratefully acknowledges financial support from SSHRC. This work was presented at a Symposium at Indiana University, the Canadian Economic Association Meeting 2008, the European Meeting of the Econometric Society 2008, and in seminars at the University of Southampton, the Weierstrass Institute for Applied Analysis and Stochastics, Queen Mary University at London, and the University of Warwick. We thank participants, and in particular Russell Davidson, Vladimir Spokoiny, and Quang Vuong, for helpful comments.

Correspondence to: Pascal Lavergne, TSE GREMAQ, 21 Allées de Brienne, 31000 TOULOUSE, FRANCE. Emails: pascal.lavergne@univ-tlse1.fr, patilea@ensai.fr

1 Introduction

We address here inference on parameters identified by a set of conditional estimating equations. Common models that fit into this framework are regression models, conditional quantile models, linear or nonlinear simultaneous equations, and econometric models of optimizing agents. Earlier work on such models develops inference based on a finite number of unconditional estimating equations using instruments, as in Generalized Method of Moments (GMM), see Hansen (1982), or Empirical Likelihood (EL), see Qin and Lawless (1994). More recent work focuses on accounting for conditional equations at the outset to obtain more efficient estimators. Some methods rely on increasing the number of considered unconditional estimating equations (or instruments) with the sample size, such as the minimum distance approach of Ai and Chen (2003), or generalizations of GMM and EL by Donald, Imbens and Newey (2003) and Hjort, McKeague and Van Keilegom (2009). Carrasco and Florens (2000) uses a regularization approach to generalize the GMM approach to a continuum of estimating equations. Other EL-type estimators use nonparametric smoothing to estimate conditional equations, such as Antoine, Bonnal and Renault (2007), Kitamura, Tripathi and Ahn (2004), and Smith (2007a,b). In these different approaches, the estimators' sensitivity to the user-chosen parameter (number of estimating equations, regularization parameter, or smoothing parameter) remains largely unknown. This was one key motivation for the alternative estimator of Dominguez and Lobato (2004), that does not require any user-chosen parameter but is still \sqrt{n} -consistent and asymptotically normal.

Our central aim is to tackle the issue of the influence of the smoothing parameter when used for inference with conditional estimating equations. For doing so, we introduce a new estimator, labeled *Smooth Minimum Distance* (SMD). Heuristically, our method can be explained as follows. Let

$$\mathbb{E}[g(Z, \theta_0)|X] = 0 \quad \text{a.s.} \quad (1.1)$$

be conditional estimating equations, where $g(Z, \theta)$ is a known r -vector valued function, $r \geq 1$, of a random vector $Z = (Y', X')' \in \mathbb{R}^{d+q}$ and of a parameter $\theta \in \Theta \subset \mathbb{R}^p$. We assume that the true parameter value θ_0 is identified by (1.1). Consider the contrast

$$\mathbb{E}[g'(Z, \theta)\mathbb{E}(g(Z, \theta)|X) f(X)] = \mathbb{E}[\mathbb{E}(g'(Z, \theta)|X) \mathbb{E}(g(Z, \theta)|X) f(X)] , \quad (1.2)$$

where $f(\cdot)$ is the density of X and $g'(\cdot, \theta)$ denotes the transpose of $g(\cdot, \theta)$. This contrast is non-negative for any θ , and is zero if and only if $\theta = \theta_0$. Using a kernel estimator of

$\mathbb{E}(g(Z, \theta)|X) f(X)$, an empirical equivalent of (1.2) based on cross-section observations is given by

$$M_{n,h}(\theta) = \frac{1}{2n(n-1)} \sum_{1 \leq i \neq j \leq n} g'(Z_i, \theta)g(Z_j, \theta)K_{ij}, \quad K_{ij} = \frac{1}{h^q} K\left(\frac{X_i - X_j}{h}\right), \quad (1.3)$$

where $K(\cdot)$ is a multivariate kernel and $h = h_n$ a sequence of bandwidths. Zheng (1996) used such a statistic to construct a specification test for a regression model, and Delgado, Dominguez and Lavergne (2006) generalizes his idea to conditional moment restrictions. Here we use the above criterion for estimation instead. Intuitively the minimizer of $M_{n,h}(\theta)$ should be a consistent estimator of θ_0 , but will be dependent of the bandwidth's choice. However, as will be shown, *one does not need to estimate consistently $\mathbb{E}(g(Z, \theta)|X) f(X)$ to estimate consistently θ_0 , and our estimator is \sqrt{n} -consistent and asymptotically normal for a fixed, non-vanishing bandwidth, as well as a vanishing one.* Hence our estimator bridges a gap between Dominguez and Lobato's method, which does not require a user-chosen parameter, and the competing EL and GMM-type methods that rely on smoothing.

One may wonder why we introduce a new class of estimators when many are already available. We believe that our approach has some interesting features, both from a theoretical as well as a practical viewpoint as detailed below. In particular, SMD is consistent when the bandwidth is fixed, while estimators based on instruments may be inconsistent if their number is kept fixed, as pointed out by Dominguez and Lobato (2004). Though our approach is not a generalization of any existing one, it shares some common features with many as explained below. Hence, our main conclusion that the bandwidth's influence should and can be accounted for in inference is likely to carry over to other estimators.

To derive the properties of SMD, we do not adopt the usual simultaneous asymptotics, where the bandwidth is assumed to decrease to zero as the sample size increases. Instead we consider our estimator as indexed by the bandwidth h . Our analysis is thus akin to recent work on uniform in bandwidth properties of kernel estimators, see Einmahl and Mason (2005) and the references therein. But to our knowledge our work is the first of its kind for a semiparametric estimator. As we show, SMD is \sqrt{n} -consistent and asymptotically normal uniformly in h for a large range of bandwidths *including bandwidths independent of the sample size*. However, its asymptotic variance depends on the bandwidth. In that respect, our theory sheds light on the bandwidth's influence. Moreover, it allows for a data-dependent bandwidth, which is not available at the moment for alternative estimators. To obtain reliable inference that accounts for the bandwidth's influence, we extend a bootstrap method proposed by Jin,

Ying and Wei (2001) and Bose and Chatterjee (2003) and we show that a bootstrap testing procedure based on the SMD objective function is valid uniformly in the bandwidth.

We also derive a number of desirable results of SMD. First, its main properties are robust to potential misspecification, i.e. when there exists no θ_0 such that (1.1) holds. This is valuable as misspecified models can still provide useful information, see e.g. White (1981). Schennach (2007) shows that the excellent properties of the EL estimator degrades enormously under the slightest misspecification, causing the loss of \sqrt{n} -consistency. Little is known on the behavior of the EL generalizations to (1.1), but one should fear that a similar phenomenon occurs. Second, the SMD objective function can be tailored to obtain a semiparametrically efficient estimator, using a vanishing bandwidth and weighting optimally the functions in $g(Z, \theta)$.

The other practical advantages of our estimator are as follows. Our estimator is as simple to implement as Dominguez and Lobato's estimator. It is versatile and applies even when $g(\cdot, \theta)$ are non-differentiable, as in quantile regression. When the bandwidth is fixed, SMD is similar in spirit to Dominguez and Lobato's (2004) estimator and both are \sqrt{n} -asymptotically normal. Their variances generally depend on the unknown data generating process, so that it is generally not possible to rank them. Our contribution is to provide an alternative to Dominguez and Lobato's approach and to develop bootstrap tests. However, our current theory only applies to independent observations, while the latter allow for dependent ones. When the bandwidth goes to zero, our estimator can be rendered efficient and is thus comparable to alternative estimators. By contrast with the EL-type estimators of Donald and al. (2003) and Kitamura and al. (2004), efficient SMD does not rely on a double optimization problem whose complexity increases with the sample size. It generally obtains from a two-step procedure and requires estimation of the conditional variance of $g(\cdot, \theta_0)$, but avoids estimation of the optimal instruments, involving derivatives of $\mathbb{E}[g(Z, \theta_0)|X]$ with respect to θ , as in efficient GMM, see e.g. Newey (1993).

Our paper is structured as follows. In Section 2, we present our estimation method in relation with previous approaches. We then develop our uniform-in-bandwidth theory, including extensions to possibly misspecified models and efficient estimation. In Section 3, we investigate a distance-metric procedure for testing restrictions on parameters and the validity of a bootstrap method to determine critical values. In Section 4, we study the small sample behavior of our estimator via a simulation study. Our estimator performs well in our experiments, and the bootstrap test is found to yield accurate levels and good power in moderate samples. Technical assumptions are gathered in Section 5. Two Appendices provides some sufficient conditions for our technical assumptions to hold. To keep our paper short, proofs

are made available on the web at <http://gremaq.univ-tlse1.fr/perso/lavergne>.

2 SMD Estimation

2.1 The Estimator

We set up a general framework that allows to review the literature and to introduce our estimator. Accounting for conditional moments (1.1) is key for identification and efficiency. If one transforms the conditional moments into a finite number of unconditional ones, then specifying primitive conditions for identification may become difficult for a nonlinear in parameters function $g(\cdot, \cdot)$. This is illustrated by Newey and McFadden (1994) for a simple consumption-based CAPM model, where $Z = (Y_1, Y_2, X)' \in \mathbb{R}^{2+q}$, $\theta = (\beta, \gamma)'$, and $g(Z, \theta) = \beta Y_1 Y_2^\gamma - 1$. Dominguez and Lobato (2004) discuss a nonlinear regression where similar issues appear, see our simulations section. Moreover, accounting for conditional moments is expected to increase estimation accuracy. Some existing approaches directly deal with conditional estimating equations and gather them in a contrast as

$$\int \|\mathbb{E}[g(Z, \theta)|X = t]\|^2 d\mu(t),$$

where $\|u\|^2 = \sum_{i=1}^r |u_i|^2$. An estimator of the contrast is then minimized over the parameter space. Cristobal Cristobal, Faraldo Roca, and Gonzalez Manteiga (1987) is an early work using such an approach in a regression context.¹ The Conditional Continuously Updated Estimator studied by Antoine and al. (2007) correspond to $\mu = P_X$, the distribution of X , and appropriate reweighting of $g(Z, \theta)$ through its conditional variance. As noted in the Introduction, the SMD objective function (1.3) estimates the contrast (1.2) when the bandwidth tends to zero. Other existing approaches can be embedded into a framework that replaces the conditional equations (1.1) by a set of unconditional ones, so that

$$\mathbb{E}[g(Z, \theta_0)|X] = 0 \quad \text{a.s.} \Leftrightarrow \mathbb{E}[g(Z, \theta_0)l(X, t)] = 0 \quad \forall t \in T, \quad (2.4)$$

for a well-chosen family of functions $l(\cdot, t)$ indexed by $t \in T$.² For instance, one can consider polynomial or Fourier series indexed by $t \in \mathbb{N}$. Standard GMM accounts only for a finite

¹We thank the Associate Editor for pointing out this reference.

²Some EL-type methods are not included, such as Kitamura and al. (2004).

number of such unconditional moments. Donald, Imbens, and Newey (2003) consider GMM when the number of moments increases with the sample size, where the contrast is

$$\int_{\mathbb{R}^q} \|\text{Var}^{-1/2}[g(Z, \theta_0)l(X, t)]\mathbb{E}[g(Z, \theta)l(X, t)]\|^2 d\mu_n(t),$$

with μ_n a measure with point masses at $\{1, \dots, K_n\}$ with $K_n \rightarrow \infty$. The estimator of Ai and Chen (2003) proceeds of the same logic, as they themselves point out. However, as shown by Stinchcombe and White (1998), there exists many families of functions $l(\cdot, t)$ such that (2.4) holds. Carrasco and Florens (2000) suggest to use $\exp(tX)$ for X univariate and t in an interval around zero. Dominguez and Lobato (2004) choose $\mathbb{I}(X \leq t)$ and consider the theoretical contrast

$$\int_{\mathbb{R}^q} \|\mathbb{E}[g(Z, \theta)\mathbb{I}(X \leq t)]\|^2 dP_X(t).$$

It should be clear that an incredible variety of estimators could be considered, each based on a contrast of the form

$$\int_T \|\mathbb{E}[g(Z, \theta)l(X, t)]\|^2 d\mu(t)$$

for suitable choices of $l(X, t)$ and μ . One could even generalize further by considering contrasts based on a L^p measure instead of a L^2 one. Each estimator is expected to be \sqrt{n} -consistent, but none can be expected to be best in every particular case.

Our new estimator corresponds to the particular choice $l(X, t) = e^{it'X}$, see Bierens (1982). For a probability measure μ with support \mathbb{R}^p and $h > 0$, θ_0 is the unique minimizer of

$$\int_{\mathbb{R}^q} \|\mathbb{E}[g(Z, \theta)e^{it'X}]\|^2 d\mu(ht). \quad (2.5)$$

For a typical measure μ such as the standard normal, low frequency moments (corresponding to small t) have more importance than high frequency moments (large t). The bandwidth h allows to vary the dispersion of the measure, so that for h going to zero, the measure tends to equalize weights across frequencies. The above contrast writes

$$\mathbb{E} \left[g'(Z_i, \theta)g(Z_j, \theta) \int_{\mathbb{R}^q} e^{it'(X_i - X_j)} d\mu(ht) \right] = \mathbb{E} \left[g'(Z_i, \theta)g(Z_j, \theta)h^{-q} \int_{\mathbb{R}^q} e^{it'(X_i - X_j)/h} d\mu(t) \right],$$

where $Z_i = (Y_i', X_i)'$ and Z_j are independent copies of Z . If we denote by $K((X_i - X_j)/h)$ the last integral (with respect to μ), the contrast becomes

$$\mathbb{E} [g'(Z_i, \theta)g(Z_j, \theta)h^{-q}K((X_i - X_j)/h)]. \quad (2.6)$$

It is uniquely minimized at θ_0 , irrespective of the choice of the parameter h .³ Now, with at hand independent copies $\{Z_1, \dots, Z_n\}$ from Z , a natural estimator of our contrast is obtained after replacing the expectation by a double average, leading to $M_{n,h}(\theta)$. Our SMD estimator is then defined as

$$\tilde{\theta}_{n,h} = \arg \min_{\Theta} M_{n,h}(\theta).$$

The resulting estimator is practically as simple to implement as Dominguez and Lobato's estimator. It depends continuously on the bandwidth h and thus allow for a convenient study of the bandwidth's influence. As an example, consider the previously mentioned nonlinear CAPM-type model where $g(Z, \theta) = \beta Y_1 Y_2^\gamma - 1$ and $K(\cdot)$ is the standard normal multivariate density $\phi(\cdot)$. Then

$$\begin{aligned} M_{n,h}(\theta) &= \frac{1}{2n(n-1)} \sum_{1 \leq i \neq j \leq n} (\beta Y_{1i} Y_{2i}^\gamma - 1) (\beta Y_{1j} Y_{2j}^\gamma - 1) h^{-q} \phi\left(\frac{X_i - X_j}{h}\right) \\ \mathbb{E}M_{n,h}(\theta) &= \mathbb{E} \left[\mathbb{E}(\beta Y_{1i} Y_{2i}^\gamma - 1 | X_i) \mathbb{E}(\beta Y_{1j} Y_{2j}^\gamma - 1 | X_j) h^{-q} \phi\left(\frac{X_i - X_j}{h}\right) \right] \\ &= \int_{\mathbb{R}^q} \|\mathbb{E}(\beta Y_1 Y_2^\gamma - 1 | X) f(X) e^{-it'X}\|^2 \phi(ht) dt, \end{aligned}$$

which is zero if and only if $\mathbb{E}(\beta Y_1 Y_2^\gamma - 1 | X) = 0$ a.s. Other kernels could alternatively be used, see next section.

2.2 Consistency

We use the following notations throughout the remaining of the paper. For a matrix A , A' is its transpose, $\|A\|$ is the Frobenius norm, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and the largest eigenvalue of A when A is symmetric. For a real-valued function $l(\cdot)$, $\nabla_\theta l(\cdot)$ and $\mathbb{H}_{\theta,\theta} l(\cdot)$ respectively denote the p column vector of first partial derivatives and the $p \times p$ matrix of second derivatives with respect to $\theta \in \mathbb{R}^p$. For a vector-valued function $l(\cdot) \in \mathbb{R}^r$, $\nabla_\theta l(\cdot)$ denotes the $p \times r$ matrix of first derivatives of the entries of $l(\cdot)$ with respect to entries of θ .

We introduce a more general version of our estimator, where

$$M_{n,h}(\theta) = \frac{1}{2n(n-1)} \sum_{1 \leq i \neq j \leq n} g'(Z_i, \theta) W_n^{-1/2}(X_i) W_n^{-1/2}(X_j) g(Z_j, \theta) K_{ij},$$

³The relationship between (2.5) and (2.6) was first noted in Fan and Li (2000) in the context of specification testing.

and $W_n(\cdot) \rightarrow W(\cdot)$ is a sequence of non-random positive definite (p.d.) weighting matrices that may depend on θ . In practice the typical use of our estimator involves first estimation with an identity weighting matrix and fixed bandwidth, and second efficient estimation with an estimated optimal weighting matrix and a vanishing bandwidth, as detailed in Section 3.4. To obtain scale-invariance with respect to X , we recommend that the observations X be scaled in practice.

To avoid technicalities, our assumptions are spelled out in Section 5, and only the central ones are discussed in the text. From our previous discussion, a restriction on the kernel $K(\cdot)$ is that

$$K(u) = \int_{\mathbb{R}^q} e^{it'u} d\mu(t)$$

for a measure $\mu(\cdot)$ that is strictly positive (but for a set of isolated points). This is true for products of the triangular, normal, Laplace, and logistic densities, see Johnson, Kotz, and Balakrishnan (1995, Section 23.3), and for a Student density, see Hurst (1995).⁴ One can also allow for other kernels if X has a bounded support. In that situation indeed, it is sufficient to consider a continuum of moment conditions $\mathbb{E}[g(Z, \theta_0)e^{it'X}] = 0$ for t in an neighborhood of the origin, see Bierens (1982), so that any positive measure μ whose support includes a neighborhood of 0 is suitable. This potentially yields higher-order kernels taking negative values, such as the normalized sinc kernel corresponding to a uniform μ . We also assume that the class of functions $\{(x, \bar{x}) \mapsto K((x - \bar{x})/h), x, \bar{x} \in \mathbb{R}^q, h > 0\}$ is Euclidean for a constant envelope. We give a formal definition of an Euclidean class in Section 5, but we here explain what this entails. Sufficient conditions for the above class of functions to be Euclidean in the case of product kernels is that the univariate kernel is bounded and of bounded variation, see e.g. Nolan and Pollard (1987). Any of the above kernels satisfies these requirements. We also assume that the families $\mathcal{G}_k = \{g^{(k)}(\cdot, \theta) : \theta \in \Theta\}$, $1 \leq k \leq r$, are Euclidean for a squared integrable envelope. Again the Euclidean property is a mild one for parametric families of functions. Practically, any function $g(\cdot, \theta)$ which is continuous almost everywhere for any θ with an uniformly bounded second moment fulfills the Euclidean condition. We are now ready to state our first result.

Theorem 2.1. *For an i.i.d. sample and under Assumptions 1(i)–(ii), 2, 3, and 4(i)–(ii), $\sup_{h_0 \geq h > 0, nh^{2q} \geq \ln(n+1)} \|\tilde{\theta}_{n,h} - \theta_0\| = o_p(1)$ for arbitrary finite $h_0 > 0$.*⁵

⁴Some of the sufficient conditions listed in Appendix A require a number of finite moments, so that the Student density should be chosen with enough degrees of freedom.

⁵Here and in what follows, we abstract from measurability issues of the suprema with respect to h .

Consistency is obtained under more general conditions than the ones imposed for most competing estimators, see e.g. Kitamura et al. (2004) who impose smoothness of the function $g(\cdot, \cdot)$, a vanishing bandwidth, and more stringent conditions on its behavior. It is also obtained irrespective of the sequence of matrices $W_n(\cdot)$, assuming it is well-behaved. This will prove useful in Section 2.4.

2.3 Uniform in Bandwidth Asymptotic Normality

This section contains our central result. We assume smoothness of the functions $\tau(x, \theta) = \mathbb{E}[g(Z, \theta)|X = x]$, specifically that all second partial derivatives with respect to the components of θ exist in a neighborhood \mathcal{N} of θ_0 and that $\|\mathbb{H}_{\theta, \theta} \tau^{(k)}(X, \theta) - \mathbb{H}_{\theta, \theta} \tau^{(k)}(X, \theta_0)\| \leq H(X) \|\theta - \theta_0\|^a$, $\forall \theta \in \mathcal{N}$, $k = 1, \dots, r$, for some $a \in (0, 1]$ and $H(\cdot)$ with $\mathbb{E}H^4 < \infty$. While this is implied by a similar condition on $g(Z, \theta)$, our assumption does not require differentiability of $g(Z, \theta)$ and thus allows to cover the case of quantile regression. Let $\mathcal{F}_n = \{\phi_{n,h}(\cdot) : h \in [0, h_0]\}$ be the family of functions defined by

$$\phi_{n,h}(z) = \mathbb{E} [\nabla_{\theta} \tau(X, \theta_0) W_n^{-1/2}(X) h^{-q} K((x - X)/h)] W_n^{-1/2}(x) g(z, \theta_0), \quad \text{for } h \in (0, h_0],$$

and $\phi_{n,0}(z) = \nabla_{\theta} \tau(x, \theta_0) W_n^{-1}(x) g(z, \theta_0) f(x)$. We introduce the following condition

Condition E. *The class of functions $\psi_n(\cdot)$ is such that*

$$\{x \mapsto \int \psi_n(x - uh) K(u) du : h \in [0, h_0]\}$$

*is uniformly Euclidean for the envelope $\Psi(\cdot)$. Here, uniformly means that the envelope and the constants in the definition of the Euclidean family are independent of n .*⁶

If Condition E holds for $\nabla_{\theta} \tau(\cdot, \theta_0) W_n^{-1/2}(\cdot) f(\cdot)$, then the sequence of (centered) empirical processes $\{\mathbb{G}_n \phi_{n,h} : h \in [0, h_0]\}$ indexed by \mathcal{F}_n , where $\mathbb{G}_n \phi_{n,h} = n^{-1/2} \sum_{i=1}^n \phi_{n,h}(Z_i)$, weakly converges to a tight zero-mean Gaussian process with covariance function

$$\begin{aligned} \Delta_{h_1, h_2} &= \mathbb{E} [\nabla_{\theta} \tau(X_1, \theta_0) W^{-1/2}(X_1) W^{-1/2}(X_2) \text{Var} [g(Z_2, \theta_0)|X_2] W^{-1/2}(X_2) \\ &\quad W^{-1/2}(X_3) \nabla'_{\theta} \tau(X_3, \theta_0) h_1^{-q} K((X_1 - X_2)/h_1) h_2^{-q} K((X_3 - X_2)/h_2)] . \end{aligned}$$

⁶Condition (E) can be weakened to a uniform entropy condition, as in van der Vaart (1998, Theorem 19.28) or van der Vaart and Wellner (1996, Theorem 2.11.22).

Sufficient mild conditions that guarantee Condition E are provided in Appendix A. In particular, it is sufficient that the functions belong to some Sobolev space, or are Hölder continuous on their support.⁷ Let us finally define $V_{n,h} = \mathbb{H}_{\theta,\theta} \mathbb{E} M_{n,h}(\theta_0)$ and

$$V_h = \lim_{n \uparrow \infty} V_{n,h} = \mathbb{E} \left[\nabla_{\theta} \tau(X_1, \theta_0) W^{-1/2}(X_1) W^{-1/2}(X_2) \nabla'_{\theta} \tau(X_2, \theta_0) h^{-q} K((X_1 - X_2)/h) \right].$$

We assume that $\inf_{n,h} \lambda_{\min}(V_{n,h}) > 0$. This holds as soon as $\mathbb{P}[a' \nabla_{\theta} \tau(X, \theta_0) = 0] < 1$ for all $a \neq 0$, as shown in our proofs. This is the natural extension to conditional moments of the usual assumption of a full rank matrix $\nabla_{\theta} \mathbb{E} g(Z, \theta_0)$ for unconditional moments $\mathbb{E} g(Z, \theta_0) = 0$.

Theorem 2.2. *Let $h \in \mathcal{H}_n = \{h_0 \geq h > 0 : nh^{4q/\alpha} \geq C\}$ for arbitrary constants $h_0, C > 0$, and $0 < \alpha < 1$. For an i.i.d. sample, under Assumptions 1–6 and $\inf_{n,h} \lambda_{\min}(V_{n,h}) > 0$, $\sqrt{n}(\tilde{\theta}_{n,h} - \theta_0) = -V_{n,h}^{-1} \mathbb{G}_n \phi_{n,h} + o_p(1)$ uniformly in $h \in \mathcal{H}_n$, and thus converges in distribution to a tight random process indexed by h whose marginal distributions are zero-mean normal with covariance function $V_{h_1}^{-1} \Delta_{h_1, h_2} V_{h_2}^{-1}$.*

By contrast to existing results, we do not consider a deterministic sequence of bandwidths nor do we assume away the bandwidth's influence by assuming that it converges to zero. As shown in our proofs, if one assumes differentiability of $g(Z, \theta)$, a similar result obtains uniformly for h in $\{h_0 \geq h > 0 : nh^{2q/\alpha} \geq C\}$. This lower bound on h is similar or weaker than the one found in other work. For instance, Andrews (1994) studies a general class of estimators depending on a preliminary kernel estimator and notes that the latter should converge faster than $n^{-1/4}$, which is equivalent to the requirement that $nh^{q/2}$ diverges. The same restriction is imposed by Donald, Imbens and Newey (2003) for GMM with an increasing number of moments, and a stronger one is required for their EL estimator.

Our uniform-in-bandwidth theory sheds light on the bandwidth's role on the estimator's distribution: it does not affect its first-order unbiasedness, nor its rate of convergence, but it does affect its variance. The \sqrt{n} -asymptotic variance $V_h^{-1} \Delta_{h,h} V_h^{-1}$ of $\tilde{\theta}_{n,h}$ can be easily estimated. When $g(Z, \theta)$ is differentiable, we can estimate V_h and $\Delta_{h,h}$ respectively by

$$\frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \nabla_{\theta} g(Z_i, \tilde{\theta}_{n,h}) W_n^{-1/2}(X_i) W_n^{-1/2}(X_j) \nabla'_{\theta} g(Z_j, \tilde{\theta}_{n,h}) K_{ij},$$

and

$$\frac{1}{n(n-1)(n-2)} \sum_{1 \leq i \neq j \neq k \leq n} \nabla_{\theta} g(Z_i, \tilde{\theta}_{n,h}) W_n^{-1/2}(X_i) W_n^{-1/2}(X_j) \hat{\Omega}_n(X_j) \times W_n^{-1/2}(X_j) W_n^{-1/2}(X_k) \nabla'_{\theta} g(Z_k, \tilde{\theta}_{n,h}) K_{ij} K_{jk}.$$

⁷In estimation with an identity weighting matrix, the family of functions $\nabla_{\theta} \tau(\cdot, \theta_0) f(\cdot)$ do not depend on n , so that the uniform part in Condition E can be discarded.

Here $\widehat{\Omega}_n(X_j)$ is a $r \times r$ matrix whose typical element “approximates” the covariance of different components of $g(Z_i, \theta_0)$ conditional on X_i , such as a nonparametric estimator of $\text{Var}[g(Z_i, \theta_0) \mid X_i]$. Alternatively, one can use the elements of $g(Z_i, \tilde{\theta}_{n,h})$ to approximate conditional covariances, in the spirit of the Eicker-White variance estimator. Consistency of the above estimators is pretty straightforward to establish. If $g(\cdot, \cdot)$ is not differentiable, one can use numerical methods similar to the ones in Pakes and Pollard (1989).

2.4 Study Under Misspecification

We study our estimator when there is no θ_0 such that (1.1) holds. As previously argued, this is useful at least as a “robustness” check. Denote the probability limit of $\tilde{\theta}_{n,h}$ as $\bar{\theta}_{n,h}(W_n) = \bar{\theta}_{n,h} = \arg \min_{\Theta} \mathbb{E} M_{n,h}(\theta)$ and assume it is unique. Because $\tilde{\theta}_{n,h}$ is not constant, we need to extend our smoothness assumptions on the different function entering our analysis to any value that it can take. We also need a mild strengthening of Condition E to account for the non-constancy of $\bar{\theta}_{n,h}$. We say that a sequence of real-valued functions $\psi_n(\cdot, \cdot)$ satisfies Condition ME with kernel $K(\cdot)$ for an envelope $\Psi(\cdot)$ if for each $n \geq 1$ the class of functions

$$\{x \mapsto \int \psi_n(x - uh, \theta) K(u) du : h \in [0, h_0], \theta \in \Theta\}$$

is uniformly Euclidean for the envelope $\Psi(\cdot)$. Let $\bar{\theta}_h = \lim_{n \uparrow \infty} \bar{\theta}_{n,h}$,

$$\begin{aligned} \bar{\Delta}_{h_1, h_2} &= \mathbb{E} \left[\nabla_{\theta} \tau(X_1, \bar{\theta}_{h_1}) W^{-1/2}(X_1) W^{-1/2}(X_2) \mathbb{E} \left[g(Z_2, \bar{\theta}_{h_1}) g'(Z_2, \bar{\theta}_{h_2}) \mid X_2 \right] W^{-1/2}(X_2) \right. \\ &\quad \left. W^{-1/2}(X_3) \nabla_{\theta} \tau(X_3, \bar{\theta}_{h_2}) h_1^{-q} K((X_1 - X_2)/h_1) h_2^{-q} K((X_3 - X_2)/h_2) \right], \\ \bar{V}_{n,h} &= H_{\theta, \theta} \mathbb{E} M_{n,h}(\bar{\theta}_h), \quad \text{and} \quad \bar{V}_h = \lim_{n \uparrow \infty} \bar{V}_{n,h}. \end{aligned}$$

Theorem 2.3. *For an i.i.d. sample and $h \in \mathcal{H}_n$, under Assumptions 1-(i), 2, 3, 4-(i) to (iii), M4, M5, and M6, if $\sup_{n,h} \lambda_{\max}(\bar{V}_{n,h}) < \infty$ and $\inf_{n,h} \lambda_{\min}(\bar{V}_{n,h}) > 0$, $\sqrt{n} (\tilde{\theta}_{n,h} - \bar{\theta}_{n,h})$ converges in distribution to a tight random process whose marginal distributions are zero-mean normal with covariance function $\bar{V}_{h_1}^{-1} \bar{\Delta}_{h_1, h_2} \bar{V}_{h_2}^{-1}$.*

2.5 Efficient SMD Estimation

We now turn to rendering our estimator semiparametrically efficient: this is desirable from a theoretical viewpoint and indicates that our SMD estimator compares well to competitors.

Let $\check{\theta}_n$ be a \sqrt{n} -consistent SMD estimate of θ_0 , computed for instance by choosing $W_n(\cdot) = I$ and $h = h_0$ fixed. Consider the nonparametric estimator of $\text{Var}[g(Z, \theta_0) | X = x]f(x)$

$$\widehat{W}_n(x, \theta) = \frac{1}{nb^q} \sum_{1 \leq k \leq n} g(Z_k, \theta)g'(Z_k, \theta)L((x - X_k)/b), \quad (2.7)$$

where $L(x)$ is a kernel and b is a vanishing bandwidth. The efficient SMD is $\widehat{\theta}_{n,h,b} = \arg \min_{\Theta} \widehat{M}_{n,h,b}(\theta)$, where

$$\widehat{M}_{n,h,b}(\theta) = \frac{1}{2n(n-1)} \sum_{1 \leq i \neq j \leq n} g'(Z_i, \theta)\widehat{W}_n^{-1/2}(X_i, \check{\theta}_n)\widehat{W}_n^{-1/2}(X_j, \check{\theta}_n)g(Z_j, \theta)K_{ij}.$$

It is thus in general a two-step estimator, but a single quasi-Newton step around the preliminary estimator can also be used when $g(\cdot)$ is differentiable. In non-differentiable cases, one would numerically optimize the new objective function, in which case the inefficient estimator would provide a natural starting value. For quantile regression, a leading non-differentiable case where $g(Z, \theta) = \mathbb{I}[Y - \mu(X, \theta) \leq 0] - \rho$, $W(x) = \rho(1 - \rho)f(x)$, so that no preliminary estimator is needed, and a one-step efficient estimator obtains, as the ones recently proposed by Otsu (2008) and Komunjer and Vuong (2010).

We consider our estimator as a process indexed by h and b . It is easy to show that $\widehat{\theta}_{n,h,b}$ is consistent by adapting the proof of Theorem 2.1. As efficiency requires a vanishing bandwidth, our following analysis assumes that h goes to zero, and in addition that the bandwidth b is in the same range than h . No relationship between the two bandwidths is required, though in practice they can be chosen to be equal. By convention, $\widehat{W}_n(x, \theta) = I$ when the right-hand side of (2.7) is not positive definite. However, the probability of this event vanishes when n grows if $L(\cdot)$ is a density with bounded support which is strictly positive around the origin. We use a generalization of a result from Einmahl and Mason (2005) on kernel estimators to control the behavior of the variance estimator. The main supplementary assumptions needed for our next results are of a bounded support for X and some smoothness of $\mathbb{E}[g(Z, \theta)g'(Z, \theta)|X = x]$ in x and in θ around θ_0 . Allowing for an unbounded support would involve introducing some trimming into the objective function, as done by Kitamura et al. (2004), but this is outside the scope of this paper. They also note that trimming does not affect their estimator in practice and in view of our following simulations results we feel confident that the same would apply here.

Theorem 2.4. For an i.i.d. sample, under Assumptions 1, 2, E2, E4, 5, E6, and E7, $\sqrt{n}(\hat{\theta}_{n,h,b} - \theta_0)$ is asymptotically $N(0, \Sigma^{-1})$ uniformly in $h, b \in \mathcal{H}'_n$ with

$$\Sigma = \mathbb{E} [\nabla_{\theta} \mathbb{E} [g(Z, \theta_0) | X] \text{Var}^{-1} [g(Z, \theta_0) | X] \nabla'_{\theta} \mathbb{E} [g(Z, \theta_0) | X]] .$$

The asymptotic variance Σ^{-1} is the semiparametric efficiency bound characterized by Chamberlain (1987).

3 SMD-Based Testing for Parameter Restrictions

The main message of our previous section is that the bandwidth influences the behavior of the estimator. Hence this should be accounted for in inference. We now develop a testing theory with that aim in mind. In what follows, we do not assume that one uses the efficient version of SMD, though our following results also apply to this case.

3.1 Asymptotics

Suppose we want to test the parametric restriction in explicit form

$$H_0 : \theta_0 = R(\gamma_0) , \tag{3.8}$$

where $\gamma_0 \in \mathbb{R}^s$ with $s \leq p$ and $R(\cdot)$ is a function from $\Gamma \subset \mathbb{R}^s$ on Θ . We assume that $R(\cdot)$ is twice continuously differentiable and that $\nabla_{\gamma} R(\gamma_0)$ has rank $\bar{r} = s \geq 1$ or $\bar{r} = 0$. The latter case corresponds to the situation where all parameters values are completely determined under H_0 . Under these regularity assumptions, explicit restrictions can be turned into equivalent implicit ones of the form $r(\theta_0) = 0$, see e.g. White (1994, Section 8.1). The explicit form considered here links the parameter θ to a parameter γ of smaller dimension. Such a restricted parametrization is computationally attractive because it reduces the dimensionality of the optimization problem. The constrained SMD estimator is $\tilde{\theta}_{n,h}^R = \arg \min_{\theta \in \Theta, \theta = R(\gamma)} M_{n,h}(\theta)$. A distance metric statistic for testing H_0 is

$$DM_{n,h} = 2n \left[M_{n,h} \left(\tilde{\theta}_{n,h}^R \right) - M_{n,h} \left(\tilde{\theta}_{n,h} \right) \right] .$$

One could alternatively consider tests of the Wald or Score type, but a theoretical advantage of the distance metric test is that it is automatically invariant to the formulation of the null hypothesis. For $h \in [0, h_0]$, let

$$\Lambda_{n,h} = \left[I_p - V_{n,h}^{1/2} \nabla'_{\gamma} R(\gamma_0) \left[\nabla_{\gamma} R(\gamma_0) V_{n,h} \nabla'_{\gamma} R(\gamma_0) \right]^{-1} \nabla_{\gamma} R(\gamma_0) V_{n,h}^{1/2} \right] V_{n,h}^{-1/2} \Delta_{n,h,h} V_{n,h}^{-1/2} ,$$

when $\bar{r} = s$ and $\Lambda_{n,h} = V_{n,h}^{-1/2} \Delta_{n,h} V_{n,h}^{-1/2}$ when $\bar{r} = 0$.

Theorem 3.1. *Under the assumptions of Theorem 2.2, then uniformly in $h \in \mathcal{H}_n$, (i) under H_0 , $DM_{n,h} - (\mathbb{G}_n \phi_{n,h})' \Lambda_{n,h} (\mathbb{G}_n \phi_{n,h}) = o_p(1)$ and (ii) $\mathbb{P}[n^{-1} DM_{n,h} > c] \rightarrow 1$ for some $c > 0$ if H_0 does not hold.*

The process $(\mathbb{G}_n \phi_{n,h})' \Lambda_{n,h} (\mathbb{G}_n \phi_{n,h})$ is asymptotically tight and for each h behaves asymptotically as a weighted sum of $p - \bar{r}$ independent chi-squares, where the weights λ_h are the positive eigenvalues of $\Lambda_h = \lim_{n \uparrow \infty} \Lambda_{n,h}$, see Johnson, Kotz, and Balakrishnan (1995). The distribution of our distance-metric statistic is thus in general non-pivotal. The usual $p - \bar{r}$ chi-square distribution reappears when we use an efficient estimator, that is for the optimal weighting matrix and h tending to zero. However the result obtained without imposing this restriction likely provides a more accurate approximation as it accounts for the bandwidth's influence. Determining critical values requires estimation of Λ_h , which rely on estimators of V_h and $\Delta_{h,h}$ such as the ones given in Section 2.3. In what follows, we shall propose another route based on the bootstrap.

3.2 Bootstrapping SMD

Application of standard bootstrap methods here would require to generate resamples that mimic the behavior of the data under the null hypothesis with the same values of X , but new observations of Y . This can be done easily in specific cases such as regression models, but may be difficult or infeasible, as for instance in a simultaneous nonlinear equations system where a reduced form may not be available. We here propose a simple general method that can circumvent these difficulties. Instead of resampling observations, we perturb the objective function and recompute our test statistic using the perturbed objective function

$$M_{n,h}^*(\theta) = \frac{1}{2n(n-1)} \sum_{1 \leq i \neq j \leq n} w_i w_j g'(Z_i, \theta) W_n^{-1/2}(X_i) W_n^{-1/2}(X_j) g(Z_j, \theta) K_{ij},$$

where $w_i, i = 1, \dots, n$, are n independent copies of a known positive random variable w with $\mathbb{E}(w) = \text{Var}(w) = 1$ and $\mathbb{E}w^4 < \infty$. To avoid repeated optimizations, one can use in practice iteration(s) of a numerical algorithm, such as Newton-Raphson, with the initial estimator $\tilde{\theta}_n$ as a starting point. In case of a non-differentiable objective function, one can rely on numerical approximations of derivatives.

Jin, Ying and Wei (2001), Bose and Chatterjee (2003), Chatterjee and Bose (2005), and Chen and Pouzo (2009) have studied such a weighted bootstrap technique in different contexts.

They do not investigate its use for testing and impose conditions that do not hold in our context. In what follows, we show that this method consistently approximates the distribution of $\tilde{\theta}_{n,h}$ and $M_{n,h}(\tilde{\theta}_{n,h})$ uniformly in the bandwidth.

Theorem 3.2. *Under the Assumptions of Theorem 2.2, then conditionally on the sample and uniformly over $h \in \mathcal{H}_n$*

- i. $\sqrt{n} \left(\tilde{\theta}_{n,h}^* - \tilde{\theta}_{n,h} \right)$ has asymptotically the same distribution as $\sqrt{n} \left(\tilde{\theta}_{n,h} - \theta_0 \right)$, that is

$$\sup_{h \in \mathcal{H}_n} \sup_u \left| \mathbb{P} \left[\sqrt{n} \left(\tilde{\theta}_{n,h}^* - \tilde{\theta}_{n,h} \right) \leq u \mid Z_1, \dots, Z_n \right] - \mathbb{P} \left[\sqrt{n} \left(\tilde{\theta}_{n,h} - \theta_0 \right) \leq u \right] \right| = o_p(1).$$
- ii. $n \left(M_{n,h}^*(\tilde{\theta}_{n,h}^*) - M_{n,h}^*(\tilde{\theta}_{n,h}) \right)$ has asymptotically the same distribution as

$$n \left(M_{n,h}(\tilde{\theta}_{n,h}) - M_{n,h}(\theta_0) \right).$$

An heuristic for this result is as follows. Since $\mathbb{E} \left(M_{n,h}^*(\theta) \mid Z_1, \dots, Z_n \right) = M_{n,h}(\theta)$ is minimized at $\tilde{\theta}_{n,h}$, $\tilde{\theta}_{n,h}^*$ is expected to tend to $\tilde{\theta}_{n,h}$ conditionally on the sample. Now, as shown in the proofs, the perturbed and the original function have a similar quadratic expansion in θ . Therefore, the distribution of $n \left(M_{n,h}^*(\tilde{\theta}_{n,h}^*) - M_{n,h}^*(\tilde{\theta}_{n,h}) \right)$ is close to the one of $n \left(M_{n,h}(\tilde{\theta}_{n,h}) - M_{n,h}(\theta_0) \right)$, and similarly for $\sqrt{n} \left(\tilde{\theta}_{n,h}^* - \tilde{\theta}_{n,h} \right)$ and $\sqrt{n} \left(\tilde{\theta}_{n,h} - \theta_0 \right)$. Our result allows the use of the weighted bootstrap to approximate the distribution of $\tilde{\theta}_{n,h}$ and to determine confidence intervals. It is likely that a studentized version would yield more accurate results, but such an investigation is beyond the scope of this paper. As confidence regions are sets of values that are not rejected by a test, the following procedure can also be used to construct such regions.

The determination of bootstrap critical values for hypothesis testing is based on Theorem 3.2-(ii). Consider the decomposition

$$\begin{aligned} DM_{n,h} &= 2n \left[M_{n,h} \left(\tilde{\theta}_{n,h}^R \right) - M_{n,h} \left(R(\gamma_0) \right) - \left(M_{n,h}(\tilde{\theta}_{n,h}) - M_{n,h}(\theta_0) \right) \right] \\ &\quad + 2n \left[M_{n,h} \left(R(\gamma_0) \right) - M_{n,h}(\theta_0) \right]. \end{aligned}$$

The distribution of $DM_{n,h}$ under H_0 is determined by the first term, while consistency is ensured because the last term diverges under the alternative. Hence to approximate the behavior of the statistic under H_0 , we need to approximate the first term only. We thus repeat estimation under the constraint (3.8), which yields $\tilde{\theta}_{n,h}^{R*} = \arg \min_{\theta, \theta = R(\gamma)} M_{n,h}^*(\theta)$, and we define the bootstrap test statistic as

$$DM_{n,h}^* = 2n \left[M_{n,h}^*(\tilde{\theta}_{n,h}^{R*}) - M_{n,h}^*(\tilde{\theta}_{n,h}^R) - \left(M_{n,h}^*(\tilde{\theta}_{n,h}^*) - M_{n,h}^*(\tilde{\theta}_{n,h}) \right) \right].$$

Theorem 3.3. *Under the Assumptions of Theorem 2.2, then conditionally on the sample and uniformly over $h \in \mathcal{H}_n$ (i) $DM_{n,h}^*$ has asymptotically the same distribution as $DM_{n,h}$ under H_0 , and (ii) $DM_{n,h}^* = o_p(n)$ when H_0 does not hold.*

The last part suffices to obtain a consistent test, since $DM_{n,h}$ diverges at rate n . One could also use Theorem 2.3 to show that DM_n^* is bounded in probability whether H_0 holds or not, and thus that the bootstrap test has local power similar to the asymptotic one.

4 Small sample study

We investigated the comparative small sample properties of our estimator in three different setups. First, we consider the linear structural model

$$\begin{aligned} Y_1 &= \alpha_0 + Y_2\beta_0 + s(X)\varepsilon, \\ Y_2 &= 2X + U, \end{aligned} \tag{4.9}$$

where Y_2 is univariate and X follows a standard univariate normal distribution. We set $\alpha_0 = \beta_0 = 0$ and considered an homoscedastic and an heteroscedastic model with $s(x) = \sqrt{(1+x^2)/2}$. In both cases, $(s(X_i)\varepsilon_i, U_i)$ has mean zero, unit unconditional variances, and unconditional correlation 0.5. We computed our estimator with normal $K(\cdot)$ and $h = 1$, and its efficient version with $h = cn^{-1/5}$ where c varies. Reported results consider $c = 2/3, 1, 4/3$ and are based on 5000 replications. We only report results for $n = 100$, since varying the sample size did not affect our main findings. We compare them to the two-stage least squares estimator using the intercept and X as instrument as well as the feasible efficient IV estimator based on kernel estimation of optimal instruments with normal kernel and $h = cn^{-1/5}$. Our estimator performs similarly to TSLS in the homoscedastic case and is better than feasible efficient IV in the heteroscedastic case. The latter is not well centered and much more variable.

The second set of experiments aimed at evaluating the small sample behavior of our methods within the setup considered by Dominguez and Lobato (2004, hereafter DL), where

$$Y = \theta_0^2 X + \theta_0 X^2 + \varepsilon, \tag{4.10}$$

with $\theta_0 = 5/4$, $X \sim N(\mu, 1)$, and $\varepsilon \sim N(0, 1)$ independently of X . We considered SMD with a Gaussian kernel and $W_n = I$ and we made the bandwidth varying to investigate its influence. As a benchmark, we computed the Nonlinear Least-Squares (NLS) estimator, which

is efficient given that the error term is homoscedastic. We also computed DL's estimator, which minimizes

$$\frac{1}{n^3} \sum_{k=1}^n \left[\sum_{i=1}^n g(Z_i, \theta) \mathbb{I}(X_i \leq X_k) \right]^2 = \frac{1}{n^2} \sum_{i,j=1}^n g(Z_i, \theta) g(Z_j, \theta) \left[\frac{1}{n} \sum_{k=1}^n \mathbb{I}(X_i \leq X_k) \mathbb{I}(X_j \leq X_k) \right],$$

with $g(Z, \theta) = Y - \theta^2 X - \theta X^2$. All results are based on 5000 replications. Figures 1 and 2 compare the densities of the different estimators centered at $\theta_0 = 5/4$ and scaled by \sqrt{n} . They illustrate that the asymptotic normal approximation is pretty accurate. Our SMD estimator outperforms DL's one for the range of considered bandwidths, though unreported simulations show that this ranking may be reversed for a very large bandwidth. SMD also compares well to NLS, though we made not attempt to adapt the bandwidth to the sample size.

To investigate the behavior of our bootstrap test, we used the two-point distribution defined through $\Pr(w = \frac{3-\sqrt{5}}{2}) = \frac{5+\sqrt{5}}{10}$ and $\Pr(w = \frac{3+\sqrt{5}}{2}) = \frac{5-\sqrt{5}}{10}$. We chose this simple distribution with third central moment equal to one in the hope to better approximate the distribution of the statistic, as is the case in simpler setups, see e.g. Mammen (1992). In Table 1, we report empirical rejections for our test, where 199 bootstrap statistics were used for each replication. Our Theorem 3.1 implies that, in this simple setup with only one parameter, $DM_{n,h}/\lambda_h$ is asymptotically χ_1^2 where λ_h is a real number. Hence we estimated λ_h to derive an asymptotic test. Our results indicate that the asymptotic test has empirical levels close to the nominal ones, but often slightly overrejects. The bootstrap test has in most cases empirical levels closer to the nominal ones. Figures 3 and 4 report the empirical rejection of our bootstrap tests and the Wald-type tests based on NLS and DL's estimators when testing $H_0 : \theta_0 = \theta_{00}$ for varying θ_{00} . Our test's power is a bit smaller than NLS but larger than DL, as could be expected from the estimators' densities.

The third set of experiments focuses on the efficient SMD in the setup of Cragg (1983), Newey (1993), and Kitamura, Tripathi and Ahn (2004, hereafter KTA), where

$$Y = \beta_1 + \beta_2 X + \varepsilon, \quad \mathbb{E}(\varepsilon|X) = 0, \quad \text{Var}(\varepsilon|X) = .1 + .2X + .3X^2, \quad (4.11)$$

with $\beta_1 = \beta_2 = 1$, $\ln X \sim N(0, 1)$, and ε is normally distributed. KTA concluded that in this setup their Smoothed Empirical Likelihood (SEL) works best than the two-step optimal GMM estimator, so we didn't repeat this comparison. As a benchmark, we considered the generalized least squares (GLS) estimator based on the true variance function and we also computed the feasible GLS estimator based on the knowledge of the variance functional form. We considered the efficient SMD and we adapt the bandwidth to the sample size as in KTA.

At the request of a referee, we also investigated the influence of b , the bandwidth used in the optimal weighting matrix estimation, on the final estimator. As an illustration, we report results for the case where $h = b$, simply labelled SMD, and the case where b is fixed at $n^{-1/5}/3$, labelled SMDb. Results for SMD are based on 5000 replications, while results for SEL are based on 500 replications as reported by KTA. For each estimator we computed the ratios of root mean squared error (RMSE) and mean absolute deviation (MAE) with respect to the ones of GLS. Since considering either made little difference, we focus on the former. Figure 5 reports the RMSE as a function of the bandwidth on the grid $n^{-1/5} \times 1/3$ through $8/3$ for SMD and SEL.⁸ Both SEL and efficient SMD performs well compared to the feasible GLS, though the latter relies on the parametric form of the variance. Both perform better with increasing sample size, but their relative performances depend on the bandwidth h . The shape of RMSE with respect to the bandwidth is strikingly different for the two estimators. For SEL, RMSE of both parameters is smaller for pretty large bandwidths. For instance, when $n = 100$, the RMSE minimizing bandwidth is 0.93, to be compared with the interquartile range of X , which is 1.45. Moreover, the RMSE-minimizing bandwidth does not seem to decrease with the sample size. For SMD, RMSE of the intercept is always minimum at the smallest considered bandwidth, while for the slope the RMSE-minimizing bandwidth is small, about 0.27 for $n = 100$, and decreases with the sample size. The choice of b has little influence.

We also investigated the behavior of our bootstrap distance-metric statistic under the null hypothesis. We focused on small to medium bandwidths, that is $h = cn^{-1/5}$ with $c = 1/3, 2/3, 4/3$, $h = b$, and ran 5000 replications with 199 bootstrap samples each. Results are reported in Table 2. Using asymptotic critical values from the chi-square distribution with one degree of freedom for our test yields rejection percentages that are higher than nominal values and our bootstrap corrects for overrejection. Tests based on FGLS (Wald and LR tests yield identical results) are severely oversized and are then not reliable.

To sum up, our SMD estimator performs well in our simulation experiments: it is competitive with alternative methods, while it exhibits a different behavior with respect to the bandwidth than SEL. Our bootstrap technique yields reliable test levels for moderate sample sizes.

⁸RMSE figures for SEL were kindly provided by Yuichi Kitamura.

5 Assumptions

5.1 General Assumptions

Assumption 1. (i) The parameter space Θ is compact. (ii) θ_0 is the unique value in Θ satisfying (1.1), that is $\mathbb{E}[g(Z, \theta)|X] = 0$ a.s. $\Rightarrow \theta = \theta_0$. (iii) θ_0 belongs to the interior of Θ .

Definition 1. An envelope for a class \mathcal{F} of functions is any function F such that $|f| \leq F \forall f \in \mathcal{F}$. A class \mathcal{F} of functions is Euclidean for an envelope F if there exists positive constants A and V such that for any $0 < \varepsilon \leq 1$ and any measure μ such that $\int F d\mu < \infty$, there are functions f_1, \dots, f_k in \mathcal{F} such that (i) $k \leq A\varepsilon^{-V}$ and (ii) for any f in \mathcal{F} , there is an f_i such that $\int |f - f_i| d\mu \leq \varepsilon \int F d\mu$.

We refer to Nolan and Pollard (1987), Pakes and Pollard (1989), and Sherman (1994) for more details on Euclidean families.

Assumption 2. (i) $K(\cdot)$ is a symmetric, bounded function, with integral equal to one and strictly positive Fourier transform on \mathbb{R}^q . (ii) The class of all functions $(x, \bar{x}) \mapsto K((x - \bar{x})/h)$, $x, \bar{x} \in \mathbb{R}^q$, $h > 0$, is Euclidean for a constant envelope.

Assumption 2-(i) implies that the Fourier transform of $K(\cdot)$ belongs to $L^1(\mathbb{R}^q) \cap L^2(\mathbb{R}^q)$. The Fourier transform of $K(\cdot)$ is formally defined as

$$\mathcal{F}[K](t) = (2\pi)^{-q/2} \int \exp^{-it'u} K(u) du.$$

Assumption 2-(ii) is also needed when studying the uniform in bandwidth properties of kernel-type estimators, see the definition of “regular kernels” in Einmahl and Mason (2005).

Assumption 3. For all n , $W_n(\cdot)$ is a $r \times r$ symmetric p.d. non-random matrix function with $0 < \inf_n \inf_u \lambda_{\min}(W_n(u)) \leq \sup_n \sup_u \lambda_{\max}(W_n(u)) < \infty$. There exists a symmetric p.d. matrix function $W(\cdot)$ such that $W_n(u) - W(u) = o(1)$ for all u in the interior of the support of X .

Assumption 3 ensures that $W_n^{-1/2}(\cdot)$ is well-defined and the spectral radius of $W_n^{-1/2}(\cdot)$ is uniformly bounded. It implies that $0 < \inf_u \lambda_{\min}(W(u)) \leq \sup_u \lambda_{\max}(W(u)) < \infty$.

Assumption 4. (i) The function $\sup_{\theta} \|\mathbb{E}[g(Z, \theta) | X = \cdot]\| f(\cdot)$ is in $L^1 \cap L^2$. For all x , the map $\theta \mapsto \mathbb{E}[g(Z, \theta) | X = x]$ is continuous. (ii) The families $\mathcal{G}_k = \{g^{(k)}(\cdot, \theta) : \theta \in \Theta\}$, $1 \leq k \leq r$, are Euclidean for an envelope G with $\mathbb{E}G^2 < \infty$. (iii) $\mathbb{E}G^4 < \infty$. (iv) There exists a neighborhood of θ_0 and a constant $c > 0$ such that for all θ in that neighborhood, $\mathbb{E}\|g(Z, \theta) - g(Z, \theta_0)\|^2 \leq c\|\theta - \theta_0\|$. (v) The components of $\nabla_{\theta}\tau(\cdot, \theta_0)f(\cdot)$ are in $L^1 \cap L^2$. (vi) The components of $\text{Var}[g(Z, \theta_0)|X = \cdot]f(\cdot)$ are in $L^1 \cap L^2$.

Assumption 4 as a whole does not require the continuity of the functions $\theta \mapsto g(z, \theta)$. Assumption 4-(i) ensures that $\mathbb{E}M_{n,h}(\theta)$ is continuous as a function of θ and h . Assumptions 2-(ii), 4-(ii), and the good behavior of the spectral radius of $W_n^{-1/2}(\cdot)$ guarantee that the family of functions

$$\{(z, \bar{z}) \mapsto g'(z, \theta)W_n^{-1/2}(x)W_n^{-1/2}(\bar{x})g(\bar{z}, \theta)K((x - \bar{x})/h) : \theta \in \Theta, h > 0\}$$

is uniformly Euclidean for a squared integrable envelope from Pakes and Pollard (1989, Lemma 2.14-(ii)).

Assumption 5. (i) For any x , all second partial derivatives of $\tau(x, \cdot) = \mathbb{E}[g(Z, \cdot) | X = x]$ exist on a neighborhood \mathcal{N} of θ_0 independent on x . (ii) There exists a real-valued function $H(\cdot)$ with $\mathbb{E}H^4 < \infty$ and some $a \in (0, 1)$ such that

$$\|\mathbb{H}_{\theta, \theta} \tau^{(k)}(X, \theta) - \mathbb{H}_{\theta, \theta} \tau^{(k)}(X, \theta_0)\| \leq H(X) \|\theta - \theta_0\|^a \quad \forall \theta \in \mathcal{N} \quad k = 1, \dots, r.$$

Assumption 5 is implied by the following Condition 2.

Condition 2. (i) For all z , all second partial derivatives of $g(z, \cdot)$ exist on a neighborhood \mathcal{N} of θ_0 independent on z . (ii) There exists a real-valued function $\tilde{H}(\cdot)$ with $\mathbb{E}\tilde{H}^4 < \infty$ and $a \in (0, 1]$ such that

$$\|\mathbb{H}_{\theta, \theta} g^{(k)}(Z, \theta) - \mathbb{H}_{\theta, \theta} g^{(k)}(Z, \theta_0)\| \leq \tilde{H}(Z) \|\theta - \theta_0\|^a \quad \forall \theta \in \mathcal{N} \quad k = 1, \dots, r.$$

Under Condition 2, $\mathbb{E}\|g(Z, \theta) - g(Z, \theta_0)\|^2 = O(\|\theta - \theta_0\|^2)$, so Assumption 4-(iv) is not restrictive. For our general results, we do not require differentiability of $g(x, \theta)$ and we impose only 4-(iv), which is precisely what is needed in conditional quantile restriction models where Condition 2 fails, see e.g. Zheng (1998, Equation A.11). By Assumption 3, $g_n(Z, \theta) = W_n^{-1/2}(X)g(Z, \theta)$ also satisfies Assumption 4-(iv), and $\tau_n(X, \theta) = W_n^{-1/2}(X)\tau(X, \theta)$ inherits the smoothness properties of $\tau(X, \theta)$.

Assumption 6. (i) The components of $\nabla_{\theta} \tau_n(\cdot, \theta_0)f(\cdot)$ satisfy Condition E with kernel $K(\cdot)$ for an envelope Φ_1 with $\mathbb{E}\Phi_1^a < \infty$ for some $a \geq 4$. (ii) The components of $\mathbb{H}_{\theta, \theta} \tau_n^{(k)}(\cdot, \theta_0)f(\cdot)$, $1 \leq k \leq r$ and $H(\cdot)f(\cdot)$ satisfy Condition E with kernel $|K(\cdot)|$ for an envelope Φ_2 with $\mathbb{E}\Phi_2^a < \infty$ for some $a \geq 4/3$.

5.2 Assumptions for Theorem 2.3

Each of the above Assumptions M4 to M6 replaces or completes Assumptions 4 to 6.

Assumption M4. (i) Each $\bar{\theta}_{n,h}$ is unique and there exists a subset Θ_M of the interior of Θ such that for each n, h there is a ball $B(\bar{\theta}_{n,h}, r)$ in Θ_M with r independent of n and h . (ii) There exists a constant $c > 0$ such that for all $\theta \in \Theta_M$, $\mathbb{E}\|g(Z, \theta_1) - g(Z, \theta_2)\|^2 \leq c\|\theta_1 - \theta_2\|$. (iii)

The components of $\nabla_{\theta}\tau(\cdot, \theta_1)f(\cdot)$ and of $\mathbb{E}[g(Z, \theta_1)g'(Z, \theta_2)|X = \cdot]f(\cdot)$, $\theta_1, \theta_2 \in \Theta_M$, are uniformly bounded in $L^1 \cap L^2$. (iv) The components of $\mathbb{E}[g(Z, \theta_1)g'(Z, \theta_2)|X = \cdot]$ are continuous in $\theta_1, \theta_2 \in \Theta_M$.

Assumption M5. (i) For any x , all second partial derivatives of $\tau(x, \cdot) = \mathbb{E}[g(Z, \cdot)|X = x]$ exist on Θ_M . (ii) There exists a real-valued function $H(\cdot)$ with $\mathbb{E}H^4 < \infty$ and some $a \in (0, 1]$ such that

$$\|\mathbb{H}_{\theta, \theta}\tau^{(k)}(X, \theta_1) - \mathbb{H}_{\theta, \theta}\tau^{(k)}(X, \theta_2)\| \leq H(X)\|\theta_1 - \theta_2\|^a \quad \forall \theta_1, \theta_2 \in \Theta_M \quad k = 1, \dots, r.$$

Assumption M6. (i) The components of $\nabla_{\theta}\tau_n(\cdot, \cdot)f(\cdot)$ satisfy Condition ME with kernel $K(\cdot)$ for an envelope Φ_1 with $\mathbb{E}\Phi_1^a < \infty$ for some $a \geq 4$. (ii) The components of $\mathbb{H}_{\theta, \theta}\tau_n^{(k)}(\cdot, \cdot)f(\cdot)$, $1 \leq k \leq r$, and $H(\cdot)f(\cdot)$ satisfy Condition ME with kernel $|K(\cdot)|$ for an envelope Φ_2 with $\mathbb{E}\Phi_2^a < \infty$ for some $a \geq 4/3$.

5.3 Assumptions for Theorem 2.4

Each of the above Assumptions E2, E4, and E6 replaces or completes Assumptions 2, 4, and 6.

Assumption E2. (i) $L(\cdot)$ is a density of bounded variation with bounded support and is strictly positive around the origin. (ii) Assumption 2-(ii) holds for $L(\cdot)$.

Assumption E4. Assumption 4 holds with $\sup_{x \in \mathbb{R}^q} \mathbb{E}[G^8(Z) | X = x] < \infty$.

Assumption E7. (i) $f(\cdot)$ is bounded away from zero and infinity with bounded support D that can be written as finite unions and/or intersections of sets $\{x : p(x) \geq 0\}$, where $p(\cdot)$ is a polynomial function. (ii) $W(\cdot) = \mathbb{E}[g(Z, \theta_0)g'(Z, \theta_0) | X = \cdot]f(\cdot)$ is such that $0 < \inf_u \lambda_{\min}(W(u)) \leq \sup_u \lambda_{\max}(W(u)) < \infty$. (iii) $W(\cdot)$ is Hölder continuous on D . (iv) Let $\omega^2(\cdot, \theta) = \mathbb{E}[g(Z, \theta)g'(Z, \theta) | X = \cdot]$. For θ in a neighborhood of θ_0 , some $\nu > 2/3$, and $c > 0$, $\|\omega^2(x, \theta) - \omega^2(x, \theta_0)\| \leq c\|\theta - \theta_0\|^\nu$ for all x .

Parts (ii) and (iii) ensure that Assumption 3 holds in probability for $W_n(\cdot) = \mathbb{E}[\widehat{W}_n(\cdot, \theta_0)]$ and that its entries as indexed by b are Euclidean for a constant envelope. Part (iv) allows to control the bias of $\widehat{W}_n(\cdot, \check{\theta})$.

Assumption E6. Each of the entries of $\nabla_{\theta}\tau(\cdot, \theta_0)f(\cdot)$, $\mathbb{H}_{\theta, \theta}\tau^{(k)}(\cdot, \theta_0)f(\cdot)$, $1 \leq k \leq r$ and $H(\cdot)f(\cdot)$ is Hölder continuous on D , with possibly different exponents.

REFERENCES

- ADAMS, R.A., AND FOURNIER, J.J.F (2003). *Sobolev Spaces (2nd edition)*. Academic Press, Elsevier Science: Oxford.
- AI, C., AND CHEN, X. (2003). Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions. *Econometrica* **71(6)**, 1795-1843.
- ANDREWS, D.K.W. (1994). Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica* **62(1)**, 43-72.
- ANTOINE, B., BONNAL, H., AND RENAULT, E. (2007). On the efficient use of the informational content of estimating equations: Implied probabilities and Euclidean empirical likelihood. *J. Econometrics* **138(2)**, 461-487.
- BIERENS, H.J. (1982). Consistent model specification tests. *J. Econometrics* **20(1)**, 105-134.
- BOSE, A., AND CHATTERJEE, S. (2003). Generalized bootstrap for estimators of minimizers of convex functions. *J. Statist. Plann. Inference* **117(2)**, 225-239.
- CARRASCO, M., AND FLORENS, J.P. (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory* **16(6)**, 797-834.
- CHAMBERLAIN, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *J. Econometrics* **34(3)**, 305-334.
- CHATTERJEE, S., AND BOSE, A. (2005). Generalized bootstrap for estimating equations. *Ann. Statist.* **33(1)**, 414-436.
- CHEN, X., LINTON, O., AND VAN KEILEGOM, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* **71(5)**, 1591-1608.
- CHEN, X. AND POUZO, D. (2009). Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *Journal of Econometrics* **152(1)**, 46-60.
- CRAGG, J. (1983). More efficient estimation in the presence of heteroscedasticity of unknown form. *Econometrica* **51(3)**, 751-764.
- CRISTOBAL CRISTOBAL, J.A., FARALDO ROCA, P., AND GONZALEZ MANTEIGA, W. (1987). A Class of Linear Regression Parameter Estimators Constructed by Nonparametric Estimation. *Ann. Statist.* **15(2)**, 603-609.
- DOMINGUEZ, M.A., AND LOBATO, I.N. (2004). Consistent estimation of models defined by conditional moment restrictions. *Econometrica* **72(5)**, 1601-1615.
- DONALD, S.G., IMBENS, G.W., AND NEWEY, W.K. (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. *J. Econometrics* **117(1)**, 55-93.
- EINMAHL, U., AND MASON, D.M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* **33(3)**, 1380-1403.

- FAN, Y., AND LI, Q. (2000). Consistent model specification tests: nonparametric versus Bierens' tests. *Econometric Theory* **16**, 1016-1041.
- HANSEN, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50(4)**, 1029-1054.
- HANSEN, L.P., AND SINGLETON, K.J. (1982). Generalized instrumental variable estimation of nonlinear rational expectations models. *Econometrica* **50(4)**, 1269-1286.
- HJORT, N.L., MCKEAGUE, I.W., AND VAN KEILEGOM, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37(3)** 1079-1111.
- HORN, E.A., AND JOHNSON, C.R. (1991). *Topics in Matrix Analysis*. Cambridge University Press: New-York.
- HURST, S. (1995). The characteristic function of the Student t distribution. Statistics Research Report SRR044-85, Center for Financial Mathematics, Canberra.
- JIN, Z., YING, Z., AND WEI, L.J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88(2)**, 381-390.
- JOHNSON, N.L., KOTZ, S., AND BALAKRISHNAN, N. (1995). *Continuous Univariate Distributions*, 2nd edition. Wiley: New-York.
- KITAMURA, Y., TRIPATHI, G., AND AHN, H. (2004). Empirical likelihood-based inference in conditional moment restriction models. *Econometrica* **72(6)**, 1667-1714.
- KOMUNJER, I., AND VUONG, Q. (2010). Efficient estimation in dynamic conditional quantile models *J. Econometrics* **157(2)**, 272-85.
- LE GRUYER, E., AND ARCHER, J.C. (1998). Harmonious extensions. *SIAM J. Math. Anal.* **29(1)**, 279-292.
- MALLIAVIN, P. (1995). *Integration and Probability*. Springer-Verlag: New-York.
- MAMMEN, E. (1992). *When does bootstrap work? Asymptotic results and simulations*, Lecture Notes in Statistics, 77. Springer-Verlag: New-York.
- MC SHANE E.J. (1934). Extension of range of functions. *Bull. Amer. Math. Soc.* **40**, 837-842.
- NEWBY, W.K. (1993). Efficient estimation of models with conditional moment restrictions. *Handbook of Statistics* vol. 11, G.S. Maddala, C.R. Rao and H.D. Vinod eds, 419-454.
- NEWBY, W.K., AND MCFADDEN, D.L. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics* vol. 4, R.F. Engle and D.L. McFadden eds, 2111-2245.
- NOLAN, D., AND POLLARD, D. (1987). *U*-processes : Rates of convergence. *Ann. Statist.* **15(2)**, 780-799.
- OTSU, T. (2008). Conditional empirical likelihood estimation and inference for quantile regression models. *J. Econometrics* **142(2)**, 508-538.
- PAKES, A., AND POLLARD, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57(5)**, 1027-1057.

- QIN, J. AND LAWLESS, J. (1994). Empirical likelihood and generalized estimating equations. *Ann. Statist.* **22(1)**, 300-325.
- SCHENNACH, S.M. (2007). Point estimation with exponentially tilted empirical likelihood. *Ann. Statist.* **35(2)**, 634-672.
- SHERMAN, R.P. (1994). Maximal inequalities for degenerate U -processes with applications to optimization estimators. *Ann. Statist.* **22(1)**, 439-459.
- SMITH, R.J. (2007a). Efficient information theoretic inference for conditional moment restrictions. *J. Econometrics* **138(2)**, 430-460.
- SMITH, R.J. (2007b). Local GEL estimation with conditional moment restrictions. Chapter 4 in *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*, eds. G.D.A. Phillips and E. Tzavalis, 100-122. Cambridge University Press: Cambridge.
- VAN DER VAART, A.W. (1998). *Asymptotic Statistics*. Cambridge University Press: New-York.
- VAN DER VAART, A.W., AND WELLNER, J.A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag: New-York.
- WHITE, H. (1981). Consequences and detection of misspecified nonlinear regression models. *J. Amer. Statist. Assoc.* **76 (374)**, pp. 419-433.
- WHITE, H. (1994): *Estimation, Inference and Specification Analysis*. New-York: Cambridge University Press.
- ZHENG, J.X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75**, 263-289.
- ZHENG, J.X. (1998). A consistent nonparametric test of parametric regression models under conditional quantile restriction. *Econometric Theory* **14(1)**, 123-138.

Appendix A

Lemmas A.1 and A.2 provide different sets of general sufficient conditions that guarantee Condition (E) and are followed by some discussion. We note that since $\int \phi_n(x - uh)K(u)du$ is the expectation of a kernel estimator, those conditions are of independent interest.

Lemma A.1. *Assume that $K(\cdot)$ is integrable and its Fourier transform $\mathcal{F}[K](\cdot)$ is Hölder continuous with exponent a . If the sequence of functions $\phi_n : \mathbb{R}^q \rightarrow \mathbb{R}$, $n \geq 1$ have integrable envelope $\Phi(\cdot)$, they satisfy Condition (E) with kernel $K(\cdot)$ for an envelope $\Phi(\cdot) + C$, $C > 0$, whenever*

$$\sup_n \int \|t\|^a |\mathcal{F}[\phi_n](t)| dt < \infty. \tag{A.1}$$

Proof. For any ϕ_n , write

$$\begin{aligned} \int \phi_n(x - hu)K(u)du &= (2\pi)^{-q/2} \int \int \phi_n(v) \exp(it'(x - v))\mathcal{F}[K](ht)dvdt \\ &= \int \mathcal{F}[\phi_n](t) \exp(it'x)\mathcal{F}[K](ht)dt, \end{aligned}$$

for almost any x , and note that the equality holds trivially for $h = 0$. Hence for any $h_1, h_2 \in [0, h_0]$, using $|\mathcal{F}[K](t_1) - \mathcal{F}[K](t_2)| \leq c\|t_1 - t_2\|^a$,

$$\begin{aligned} \left| \int \phi_n(x - h_1 u) K(u) du - \int \phi_n(x - h_2 u) K(u) du \right| &\leq \int |\mathcal{F}[\phi_n](t)| |\mathcal{F}[K](h_1 t) - \mathcal{F}[K](h_2 t)| dt \\ &\leq c|h_1 - h_2|^a \int \|t\|^a |\mathcal{F}[\phi_n](t)| dt. \end{aligned}$$

Use Lemma 2.13 of Pakes and Pollard (1989) to conclude. \square

As most common kernels have bounded moment of order 1, the Hölder continuity of $\mathcal{F}[K](\cdot)$ is satisfied with $a = 1$, so we assume this from now on without much loss of generality. Condition (A.1) is fulfilled when $\phi_n(\cdot)$ belongs to $W^{m,1}$, the subspace of functions of L^1 such that their weak partial derivatives belongs to L^1 up to integer order $m \geq 3$, see e.g. Malliavin (1995, Section III.3). Another possible space is the Sobolev space of functions H^s . Indeed,

$$\int \|t\| |\mathcal{F}[\phi_n](t)| dt \leq \int_{\|t\| \leq 1} |\mathcal{F}[\phi_n](t)| dt + \int_{\|t\| > 1} \|t\| |\mathcal{F}[\phi_n](t)| dt \leq \int \Phi(x) dx + I_2.$$

By Cauchy-Schwarz inequality, for any $b > 1$

$$I_2 \leq \left[\int \left(1 + \|t\|^2\right)^{1+b/2} |\mathcal{F}[\phi_n](t)|^2 dt \right]^{1/2} \left[\int_{\|t\| > 1} \|t\|^{-b} dt \right]^{1/2}.$$

Condition (A.1) then holds for a sequence $\phi_n(\cdot)$ from the Sobolev space of functions H^s with $s > 3/2$ endowed with the norm

$$\|\phi\|_{H^s}^2 = \int_{\mathbb{R}^d} \left(1 + \|t\|^2\right)^s |\mathcal{F}[\phi](t)|^2 dt.$$

For any integer $s \geq 1$, H^s is isomorph to $W^{s,2}$ endowed with the norm $\|\phi\|_{W^{s,2}}^2 = \sum_{0 \leq |\alpha| \leq s} \|D^\alpha \phi\|_{L^2}^2$, where for a multi-index $\alpha = (\alpha_1, \dots, \alpha_q)$ of degree $|\alpha| = \alpha_1 + \dots + \alpha_q$, $D^\alpha \phi$ denotes the weak partial derivative of ϕ , see Malliavin (1995, Section III.3) or Adams and Fournier (2003, Chapter 3). Finally, note that if two sequences of functions belongs to $W^{m,2}$ with $m \geq 3$, their product belongs to $W^{m,1}$ and thus also fulfills Condition (E).

Lemma A.2. *For $K(\cdot)$ integrable, any of the following conditions ensures that Condition (E) holds for a constant envelope.*

- i. $\phi_n(x) = \psi_n(p(x))$, where $p(x)$ is a polynomial in q variables and $\psi_n(\cdot)$ is a uniformly bounded sequence of functions of bounded variation on \mathbb{R} .
- ii. The $\phi_n(\cdot)$ are uniformly bounded and Hölder continuous with exponent a , and $\int \|u\|^a |K(u)| du < \infty$.
- iii. The functions ϕ_n are finite addition, multiplication, min, or max of functions satisfying one of (i) or (ii) (for finite multiplication under (ii), assume that $K(\cdot)$ has enough finite moments).

Proof. The proof follows by showing in each case that $\{(x, u) \mapsto \phi_n(x - hu) : h \in [0, 1]\}$ is Euclidean for a constant envelope and using that the Euclidean property is preserved by integration with respect to a finite measure, see Nolan and Pollard (1987, Lemma 20).

(i) For each n , the class of subgraphs $\{(x, u) \mapsto \text{subgraph } (\phi_n(x - uh)) : h \in [0, 1]\}$ is a VC class of sets by the arguments of Lemma 22 of Nolan and Pollard (1987). A careful inspection of their proof shows that the index of this class of subgraphs is independent on n provided the functions ϕ_n are uniformly bounded, and the class of functions is thus Euclidean.

(ii) As for all n , $|\phi_n(x_1) - \phi_n(x_2)| \leq c\|x_1 - x_2\|^a$ for some $c > 0$, $|\phi_n(x - uh_1) - \phi_n(x - uh_2)| \leq c\|u\|^a|h_1 - h_2|^a$. Lemma 2.13 of Pakes and Pollard (1989) thus implies that the class of $\phi_n(x - hu)$ as functions of (x, u) is Euclidean for an envelope $C_1 + C_2\|u\|^a$ for some $C_1, C_2 > 0$.

(iii) From the above proofs, each of the class of functions $\phi_n(x, u; h) = \phi_n(x - hu)$ as functions of (x, u) is Euclidean for a constant envelope in Case (i), for an integrable envelope in Case (ii). From Lemma 2.14 of Pakes and Pollard (1989), finite additions, multiplications, maximum, and minimum, of functions in such families are Euclidean with an envelope deduced by similar operations on the envelopes of each family. \square

Since the indicator function $\mathbb{I}(u \geq 0)$ is of bounded variation on \mathbb{R} , Lemma A.2-(i) implies that Condition (E) is satisfied when $\phi_n(\cdot) = \phi(\cdot) = \mathbb{I}(p(x) \geq 0)$ for any polynomial $p(x)$. Hence, $\phi(\cdot)$ can be the indicator function of a half space, a ball, a rectangle, or finite unions and intersections of such subsets of \mathbb{R}^q . Now, if the $\phi_n(\cdot)$ have a common fixed bounded support (and vanish outside this set) and the Hölder continuity condition in Lemma A.2-(ii) holds on this support, then $\phi_n(\cdot)$ can always be written as the product of the indicator function of the support and a Hölder continuous extension of $\phi_n(\cdot)$ to the whole space \mathbb{R}^q , which exists by the McShane-Whitney theorem, see McShane (1934). Lemma A.2-(iii) then ensures that the $\phi_n(\cdot)$ satisfy Condition (E).

Appendix B

The following lemmas allow to show that our Assumption E7-(ii) and (iii) implies that $\mathbb{E}[\widehat{W}_n(x, \theta_0)]$ is Euclidean for a constant envelope.

Lemma B.1. *Let $\omega(x; b)$, $b \in [0, h_0]$, be positive definite $r \times r$ matrix-valued functions on \mathbb{R}^q with eigenvalues uniformly bounded away from zero and infinity. If $\{(x, u) \mapsto \omega(x - uh; b) : h, b \in [0, h_0]\}$ is Euclidean for a constant envelope, then so is $\{(x, u) \mapsto \omega^{-s}(x - uh; b) : h, b \in [0, h_0]\}$, $s = 1/2$ or 1 .*

Proof. We treat the case $s = 1/2$, the other case similarly follows. For any p.d. A and B , and the spectral matrix norm $\|\cdot\|_2$,

$$\left\|A^{1/2} - B^{1/2}\right\|_2 \leq \frac{1}{2} \left\{\max(\|A^{-1}\|_2, \|B^{-1}\|_2)\right\}^{1/2} \|A - B\|_2,$$

see Horn and Johnson (1991, page 557). Since $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$,

$$\begin{aligned} \|A^{-1} - B^{-1}\|_2 &\leq \|A^{-1}\|_2 \|B - A\|_2 \|B^{-1}\|_2 \\ \text{and } \left\|A^{-1/2} - B^{-1/2}\right\|_2 &\leq \frac{1}{2} \left\{\max(\|A\|_2, \|B\|_2)\right\}^{1/2} \|A^{-1}\|_2 \|B^{-1}\|_2 \|A - B\|_2. \end{aligned} \quad (\text{A.2})$$

From the upper and lower bounds of the eigenvalues of $\omega(x; b)$ and the equivalence between the Euclidean norm $\|\cdot\|$ and the spectral norm $\|\cdot\|_2$, deduce that for any h_i, b_i , $i = 1, 2$,

$$\|\omega^{-1/2}(x - uh_1; b_1) - \omega^{-1/2}(x - uh_2; b_2)\| \leq C\|\omega(x - uh_1; b_1) - \omega(x - uh_2; b_2)\|.$$

for some constant C . Finally, apply the definition of the Euclidean property. \square

In what follows, $\bar{\omega}(x; b) = \int_{\mathbb{R}^q} \omega(x - bv)L(v) dv$, D is a domain that can be written as $\{x : p(x) \geq 0\}$ for some real polynomial $p(x)$, or finite unions and/or intersections of such sets.

Lemma B.2. *If $\omega(x)$ has eigenvalues uniformly bounded away from zero and infinity on D and is Hölder continuous on D (i) $\bar{\omega}(x; b)$ has eigenvalues uniformly bounded away from zero and infinity on D if $L(\cdot)$ is strictly positive in a neighborhood of the origin; (ii) $\{(x, u) \mapsto \bar{\omega}(x - hu; b) : h, b \in [0, h_0]\}$ is Euclidean entrywise for a constant envelope.*

Proof. Part (i) is straightforward, Part (ii) is shown as follows. Since $\omega(x)$ is positive definite, there exists a unique lower triangular matrix $T(x)$ with positive diagonal entries such that $\omega(x) = T(x)T'(x)$. The eigenvalues of $\omega(\cdot)$ are uniformly bounded away from zero and infinity iff the same holds for the eigenvalues of $T(\cdot)$, that is its diagonal entries. Moreover, the entries of $T(\cdot)$ are Hölder continuous functions with exponent a since they obtain recursively from the entries of $\omega(\cdot)$ through the equations

$$T_{i,i}^2(x) = \omega_{i,i}(x) - \sum_{k=1}^{i-1} T_{i,k}^2(x), \quad T_{i,j}(x) = T_{j,j}^{-1}(x) \left(\omega_{i,j}(x) - \sum_{k=1}^{j-1} T_{i,k}(x)T_{j,k}(x) \right), \quad 1 \leq i \leq r, \quad i > j.$$

By Theorem 3.3 and Remark 3.4 of Le Gruyer and Archer (1998), each entry $T_{i,j}(x)$ can be extended to \mathbb{R}^q such that its extension is Hölder continuous with the same exponent and remains between $\inf_{x \in D} T_{i,j}(x)$ and $\sup_{x \in D} T_{i,j}(x)$. The lower triangular matrix extension $\tilde{T}(\cdot)$ yields an extension $\tilde{\omega}(\cdot) = \tilde{T}(\cdot)\tilde{T}'(\cdot)$ of $\omega(\cdot)$ on \mathbb{R}^q which is positive definite with eigenvalues uniformly bounded away from zero and infinity and Hölder continuous. By Lemma 2.13 of Pakes and Pollard (1989) and the fact that multiplication preserves Euclideanity, deduce that the class of functions $\{(x, u, v) \mapsto \tilde{\omega}(x - uh - vb)\mathbb{I}(x - hu - vb \in D) = \omega(x - uh - vb), x, u, v \in \mathbb{R}^q, h, b \in [0, h_0]\}$, is Euclidean for a constant envelope. The result follows since Euclideanity is preserved by integration. \square

The two above lemma can be combined to yield a result on $\bar{\omega}^{-1/2}(x - uh; b)$.

Lemma B.3. *$\{(x, u) \mapsto \bar{\omega}^{-s}(x - hu; b)\mathbb{I}(x - hu \in D) : h, b \in [0, h_0]\}$, $s = 1/2$ or 1 , is Euclidean for a constant envelope under the assumptions of Lemma B.2.*

Proof. Lemma A.2 and the fact that Euclideanity is preserved by addition yield that the class of functions $\{(x, u) \mapsto \tilde{\omega}(x - uh; b) = \mathbb{I}(x - hu \in D^c)\mathbf{I} + \bar{\omega}(x - hu; b) : h, b \in [0, h_0]\}$ is Euclidean for a constant envelope. By definition, the eigenvalues of $\tilde{\omega}(x - uh; b)$ stay away from zero and infinity and $\tilde{\omega}(x - uh; b) = \bar{\omega}(x - uh; b)$ whenever $x - uh \in D$.

By Lemma B.1, the class $\{(x, u) \mapsto \tilde{\omega}^{-1/2}(x - uh; b) : h, b \in [0, h_0]\}$ is then Euclidean for a constant envelope, and so is $\{(x, u) \mapsto \tilde{\omega}^{-1/2}(x - uh; b)\mathbb{I}(x - hu \in D) : h, b \in [0, h_0]\}$ by Lemma A.2-(i). A similar reasoning applies when $s = 1$. \square

Proofs

In what follows we adopt the notations of Sherman (1993, 1994) concerning U -statistics. Following his use, we say that for a sequence $\theta_{n,h}$, $H_n(\theta) = o_p(1)$, respectively $O_p(1)$, uniformly over $o_p(1)$ neighborhoods of $\theta_{n,h}$ and uniformly in $h \in \mathcal{H}_n$ if for any sequence of random variables $r_n = o_p(1)$, there exist a sequence $b_n = o_p(1)$, respectively $O_p(1)$, such that $\sup_{n,h \in \mathcal{H}_n} \sup_{\|\theta - \theta_{n,h}\| \leq r_n} |H_n(\theta)| \leq b_n$. The following is an extension of Corollary 8 of Sherman (1994).

Lemma 1. *Let $\mathcal{F}_n = \{f_n(\cdot, \theta, h) : \theta \in \Theta, h > 0\}$ be a class of degenerate functions on \mathbb{R}^k , $k \geq 1$, where $f_n(\cdot, \theta_{n,h}, \cdot) \equiv 0$. If*

- i. \mathcal{F}_n is Euclidean for an envelope F satisfying $\mathbb{E}F^4 < \infty$ uniformly in n ,*
- ii. There is a ball $B(\theta_{n,h}, r)$ and positive constants a and c , with r , a , and c independent on n and h , such that $\mathbb{E}f_n^2(\cdot, \theta, h) \leq c\|\theta - \theta_{n,h}\|^a$ for all $\theta \in B(\theta_{n,h}, r)$, all $h > 0$, and all n ,*

then uniformly over $B(\bar{\theta}_{n,h}, r)$ and $h > 0$, and for any $0 < \alpha < 1$

$$n^{k/2}U_n^k f_n(\cdot, \theta, h) = \|\theta - \theta_{n,h}\|^{a\alpha/2}O_p(1) + O_p\left(n^{-\alpha/4}\right).$$

If we assume further that $f_n^2(\cdot, \theta_{n,h}, h) \leq \Phi(\cdot)\|\theta - \theta_{n,h}\|^a$ with $\mathbb{E}\Phi < \infty$, then then uniformly over $B(\bar{\theta}_{n,h}, r)$ and $h > 0$, $n^{k/2}U_n^k f_n(\cdot, \theta, h) = \|\theta - \theta_{n,h}\|^{a\alpha/2}O_p(1)$ for any $0 < \alpha < 1$.

Proof. For simplicity, write \mathcal{N} for $B(\theta_{n,h}, r_n)$. Following the proof of Sherman (1994, Corollary 8),

$$\mathbb{E} \sup_{\theta \in \mathcal{N}, h > 0} \left| n^{k/2}U_n^k f_n(\cdot, \theta, h) \right| \leq \left[\mathbb{E} \sup_{\theta \in \mathcal{N}, h > 0} U_{2n}^k f_n^2(\cdot, \theta, h) \right]^{\alpha/2}$$

for any $0 < \alpha < 1$. Under the last condition, one readily obtains the desired result. Under Conditions i and ii only,

$$\mathbb{E} \sup_{\theta \in \mathcal{N}, h > 0} U_{2n}^k f_n^2(\cdot, \theta, h) \leq \sup_{\theta \in \mathcal{N}, h > 0} \mathbb{E}f_n^2(\cdot, \theta, h) + \sum_{i=1}^k \mathbb{E} \sup_{\theta \in \mathcal{N}, h > 0} U_{2n}^i f_{n,i}(\cdot, \theta, h)$$

where the class of functions $\{f_{n,i} : \theta \in \mathcal{N}, h > 0\}$ is degenerate on \mathbb{R}^i . Deduce that these classes are uniformly Euclidean for squared-integrable envelopes F_i from Lemma 2.14 of Pakes and Pollard (1989), and that $\mathbb{E} \sup_{\theta \in \mathcal{N}, h > 0} U_{2n}^i f_{n,i}(\cdot, \theta, h) = O(n^{-i/2})$ from Corollary 4 of Sherman (1994). \square

The following lemmas are extensions of Theorems 1 and 2 of Sherman (1993) and Theorems 1 and 2 of Sherman (1994b). The proofs proceed by straightforward modifications of his.

Lemma 2. *Let $\theta_{n,h}$ be the minimizer of $M_{n,h}(\theta)$ depending on a bandwidth h , \mathcal{H}_n a set of bandwidths, and let $\bar{\theta}_{n,h}$ be a minimizer of a function $\bar{M}_{n,h}(\theta)$ that may also depend on h . If*

- i. $\theta_{n,h} - \bar{\theta}_{n,h} = o_p(1)$ uniformly in $h \in \mathcal{H}_n$,*
- ii. there is a ball $B(\bar{\theta}_{n,h}, r)$ and a constant $\kappa > 0$, with r and κ independent on n and h , such that uniformly in $h \in \mathcal{H}_n$*

$$\bar{M}_{n,h}(\theta) - \bar{M}_{n,h}(\bar{\theta}_{n,h}) \geq (\kappa + o(1))\|\theta - \bar{\theta}_{n,h}\|^2 \quad \forall \theta \in B(\bar{\theta}_{n,h}, r),$$

iii. for some $\varepsilon_n = o(1)$ and uniformly over $o_p(1)$ neighborhood of $\bar{\theta}_{n,h}$ and $h \in \mathcal{H}_n$,

$$M_{n,h}(\theta) = \bar{M}_{n,h}(\theta) + \|\theta - \bar{\theta}_{n,h}\| O_p(1/\sqrt{n}) + \|\theta - \bar{\theta}_{n,h}\|^2 o_p(1) + O_p(\varepsilon_n),$$

then $\|\theta_{n,h} - \bar{\theta}_{n,h}\| = O_p\left[\max\left(\varepsilon_n^{1/2}, n^{-1/2}\right)\right]$ uniformly in $h \in \mathcal{H}_n$.

Lemma 3. Let $\theta_{n,h}$ be as in Lemma 2. Suppose $\theta_{n,h} - \bar{\theta}_{n,h} = O_p(1/\sqrt{n})$ uniformly in $h \in \mathcal{H}_n$, that the limit points of the sequence $\bar{\theta}_{n,h}$ are in the interior of Θ , and that uniformly over $O_p(1/\sqrt{n})$ neighborhoods of $\bar{\theta}_{n,h}$,

$$M_{n,h}(\theta) = M_{n,h}(\bar{\theta}_{n,h}) + \frac{1}{2} (\theta - \bar{\theta}_{n,h})' V_{n,h} (\theta - \bar{\theta}_{n,h}) + \frac{1}{\sqrt{n}} A'_{n,h} (\theta - \bar{\theta}_{n,h}) + o_p(1/n)$$

where $V_{n,h}$ is a sequence of positive definite matrices such that $0 < c_{\min} \leq \lambda_{\min}(V_{n,h}) \leq \lambda_{\max}(V_{n,h}) \leq c_{\max} < \infty$ for some c_{\min} and c_{\max} independent on n and h , and $A_{n,h} = O_p(1)$ uniformly in $h \in \mathcal{H}_n$. Then $\sqrt{n}(\theta_{n,h} - \bar{\theta}_{n,h}) + V_{n,h}^{-1} A_{n,h} = o_p(1)$ uniformly in $h \in \mathcal{H}_n$.

Lemma 4. Under Assumptions 3 and 4(v), $\sup_{n,h} \lambda_{\max}(V_{n,h}) < \infty$.

Lemma 5. Under Assumptions 3 and 4(v), $\liminf_n \inf_h \lambda_{\min}(V_{n,h}) > 0$ if $\mathcal{F}[K](ht) \geq \mathcal{F}[K](h_0t) \forall t \in \mathbb{R}^q$, $\forall h \in [0, h_0]$ and $\Pr[a' \nabla_{\theta} \tau(X, \theta_0) = 0] \leq 1$ for all $a \neq 0$.

The assumption on the kernel is fulfilled for instance by products of normal, logistic, Laplace, and Student densities.

Proof of Lemmas 4 and 5. For any n, h , and $a \in \mathbb{R}^p$,

$$\begin{aligned} a' V_{n,h} a &= \mathbb{E} \left[a' \nabla_{\theta} \tau_n(X_1, \theta_0) \nabla'_{\theta} \tau_n(X_2, \theta_0) a h^{-q} K \left(\frac{X_1 - X_2}{h} \right) \right] \\ &= (2\pi)^{q/2} \left\{ \int_{\mathbb{R}^q} \sum_{k=1}^r \left| \mathcal{F} \left[a' \nabla_{\theta} \tau_n^{(k)}(\cdot, \theta_0) f(\cdot) \right] (t) \right|^2 \mathcal{F}[K](ht) dt \right\}, \end{aligned} \quad (1)$$

Since $\mathcal{F}[K](ht) \leq (2\pi)^{-q/2}$ for all h, t , and by Assumptions 3 and 4(v),

$$\sup_{n,h} \lambda_{\max}(V_{n,h}) = \sup_n \lambda_{\max}(V_{n,0}) \leq \lambda_{\max}(\mathbb{E}[\nabla_{\theta} \tau(X, \theta_0) \nabla'_{\theta} \tau(X, \theta_0) f(X)]) \sup_{n,u} \lambda_{\min}^{-1}(W_n(u)) < \infty.$$

If $\mathcal{F}[K](ht) \geq \mathcal{F}[K](h_0t)$ for all $t, h \in [0, h_0]$, $\liminf_n \inf_h \lambda_{\min}(V_{n,h}) = \liminf_n \lambda_{\min}(V_{n,h_0})$ from (1). Moreover, $\liminf_n \lambda_{\min}(V_{n,h_0}) \geq \lambda_{\min}(V_{h_0}) - \limsup_n \|\tilde{V}_{n,h_0}\|_2$, where $\tilde{V}_{n,h_0} = V_{n,h_0} - V_{h_0}$ and $\|\cdot\|_2$ denotes the spectral norm. From Assumption 3 and since the map $W \mapsto W^{-1/2}$ is continuous, see Equation (A.2) below, $\sup_u \lambda_{\max}(W^{-1/2}(u))$ and $\sup_{n,u} \lambda_{\max}(W_n^{-1/2}(u))$ are bounded, and $W_n^{-1/2}(u) - W^{-1/2}(u) = o(1)$ for any u . It follows from the Lebesgue dominated convergence theorem and Assumption 4(v) that $\limsup_n \|\tilde{V}_{n,h_0}\|_2 = o(1)$. Therefore $\liminf_n \lambda_{\min}(V_{n,h_0}) \geq (1/2)\lambda_{\min}(V_{h_0})$. Using (1) and the unicity of the Fourier transform,

$$\begin{aligned} \lambda_{\min}(V_{h_0}) = 0 &\Leftrightarrow \exists a \neq 0 : a' \nabla_{\theta} \tau(X, \theta_0) W^{-1/2}(X) f(X) = 0 \text{ a.s.} \\ &\Leftrightarrow \exists a \neq 0 : a' \nabla_{\theta} \tau(X, \theta_0) = 0 \text{ a.s.} \end{aligned}$$

Thus $\lambda_{\min}(V_{h_0}) > 0$. □

Main proofs

In the main proofs, we use a single index n in place of the double indices n and h , e.g. we write M_n instead of $M_{n,h}$.

Proof of Theorem 2.1. Replacing $g(Z, \theta)$ by $g_n(Z, \theta) = W_n^{-1/2}(X)g(Z, \theta)$ in (2.6) yields

$$\begin{aligned} \mathbb{E}M_n(\theta) &= (1/2) \int \left| \mathcal{F} \left[\mathbb{E} \left[g_n^{(k)}(Z, \theta) | X = \cdot \right] f(\cdot) \right] (t) \right|^2 d\mu(ht) = 0 \\ &\Leftrightarrow \mathcal{F} \left[\mathbb{E} \left[g_n^{(k)}(Z, \theta) | X = \cdot \right] f(\cdot) \right] (t) = 0 \quad \forall t \in \mathbb{R}^q, k = 1, \dots, r \\ &\Leftrightarrow W_n^{-1/2}(X) \mathbb{E} [g(Z, \theta) | X] = 0 \quad \text{a.s.} \Leftrightarrow \theta = \theta_0, \end{aligned}$$

as $W_n(X)$ is positive definite and μ is a measure with support \mathbb{R}^q . If μ is zero at isolated points, then the second implication still holds by continuity of the Fourier transform. Since $\mathbb{E}M_n(\theta)$ is continuous in θ from Assumptions 4-(ii) and 3 as well as in h , we have that $\forall \varepsilon > 0, \exists \mu > 0$ such that $\inf_{\|\theta - \theta_0\| \geq \varepsilon, 0 \leq h \leq h_0} \mathbb{E}M_n(\theta) \geq \mu$ for n large enough. The family of functions $\{g'(Z_1, \theta)W_n^{-1/2}(X_1)W_n^{-1/2}(X_2)g(Z_2, \theta)K((X_1 - X_2)/h) : \theta \in \Theta, h > 0\}$ is Euclidean for a square-integrable envelope by Assumptions 2 and 4, Lemma 22(ii) of Nolan and Pollard (1987) and Lemma 2.14(ii) of Pakes and Pollard (1989). Thus by Corollary 7 of Sherman (1994), $\sup_{\theta \in \Theta, h > 0} |h^q M_n(\theta) - \mathbb{E}h^q M_n(\theta)| = O_{\mathbb{P}}(n^{-1/2})$. Let $\bar{\mathcal{H}}_n$ the set of bandwidths from the theorem and consider a set on which $\sup_{\theta \in \Theta, h \in \bar{\mathcal{H}}_n} |h^q M_n(\theta) - \mathbb{E}h^q M_n(\theta)| \leq Cn^{-1/2} \ln \ln(n+2)$, whose probability tends to one for any constant $C > 0$. On this set,

$$\inf_{\|\theta - \theta_0\| \geq \varepsilon} \inf_{h \in \bar{\mathcal{H}}_n} [M_n(\theta) - M_n(\theta_0)] \geq \inf_{\|\theta - \theta_0\| \geq \varepsilon} \inf_{h \in \bar{\mathcal{H}}_n} \mathbb{E}M_n(\theta) - \left[2C \ln \ln(n+2) / (\ln(n+1))^{-1/2} \right]$$

so that $\inf_{\|\theta - \theta_0\| \geq \varepsilon} \sup_{h \in \bar{\mathcal{H}}_n} [M_n(\theta) - M_n(\theta_0)] \geq \mu/2$ for n large enough. Since $M_n(\bar{\theta}_n) \leq M_n(\theta_0)$, it follows that $\sup_{h \in \bar{\mathcal{H}}_n} \|\bar{\theta}_n - \theta_0\| < \varepsilon$ with probability tending to one. \square

Proof of Theorem 2.2. The proof follows from Parts (ii) to (iv) of Theorem 2.3's proof, setting $\bar{\theta}_n = \theta_0$ and accounting for (1.1). \square

Proof of Theorem 2.3. When studying our estimator under misspecification, we define $\bar{\mathcal{F}}_n = \{\bar{\phi}_{n,h}(\cdot) : h \in [0, h_0]\}$, where

$$\bar{\phi}_{n,h}(z) = \mathbb{E} \left[\nabla_{\theta} \tau(X, \bar{\theta}_{n,h}) W_n^{-1/2}(X) h^{-q} K((x - X)/h) \right] W_n^{-1/2}(x) g(z, \bar{\theta}_{n,h}),$$

and $\bar{\phi}_{n,0}(z) = \nabla_{\theta} \tau(x, \bar{\theta}_{n,0}) f(x) W_n^{-1}(x) g(z, \bar{\theta}_{n,0})$. Let $\{\mathbb{G}_n \bar{\phi}_{n,h} : h \in [0, h_0]\}$ be the sequence of centered empirical processes indexed by the families $\bar{\mathcal{F}}_n$,

$$\begin{aligned} \bar{V}_{n,h} &= \mathbb{H}_{\theta, \theta} \mathbb{E}M_n(\bar{\theta}_{n,h}) = \mathbb{E} \left[\nabla_{\theta} \tau_n(X_1, \bar{\theta}_{n,h}) \nabla_{\theta}' \tau_n(X_2, \bar{\theta}_{n,h}) h^{-q} K((X_1 - X_2)/h) \right] \\ &\quad + \sum_{k=1}^r \mathbb{E} \left[\mathbb{H}_{\theta, \theta} \tau_n^{(k)}(X_1, \bar{\theta}_{n,h}) g_n^{(k)}(X_2, \bar{\theta}_{n,h}) h^{-q} K((X_1 - X_2)/h) \right], \end{aligned}$$

and $\bar{V}_{n,0} = \lim_{h \downarrow 0} \bar{V}_{n,h} = \mathbb{H}_{\theta, \theta} \mathbb{E}M_n(\bar{\theta}_{n,0})$.

(i) Consistency: Since $\bar{\theta}_n$ is the unique minimizer of $\mathbb{E}M_n(\theta)$, reason as in Theorem 2.1's proof to show that $\sup_{h \in \bar{\mathcal{H}}_n} \|\bar{\theta}_n - \bar{\theta}_n\| = o_p(1)$.

(ii) \sqrt{n} -consistency: Since $\nabla_{\theta} \mathbb{E} M_n(\bar{\theta}_n) = 0$ and $\inf_{n,h} \lambda_{\min}(V_{n,h}) > 0$, we have uniformly in $h \in \mathcal{H}_n$

$$\begin{aligned} & \mathbb{E} M_n(\theta) - \mathbb{E} M_n(\bar{\theta}_n) \\ &= (\theta - \bar{\theta}_n)' \nabla_{\theta} \mathbb{E} M_n(\bar{\theta}_n) + \frac{1}{2} (\theta - \bar{\theta}_n)' \mathbb{H}_{\theta, \theta} \mathbb{E} M_n(\bar{\theta}_n) (\theta - \bar{\theta}_n) + o(\|\theta - \bar{\theta}_n\|^2) \\ &= \frac{1}{2} (\theta - \bar{\theta}_n)' \bar{V}_{n,h} (\theta - \bar{\theta}_n) + o(\|\theta - \bar{\theta}_n\|^2) \geq \frac{1}{2} \left(\inf_{n,h} \lambda_{\min}(\bar{V}_{n,h}) + o(1) \right) \|\theta - \bar{\theta}_n\|^2. \end{aligned}$$

Now apply Hoeffding's decomposition to $M_n(\theta) - M_n(\bar{\theta}_n)$ and consider the first-order empirical process $\mathbb{P}_n \tilde{l}_{\theta}$, where $\tilde{l}_{\theta}(Z_i) = \mathbb{E}[l_{\theta}(Z_i, Z_j) \mid Z_i] + \mathbb{E}[l_{\theta}(Z_i, Z_j) \mid Z_j] - 2\mathbb{E}[l_{\theta}(Z_i, Z_j)]$,

$$\begin{aligned} l_{\theta}(Z_i, Z_j) &= (1/2) (g'_n(Z_i, \theta) g_n(Z_j, \theta) - g'_n(Z_i, \bar{\theta}_n) g_n(Z_j, \bar{\theta}_n)) h^{-q} K((X_i - X_j)/h) \\ &= (1/2) g'_n(Z_i, \bar{\theta}_n) (g_n(Z_j, \theta) - g_n(Z_j, \bar{\theta}_n)) h^{-q} K((X_i - X_j)/h) \\ &\quad + (1/2) (g_n(Z_i, \theta) - g_n(Z_i, \bar{\theta}_n))' g_n(Z_j, \bar{\theta}_n) h^{-q} K((X_i - X_j)/h) \\ &\quad + (1/2) (g_n(Z_i, \theta) - g_n(Z_i, \bar{\theta}_n))' (g_n(Z_j, \theta) - g_n(Z_j, \bar{\theta}_n)) h^{-q} K((X_i - X_j)/h) \\ &= l_{1\theta}(Z_i, Z_j) + l_{2\theta}(Z_i, Z_j) + l_{3\theta}(Z_i, Z_j), \end{aligned}$$

and $l_{1\theta}(Z_i, Z_j) = l_{2\theta}(Z_j, Z_i)$ by the symmetry of $K(\cdot)$. Now $\mathbb{E}[l_{1\theta}(Z_i, Z_j) \mid Z_j]$ and from Assumption M5,

$$\begin{aligned} 2\mathbb{E}[l_{1\theta}(Z_i, Z_j) \mid Z_i] &= g'_n(Z_i, \bar{\theta}_n) \mathbb{E} [(g_n(Z, \theta) - g_n(Z, \bar{\theta}_n)) h^{-q} K((X_i - X)/h) \mid Z_i] \\ &= g'_n(Z_i, \bar{\theta}_n) \left[\int_{\mathbb{R}^q} \nabla'_{\theta} \tau_n(x, \bar{\theta}_n) f(x) h^{-q} K((X_i - x)/h) dx \right] (\theta - \bar{\theta}_n) \end{aligned} \quad (2)$$

$$\begin{aligned} &+ \frac{1}{2} g'_n(Z_i, \bar{\theta}_n) \sum_{k,l=1}^p (\theta^{(k)} - \bar{\theta}_n^{(k)}) (\theta^{(l)} - \bar{\theta}_n^{(l)}) \\ &\quad \left[\int_{\mathbb{R}^q} \mathbb{H}_{\theta^{(k)} \theta^{(l)}} \tau_n(x, \bar{\theta}_n) f(x) h^{-q} K((X_i - x)/h) dx \right] + R_{1n}(Z_i, \theta) \end{aligned} \quad (3)$$

$$\text{where } \|R_n(Z_i, \theta)\| \leq G(Z_i) \|\theta - \bar{\theta}_n\|^{2+a} \left[\sum_{k=1}^r \left(\int_{\mathbb{R}^q} H_n^{(k)}(X_i - hu) f(X_i - hu) |K(u)| du \right)^2 \right]^{1/2}$$

and $H_n(\cdot) = W_n^{-1/2}(\cdot) H(\cdot)$. By Assumption M6-(i), the functions $\nabla_{\theta} \tau_n^{(k)}(\cdot, \bar{\theta}_n) f(\cdot)$, $n \geq 1$ satisfy Condition (ME) for an envelope Φ with $\mathbb{E} \Phi^a(X) < \infty$ for some $a \geq 4$. Use Assumption M4 and Lemma 2.14-(ii) in Pakes and Pollard (1989) to conclude that the family of functions $\tilde{\phi}'_{n,h}(z)$ indexed by h in (2) is uniformly Euclidean for a squared-integrable envelope. Hence $A'_n = \mathbb{G}_n \tilde{\phi}'_{n,h} = O_p(1)$ uniformly in θ and $h \in [0, h_0]$ by Corollary 4 of Sherman (1994). Similarly, the family of functions in (3) is uniformly Euclidean for an integrable envelope. By a version of the Glivenko-Cantelli for families changing with n , see e.g. van de Geer (2000, p.44), the centered empirical sum based on this family of functions is then an $o_p(1)$ uniformly in $h \in [0, h_0]$. Finally, $\left\{ G(z) \int_{\mathbb{R}^q} H_n^{(k)}(x - hu) f(x - hu) |K(u)| du : h \in [0, h_0] \right\}$ are also uniformly Euclidean for an integrable envelope, so that the (uncentered) empirical sum based on this family of functions is a $O_p(1)$

uniformly in $h \in [0, h_0]$. A similar expansion for $l_{3\theta}$ yields

$$\begin{aligned}
2\mathbb{E}[l_{3\theta}(Z_i, Z_j) \mid Z_i] &= (g_n(Z_i, \theta) - g_n(Z_i, \bar{\theta}_n))' \mathbb{E} [(g_n(Z, \theta) - g_n(Z, \bar{\theta}_n)) h^{-q} K((X_i - X)/h) \mid Z_i] \\
&= (g_n(Z_i, \theta) - g_n(Z_i, \bar{\theta}_n))' \\
&\quad \left[\int_{\mathbb{R}^q} \nabla'_\theta \tau_n(x, \bar{\theta}_n) f(x) h^{-q} K((X_i - x)/h) dx \right] (\theta - \bar{\theta}_n) \\
&\quad + \frac{1}{2} (g_n(Z_i, \theta) - g_n(Z_i, \bar{\theta}_n))' \sum_{k,l=1}^p (\theta^{(k)} - \bar{\theta}_n^{(k)}) (\theta^{(l)} - \bar{\theta}_n^{(l)}) \\
&\quad \left[\int_{\mathbb{R}^q} \mathbb{H}_{\theta^{(k)}\theta^{(l)}} \tau_n(x, \bar{\theta}_n) f(x) h^{-q} K((X_i - x)/h) dx \right] + R_{3n}(Z_i, \theta).
\end{aligned} \tag{4}$$

Since the function in (4) is such that

$$\mathbb{E} \left[(g_n(Z_i, \theta) - g_n(Z_i, \bar{\theta}_n))' \left[\int_{\mathbb{R}^q} \nabla'_\theta \tau_n(x, \bar{\theta}_n) f(x) h^{-q} K((X_i - x)/h) dx \right] \right] \rightarrow 0$$

as $\theta - \bar{\theta}_n \rightarrow 0$, the centered process based on these functions is an $o_p(1/\sqrt{n})$ uniformly in θ and h by Corollary 8 of Sherman (1994). The remaining terms can be dealt with similarly. Hence

$$\mathbb{P}_n \tilde{l}_\theta = \frac{1}{\sqrt{n}} A'_n (\theta - \bar{\theta}_n) + \|\theta - \bar{\theta}_n\|^2 o_p(1), \tag{5}$$

uniformly over $o_p(1)$ neighborhoods of $\bar{\theta}_n$ and $h \in [0, h_0]$.

Consider the second order U -process $U_n \bar{l}_\theta$ in the decomposition of $M_n(\theta) - M_n(\bar{\theta}_n)$. For $\theta \in \mathcal{N}$, $\mathbb{E} h^{2q} l_\theta^2(Z_i, Z_j) = \mathbb{E} [(g'_n(Z_i, \theta) g_n(Z_j, \theta) - g'_n(Z_i, \bar{\theta}_n) g_n(Z_j, \bar{\theta}_n)) K((X_i - X_j)/h)]^2$. Since $K(\cdot)$ is bounded, the Z_i are independent, and for any $a_1, \dots, a_r \in \mathbb{R}$, $(a_1 + \dots + a_r)^2 \leq r(a_1^2 + \dots + a_r^2)$, deduce that $\mathbb{E} h^{2q} l_\theta^2(Z_i, Z_j) = O(\|\theta - \bar{\theta}_n\|)$. From Assumption M4-(iii), $h^q l_\theta(Z_i, Z_j)$ is Euclidean for an integrable envelope with fourth moment. Use Lemma 1 to deduce that for any $0 < \alpha < 1$

$$\sup_{h>0} |U_n h^q \bar{l}_\theta| = \|\theta - \bar{\theta}_n\|^{\alpha/2} O_p(n^{-1}) + O_p(n^{-1-\alpha/4})$$

uniformly over $o_p(1)$ neighborhoods of $\bar{\theta}_n$, which yields

$$\sup_{h \in \mathcal{H}_n} |U_n \bar{l}_\theta| = \|\theta - \bar{\theta}_n\|^{\alpha/2} O_p(\sup_{h \in \mathcal{H}_n} n^{-1} h^{-q}) + O_p(\sup_{h \in \mathcal{H}_n} n^{-1-\alpha/4} h^{-q}). \tag{6}$$

Choose $\alpha < 1$ such that $nh^{4q/\alpha} \geq C$ for all $h \in \mathcal{H}_n$ from our assumption to deduce that the second term is a $O_p(n^{-1})$. For θ in a $o_p(1)$ neighborhood of $\bar{\theta}_n$, the first term is $O_p(\varepsilon_{0,n})$ with $\varepsilon_{0,n} = o(\sup_{h \in \mathcal{H}_n} n^{-1} h^{-q})$. Use Equations (5) and (6) in conjunction with Lemma 2 to obtain $\|\tilde{\theta}_n - \bar{\theta}_n\| = O_p(\varepsilon_{0,n}^{1/2})$. Plug in this result in (6), so that the first term is a $O_p(\varepsilon_{1,n})$ with $\varepsilon_{1,n} = \varepsilon_{0,n}^{1+\alpha/4}$. Apply repeatedly m times to get $\varepsilon_{m,n} = \varepsilon_{0,n}^{\alpha^m}$ with $\alpha_m = \sum_{j=0}^{m-1} (\alpha/4)^j$. When m increases, $\varepsilon_{m,n}$ decreases and α_m tends to $4/(4-\alpha)$. Since $\varepsilon_{0,n}^{4/(4-\alpha)} = o(n^{-1})$, after m iterations with m finite large enough, the first term in Equation (6) is a $O_p(n^{-1})$. Apply then again Lemma 2 to conclude that $\|\tilde{\theta}_n - \bar{\theta}_n\| = O_p(n^{-1/2})$.

Remark that under Condition 2, Equation (6) becomes $\sup_h |U_n \bar{l}_\theta| = \|\theta - \bar{\theta}_n\|^\alpha O_p(\sup_h n^{-1} h^{-q})$. Choose any $\alpha < 1$ such that $nh^{\frac{2q}{\alpha}} \geq C$ for all h and reason as above to obtain that $\sup_h |U_n \bar{l}_\theta| = O_p(n^{-1})$ and $\|\tilde{\theta}_n - \bar{\theta}_n\| = O_p(n^{-1/2})$.

(iii) Asymptotic representation: Equation (7.16) and Part (ii) imply that for any $\alpha \leq \alpha' < 1$, where α comes from our assumptions, $\sup_h |U_n \bar{l}_\theta| = O_p(\sup_h n^{-1-\alpha'/4} h^{-q})$. Conclude that $\sup_h |U_n \bar{l}_\theta| = o_p(n^{-1})$, and use (5) to obtain

$$M_n(\theta) = M_n(\bar{\theta}_n) + \frac{1}{2} (\theta - \bar{\theta}_n)' \bar{V}_n (\theta - \bar{\theta}_n) + \frac{1}{\sqrt{n}} A'_n (\theta - \bar{\theta}_n) + o_p(1/n),$$

uniformly over $O_p(1/\sqrt{n})$ neighborhoods of $\bar{\theta}_n$ and in $h \in \mathcal{H}_n$. Conclude from Lemma 3 that $\sqrt{n} (\tilde{\theta}_n - \bar{\theta}_n) + \bar{V}_n^{-1} A_n = o_p(1)$.

(iv) Behavior of $\mathbb{G}_n \bar{\phi}_{n,h}$: We consider the case $r = 1$, the multivariate case follows similarly at the cost of more cumbersome algebra. We apply Theorem 19.28 of van der Vaart (1998), where the Lindeberg condition follows from our Assumption M4 and M6. We first consider that $\bar{\theta}_{n,h} = \theta_0$, i.e. a correct model. We have to show his Condition (19.27), that is $\sup_{|h_1 - h_2| < \delta} \mathbb{E} \|\phi_{n,h_1}(Z) - \phi_{n,h_2}(Z)\|^2 \rightarrow 0$ whenever $\delta \rightarrow 0$. Let $\omega_n^2(X, \theta_0) = \mathbb{E} [g_n^2(Z, \theta_0) | X]$. Proceed as in the consistency proof to show that

$$\begin{aligned} \mathbb{E} [\phi'_{n,h_1}(Z) \phi_{n,h_2}(Z)] &= (2\pi)^{q/2} \int_{\mathbb{R}^q} \int_{\mathbb{R}^q} \mathcal{F} [\nabla'_\theta \tau_n(\cdot, \theta_0)] f(\cdot) (-t) \mathcal{F} [\omega_n^2(\cdot, \theta_0) f(\cdot)] (t - u) \\ &\quad \mathcal{F} [\nabla_\theta \tau_n(\cdot, \theta_0)] f(\cdot) (u) \mathcal{F} [K] (h_1 t) \mathcal{F} [K] (h_2 u) dt du. \end{aligned}$$

$$\begin{aligned} \text{Hence, } \mathbb{E} \|\phi_{n,h_1}(Z) - \phi_{n,h_2}(Z)\|^2 &= (2\pi)^{q/2} \int_{\mathbb{R}^q} \int_{\mathbb{R}^q} \mathcal{F} [\nabla'_\theta \tau_n(\cdot, \theta_0)] f(\cdot) (-t) \mathcal{F} [\omega_n^2(\cdot, \theta_0) f(\cdot)] (t - u) \mathcal{F} [\nabla_\theta \tau_n(\cdot, \theta_0) f(\cdot)] (u) \\ &\quad [\mathcal{F} [K] (h_1 t) \mathcal{F} [K] (h_1 u) - 2\mathcal{F} [K] (h_1 t) \mathcal{F} [K] (h_2 u) + \mathcal{F} [K] (h_2 t) \mathcal{F} [K] (h_2 u)] dt du. \end{aligned}$$

Use the uniform continuity of $\mathcal{F} [K] (\cdot)$, Assumption 4(v)-(vi), the properties of the convolution of Fourier transforms, and the Lebesgue dominated convergence theorem to conclude. The case where $h_2 = 0$ can be treated similarly.

We now turn to the general case of a misspecified model, so we make explicit θ as an argument of $\bar{\phi}_{n,h}$. The result similarly follows if we show $\sup_{|h_1 - h_2| < \delta, \|\theta_1 - \theta_2\| < \delta} \mathbb{E} \|\bar{\phi}_{n,h_1}(Z, \theta_1) - \bar{\phi}_{n,h_2}(Z, \theta_2)\|^2 \rightarrow 0$ whenever $\delta \rightarrow 0$. When only h varies in this expression, we can apply our previous reasoning, provided we use $\omega_n^2(X, \theta_1, \theta_2) = \mathbb{E} [g_n(Z, \theta_1) g_n(Z, \theta_2) | X]$ together with Assumption M4. We are left to deal with the case where only θ varies. The result follows from continuity arguments, i.e. Assumptions M4(iv), M5, and M6, and the Lebesgue dominated convergence theorem. \square

Proof of Theorem 2.4. The proof rely on the three intermediary results.

Corollary 6. *Under the Assumptions of Theorem 2.2, if $W(X) = \text{Var} [g(Z, \theta_0) | X] f(X)$, then uniformly over $h \in \mathcal{H}'_n = \{1/\ln(n+1) \geq h > 0 : nh^{4q/\alpha} \geq C\}$, $C > 0$ and $0 < \alpha < 1$ arbitrary, $\sqrt{n} (\tilde{\theta}_{n,h} - \theta_0)$ is asymptotically $N(0, \Sigma^{-1})$ with*

$$\Sigma = \mathbb{E} [\nabla_\theta \mathbb{E} [g(Z, \theta_0) | X] \text{Var}^{-1} [g(Z, \theta_0) | X] \nabla'_\theta \mathbb{E} [g(Z, \theta_0) | X]].$$

Theorem 7. *For an i.i.d. sample, under Assumptions 1, 2, E2, E4, 5, and E7,*

$$\sup_{h,b \in \mathcal{H}'_n} \left| \widehat{M}_{n,h,b}(\theta) - M_{n,h,b}(\theta) \right| = o_p(n^{-1} + \|\theta - \theta_0\|/\sqrt{n} + \|\theta - \theta_0\|^2) \quad (7)$$

uniformly over θ in $o(1)$ neighborhoods of θ_0 , where $M_{n,h,b}(\theta)$ is defined as in (2.2) with $W_n(x) = \mathbb{E} [\widehat{W}_n(x, \theta_0)]$.

This result ensures the equivalence of the two-step $\widehat{\theta}_{n,h,b}$ and $\tilde{\theta}_{n,h}$ with weighting matrix $\mathbb{E} \left[\widehat{W}_n(\cdot, \theta_0) \right]$. We can then apply Theorem 2.2 accounting for its dependence on b . Note that the SMD estimator is not affected by boundary effects in the estimation of the optimal weighting matrix, since only pointwise convergence in the interior of the support of X is necessary.

Corollary 8. *Under the assumptions of Theorem 7 and E6, $\sqrt{n} \left(\widehat{\theta}_{n,h,b} - \theta_0 \right)$ is asymptotically $N(0, \Sigma^{-1})$ uniformly in $h, b \in \mathcal{H}'_n$.*

Proof of Corollary 6. Under our assumptions, $\tilde{\theta}_{n,h}$ is asymptotically $N(0, V_0^{-1} \Delta_{0,0} V_0^{-1})$ uniformly over $h \in \mathcal{H}'_n$ where

$$V_0 = \mathbb{E} \left[\nabla_{\theta} \mathbb{E} [g(Z, \theta_0) | X] W^{-1}(X) \nabla'_{\theta} \mathbb{E} [g(Z, \theta_0) | X] f(X) \right] \quad \text{and}$$

$$\Delta_{0,0} = \mathbb{E} \left[\nabla_{\theta} \mathbb{E} [g(Z, \theta_0) | X] W^{-1}(X) \text{Var} [g(Z, \theta_0) | X] W^{-1}(X) \nabla'_{\theta} \mathbb{E} [g(Z, \theta_0) | X] f^2(X) \right].$$

Plug $W(X) = \text{Var} [g(Z, \theta_0) | X] f(X)$ to obtain the result. \square

Proof of Theorem 7. Step 1 is to obtain the uniform rate of convergence of $\widehat{W}_n(\cdot, \theta) - W_n(\cdot, \theta)$, where $W_n(\cdot, \theta) = \mathbb{E} \left[\widehat{W}_n(\cdot, \theta) \right]$. A useful result can be derived along the lines of Proposition 2 of Einmahl and Mason (2005). A careful inspection of their proof shows that the result holds not only on a compact subset, but on the whole space \mathbb{R}^q provided their Condition (1.7) on the continuity of the density $f(\cdot)$ is replaced by the assumption of a bounded density.

Lemma 9. *Let Φ denote a class of measurable functions on \mathbb{R}^{d+q} , where $d, q \geq 1$, with a finite-valued measurable envelope function F . Further assume that the kernel $L(\cdot)$ is a density of bounded variation with bounded support, the density $f(\cdot)$ is bounded and*

$$\sup_{x \in \mathbb{R}^q} \mathbb{E}[F^4(Z) | X = x] < \infty.$$

Let $\eta_{\varphi,n,b}(x) = (nb^q)^{-1} \sum_{1 \leq i \leq n} \varphi(Z_i) L((x - X_i)/b)$, $\varphi \in \Phi$ and $\|\cdot\|_{\infty}$ be the supremum norm. There exists $c > 0$ such that, with probability 1

$$\limsup_{n \rightarrow \infty} \sup_{b \in \mathcal{H}_n} \sqrt{nb^q} \frac{\sup_{\varphi \in \Phi} \|\eta_{\varphi,n,b} - \mathbb{E} \eta_{\varphi,n,b}\|_{\infty}}{\sqrt{\ln(1/b^q) \vee \ln \ln n}} = c.$$

Step 2 consists in establishing an expansion of the power $-1/2$ of a positive definite matrix. By the integral representation of the square root of a matrix, see e.g. Higham (2008), for any positive definite $q \times q$ matrix A

$$A^{-1/2} = \frac{2}{\pi} \int_0^{\infty} (t^2 A + I)^{-1} dt.$$

Moreover, for any conformable square matrices B and D and any $t > 0$,

$$(A + B)^{-1} = A^{-1} - A^{-1} (I + BA^{-1})^{-1} BA^{-1}, \quad (8)$$

and $\left[I + t^2 D (t^2 A + I)^{-1} \right]^{-1} = I - t^2 D (t^2 A + I)^{-1} + R,$

$$\begin{aligned} \text{with } \|R\| \leq \sqrt{q} \|R\|_2 &\leq \frac{\sqrt{q} \left\| t^2 D (t^2 A + I)^{-1} \right\|_2^2}{1 - \left\| t^2 D (t^2 A + I)^{-1} \right\|_2} \\ &\leq \sqrt{q} \|D\|_2^2 \left[\frac{t^2}{1 + t^2 \lambda_{\min}(A)} \right]^2 \left[1 - \frac{t^2 \|D\|_2}{1 + t^2 \lambda_{\min}(A)} \right]^{-1} \\ &\leq k(c) \|D\|_2^2 \leq k(c) \|D\|^2. \end{aligned}$$

Here and in what follows, $\|\cdot\|_2$ denotes the spectral matrix norm, $\lambda_{\min}(A)$ is as before the smallest eigenvalue of A , and $k(c)$ depends on c , $\lambda_{\min}(A)$, and \sqrt{q} , where c is assumed to be such that

$$\|D\|_2 / \lambda_{\min}(A) \leq \|D\| / \lambda_{\min}(A) \leq c < 1.$$

Use the integral representation for $(A + D)^{-1/2}$ and $A^{-1/2}$ and apply (8) with A replaced by $t^2 A + I$ and $B = t^2 D$ to deduce that

$$\begin{aligned} (A + D)^{-1/2} - A^{-1/2} &= -\frac{2}{\pi} \int_0^\infty t^2 (t^2 A + I)^{-1} D (t^2 A + I)^{-1} dt \\ &\quad + \frac{2}{\pi} \int_0^\infty t^4 (t^2 A + I)^{-1} D (t^2 A + I)^{-1} D (t^2 A + I)^{-1} dt \\ &\quad - \frac{2}{\pi} \int_0^\infty t^2 (t^2 A + I)^{-1} R D (t^2 A + I)^{-1} dt, \end{aligned} \quad (9)$$

$$\text{where } \left\| (t^2 A + I)^{-1} R D (t^2 A + I)^{-1} \right\| \leq \left[\frac{1}{1 + t^2 \lambda_{\min}(A)} \right]^2 k(c) \|D\|^3.$$

This implies that for some constant C the last integral in (9) is bounded by

$$\frac{2}{\pi} k(c) \|D\|^3 \int_0^\infty t^2 [1 + t^2 \lambda_{\min}(A)]^{-2} dt \leq C \|D\|^3.$$

Step 3 consists in applying Identity (9) to our problem, with $D = D_{n,i}(\theta_2) = \widehat{W}_n(X_i, \theta_2) - W_n(X_i, \theta_0)$ and $A = W_n(X_i, \theta_0) = W_n(X_i)$. Let $\widehat{M}_n(\theta, \theta_2)$ and $M_n(\theta)$ be the objective functions with weighting matrix $\widehat{W}_n(\cdot, \theta_2)$ and $W_n(\cdot)$, respectively. Let also $0 < \lambda \leq \inf_{x,n} \lambda_{\min}(W_n(x))$ for some fixed $\lambda > 0$, which exists by our Assumption E7. For any $\theta \in \Theta$ and θ_2 in a $O(n^{-1/2})$ neighborhood of θ_0 ,

$$\begin{aligned} \widehat{M}_n(\theta, \theta_2) &= M_n(\theta) - \frac{2}{\pi} \int_0^\infty t^2 [1 + t^2 \lambda]^{-2} [M_{1n}(t) + M'_{1n}(t)] dt \\ &\quad + \frac{2}{\pi} \int_0^\infty t^4 [1 + t^2 \lambda]^{-3} [M_{2n}(t) + M'_{2n}(t)] dt \\ &\quad + \frac{4}{\pi^2} \int_0^\infty \int_0^\infty t^2 [1 + t^2 \lambda]^{-2} s^2 [1 + s^2 \lambda]^{-2} M_{3n}(t, s) dt ds \\ &\quad + O_p \left(\sup_{x \in \mathbb{R}^q} \sup_{\|\theta_2 - \theta_0\| \leq C n^{-1/2}} \sup_{b \in \mathcal{H}'_n} \left\| \widehat{W}_n(x, \theta_2) - \widehat{W}_n(x) \right\|^3 \right). \end{aligned}$$

The last term is an $o_p(n^{-1})$ uniformly in $b \in \mathcal{H}'_n$ by Step 1 and noticing that from Assumption E7, for some $C > 0$ and $\nu > 2/3$, $\|\mathbb{E}[(\omega^2(X, \theta_2) - \omega^2(X, \theta_0)) b^{-q} L((X-x)/b)]\| \leq c \|\theta_2 - \theta_0\|^\nu \|\mathbb{E}[b^{-q} L((X-x)/b)]\| \leq C \|\theta_2 - \theta_0\|^\nu = o(n^{-1/3})$ uniformly in θ_2 in a $O(n^{-1/2})$ neighborhood of θ_0 . In the last display,

$$\begin{aligned} M_{1n}(t) &= M_{1n}(t, \theta, \theta_2, h, b) \\ &= \frac{t^{-4} (1+t^2\lambda)^2}{2n(n-1)} \sum_{i \neq j} g'(Z_i, \theta) [W_n(X_i) + t^{-2}I]^{-1} D_{n,i}(\theta_2) \\ &\quad \times [W_n(X_i) + t^{-2}I]^{-1} W_n^{-1/2}(X_j) g(Z_j, \theta) K_{ij}, \\ M_{2n}(t) &= M_{2n}(t, \theta, \theta_2, h, b) \\ &= \frac{t^{-6} (1+t^2\lambda)^3}{2n(n-1)} \sum_{i \neq j} g'(Z_i, \theta) [W_n(X_i) + t^{-2}I]^{-1} D_{n,i}(\theta_2) [W_n(X_i) + t^{-2}I]^{-1} \\ &\quad \times D_{n,i}(\theta_2) [W_n(X_i) + t^{-2}I]^{-1} W_n^{-1/2}(X_j) g(Z_j, \theta) K_{ij}, \end{aligned}$$

$$\begin{aligned} M_{3n}(t, s) &= M_{3n}(t, s, \theta, \theta_2, h, b) \\ &= \frac{(1+t^2\lambda)^2 (1+s^2\lambda)^2}{t^4 s^4 2n(n-1)} \sum_{i \neq j} g'(Z_i, \theta) [W_n(X_i) + t^{-2}I]^{-1} D_{n,i}(\theta_2) [W_n(X_i) + t^{-2}I]^{-1} \\ &\quad \times [W_n(X_j) + s^{-2}I]^{-1} D_{n,j}(\theta_2) [W_n(X_j) + s^{-2}I]^{-1} g(Z_j, \theta) K_{ij}. \end{aligned}$$

Strictly speaking, we should separate the integrals on $[0, 1)$ and $[1, \infty)$ in the following. Specifically, for $t \in [0, 1)$, the terms $[W_n(\cdot) + t^{-2}I]^{-1}$ should be replaced by $[t^2 W_n(\cdot) + I]^{-1}$, with adequate changes in the other arguments under the integral. The following arguments adapt easily.

Step 4 is to show that uniformly over θ in a $o(1)$ neighborhood of θ_0 and θ_2 in a $O(n^{-1/2})$ neighborhoods of θ_0

$$\sup_{t, s \geq 1} \sup_{b, h \in \mathcal{H}_n} \{\|M_{1n}\| + \|M_{2n}\| + \|M_{3n}\|\} = o_p(n^{-1} + \|\theta - \theta_0\|/\sqrt{n} + \|\theta - \theta_0\|^2), \quad (10)$$

which implies (7). The terms M_{1n} , M_{2n} and M_{3n} involve the family of matrix-valued functions

$$\left\{ [W_n(\cdot) + t^{-2}I]^{-1} : b \in \mathcal{H}'_n, t \geq 1 \right\} \quad \text{and} \quad \left\{ W_n^{-1/2}(\cdot) : b \in \mathcal{H}'_n \right\}.$$

For $t \in [0, 1)$, the first family has to be replaced by $\left\{ [t^2 W_n(\cdot) + I]^{-1} : b \in \mathcal{H}'_n, t \in [0, 1) \right\}$. We here focus on the case $t \geq 1$, the arguments for the other case being similar. Lemma B.2 in Appendix B shows that under our assumptions these families of functions are Euclidean entrywise for a constant envelope. For the sake of simplicity, we show (10) only for $r = 1$, since the same arguments apply componentwise for $r > 1$ at the expense of much more cumbersome algebra. Also we focus on $M_{1n}(t)$, since a similar reasoning applies to $M_{2n}(t)$ and $M_{3n}(t)$. Let

$$\begin{aligned} d_{\theta_2}(x, Z_k) &= g^2(Z_k, \theta_2) L((x - X_k)/b) - \mathbb{E}[\omega^2(X, \theta_2) L((x - X)/b)], \\ \delta_{\theta_2}(x) &= \mathbb{E}[\omega^2(X, \theta_2) L((x - X)/b)] - \mathbb{E}[\omega^2(X, \theta_0) L((x - X)/b)], \end{aligned}$$

so that $D_{n,i}(\theta_2) = \frac{1}{nb^q} \sum_{1 \leq k \leq n} [d_{\theta_2}(X_i, Z_k) + \delta_{\theta_2}(X_i)]$. We accordingly separate $M_{1n}(t)$ into two terms and we study each of them in turn.

Note that $\mathbb{E}[d_{\theta_2}(X_i, Z_k) | X_i] = 0$ for $i \neq k$ and consider the decomposition

$$\begin{aligned}
& \frac{1}{nb^q} \frac{1}{(n)_2} \sum_{1 \leq k \leq n} \sum_{i \neq j} \frac{g(Z_i, \theta)}{[W_n(X_i) + t^{-2}]^2} d_{\theta_2}(X_i, Z_k) W_n^{-1/2}(X_j) g(Z_j, \theta) K_{ij} \\
&= \frac{(n-2)}{nb^q} \frac{1}{(n)_3} \sum_{i \neq j \neq k} \frac{g(Z_i, \theta)}{[W_n(X_i) + t^{-2}]^2} d_{\theta_2}(X_i, Z_k) W_n^{-1/2}(X_j) g(Z_j, \theta) K_{ij} \\
&\quad + \frac{1}{nb^q} \frac{1}{(n)_2} \sum_{i \neq j} \frac{g(Z_i, \theta)}{[W_n(X_i) + t^{-2}]^2} d_{\theta_2}(X_i, Z_i) W_n^{-1/2}(X_j) g(Z_j, \theta) K_{ij} \\
&\quad + \frac{1}{nb^q} \frac{1}{(n)_2} \sum_{i \neq j} \frac{g(Z_i, \theta)}{[W_n(X_i) + t^{-2}]^2} d_{\theta_2}(X_i, Z_j) W_n^{-1/2}(X_j) g(Z_j, \theta) K_{ij} \\
&= \frac{(n-2)}{nb^q h^q} \frac{1}{(n)_3} \sum_{i \neq j \neq k} m_{11}(Z_i, Z_j, Z_k) + \frac{1}{nb^q h^q} \frac{1}{(n)_2} \sum_{i \neq j} m_{12}(Z_i, Z_j) \\
&\quad + \frac{1}{nb^q h^q} \frac{1}{(n)_2} \sum_{i \neq j} m_{13}(Z_i, Z_j) \\
&= \frac{(n-2)}{nb^q h^q} M_{11n} + \frac{1}{nb^q h^q} M_{12n} + \frac{1}{nb^q h^q} M_{13n},
\end{aligned}$$

where $(n)_k = n!/(n-k)!$. For the first and dominant term, write

$$\begin{aligned}
m_{11} &= m_{11}(Z_i, Z_j, Z_k) = \frac{g(Z_i, \theta_0)}{[W_n(X_i) + t^{-2}]^2} d_{\theta_2}(X_i, Z_k) W_n^{-1/2}(X_j) g(Z_j, \theta_0) h^q K_{ij} \\
&\quad + \frac{g(Z_i, \theta_0)}{[W_n(X_i) + t^{-2}]^2} d_{\theta_2}(X_i, Z_k) W_n^{-1/2}(X_j) \{g(Z_j, \theta) - g(Z_j, \theta_0)\} h^q K_{ij} \\
&\quad + \frac{\{g(Z_i, \theta) - g(Z_i, \theta_0)\}}{[W_n(X_i) + t^{-2}]^2} d_{\theta_2}(X_i, Z_k) W_n^{-1/2}(X_j) g(Z_j, \theta_0) h^q K_{ij} \\
&\quad + \frac{\{g(Z_i, \theta) - g(Z_i, \theta_0)\}}{[W_n(X_i) + t^{-2}]^2} d_{\theta_2}(X_i, Z_k) W_n^{-1/2}(X_j) \{g(Z_j, \theta) - g(Z_j, \theta_0)\} h^q K_{ij} \\
&= m_{110} + m_{111} + m_{112} + m_{113}.
\end{aligned}$$

We note that our assumptions ensure that all functions entering the above terms, as indexed by θ , θ_2 , h , and b , are Euclidean. In particular Appendix B shows that the class of functions $x \mapsto W_n^{-1/2}(x)$ is Euclidean as indexed by b for a constant envelope by Assumption E7-(iii).

By convention, for $j = 0, \dots, 3$, we denote by M_{11jn} the U -process based on m_{11j} . The term M_{110n} is a third-order degenerate U -process independent of θ and is a $O_p(n^{-3/2})$ uniformly in θ_2 , h , b , and t . Consider the Hoeffding's decomposition of M_{111n} and note that $\mathbb{E}[m_{111} | Z_i, Z_j] = \mathbb{E}[m_{111} | Z_j, Z_k] = 0$. The third order degenerate U -process in that decomposition is a uniform $o_p(n^{-3/2})$ by Corollary 8 of Sherman (1994). The remaining term to be studied is the degenerate second order U -process defined by the family of functions

$$\frac{g(Z_i, \theta_0) d_{\theta_2}(X_i, Z_k)}{[W_n(X_i) + t^{-2}]^2} \mathbb{E} \left[W_n^{-1/2}(X_j) \{g(Z_j, \theta) - g(Z_j, \theta_0)\} h^q K_{ij} \mid X_i \right].$$

By a Taylor expansion of $\mathbb{E}[g(Z_j, \theta_0) | X_j]$ around θ_0 and Assumption E7, deduce that the uniform rate of convergence of this U -process is $O_p(n^{-1} \|\theta - \theta_0\|)$. Similar arguments apply to $h^q M_{112n}$. For M_{113n} , the

different terms in Hoeffding's decomposition are the third order degenerate U -process, the two degenerate second order U -processes based on $\mathbb{E}[m_{113} | Z_j, Z_k] - \mathbb{E}[m_{113} | Z_k]$ and $\mathbb{E}[m_{113} | Z_i, Z_k] - \mathbb{E}[m_{113} | Z_k]$, and the empirical process based on $\mathbb{E}[m_{113} | Z_k]$. For the third and second order U -processes we proceed as above. For the remaining (centered) empirical process, rely again on Taylor expansions around θ to deduce that its uniform rate of convergence is $O_p(n^{-1/2}\|\theta - \theta_0\|^2)$. Gathering these facts and using $n^{-1}\{\inf \mathcal{H}'_n\}^{-4q} = o_p(1)$ show that

$$\sup_{t \geq 1} \sup_{h, b \in \mathcal{H}'_n} |b^{-q} h^{-q} M_{11n}| = o_p\left(\|\theta - \theta_0\| n^{-1/2} + \|\theta - \theta_0\|^2 + n^{-1}\right)$$

uniformly over θ and θ_2 in $o(1)$ neighborhoods of θ_0 . For $n^{-1}b^{-q}M_{12n}$ and $n^{-1}b^{-q}M_{13n}$, follow a similar (shorter) reasoning to obtain the same order.

Recall that $\|\delta_{\theta_2}(X_i)\| \leq c\|\theta_2 - \theta_0\|^\nu$ for some $\nu > 2/3$ and $c > 0$ uniformly in b and $\theta_2 - \theta_0 = O_p(n^{-1/2})$, and note that

$$\begin{aligned} & \frac{1}{b^q} \frac{1}{(n)_2} \sum_{1 \leq k \leq n} \sum_{i \neq j} \frac{g(Z_i, \theta)}{[W_n(X_i) + t^{-2}]^2} \delta_{\theta_2}(X_i) W_n^{-1/2}(X_j) g(Z_j, \theta) K_{ij} \\ &= \frac{1}{b^q} \frac{1}{(n)_2} \sum_{i \neq j} \frac{g(Z_i, \theta_0)}{[W_n(X_i) + t^{-2}]^2} \delta_{\theta_2}(X_i) W_n^{-1/2}(X_j) g(Z_j, \theta_0) K_{ij} \\ &+ \frac{1}{b^q} \frac{1}{(n)_2} \sum_{i \neq j} \frac{g(Z_i, \theta)}{[W_n(X_i) + t^{-2}]^2} \delta_{\theta_2}(X_i) W_n^{-1/2}(X_j) \{g(Z_j, \theta) - g(Z_j, \theta_0)\} K_{ij} \\ &+ \frac{1}{b^q} \frac{1}{(n)_2} \sum_{i \neq j} \frac{\{g(Z_i, \theta) - g(Z_i, \theta_0)\}}{[W_n(X_i) + t^{-2}]^2} \delta_{\theta_2}(X_i) W_n^{-1/2}(X_j) g(Z_j, \theta_0) K_{ij} \\ &+ \frac{1}{b^q} \frac{1}{(n)_2} \sum_{i \neq j} \frac{\{g(Z_i, \theta) - g(Z_i, \theta_0)\}}{[W_n(X_i) + t^{-2}]^2} \delta_{\theta_2}(X_i) W_n^{-1/2}(X_j) \{g(Z_j, \theta) - g(Z_j, \theta_0)\} K_{ij} \\ &= b^{-q} h^{-q} \left(\tilde{M}_{10n} + \tilde{M}_{11n} + \tilde{M}_{12n} + \tilde{M}_{13n} \right). \end{aligned}$$

Use Hoeffding's decomposition and the last statement of Lemma 1 to deduce that \tilde{M}_{10n} is a uniform $O_p(n^{-1-2\alpha/3})$ for any $\alpha < 1$. Use a Taylor expansion around θ_0 , Hoeffding's decomposition, and Lemma 1 to show that each of \tilde{M}_{1jn} , $j = 1, 2$, is a $O_p(\|\theta - \theta_0\| n^{-1/2-2\alpha/3})$ for any $\alpha < 1$. Use similar arguments to show that $\tilde{M}_{13n} = O_p(\|\theta - \theta_0\|^2 n^{-2\alpha/3})$ for any $\alpha < 1$. Gathering these facts and using $n^{-1}\{\inf \mathcal{H}'_n\}^{-4q/\alpha} = o_p(1)$ for some $\alpha < 1$,

$$\sup_{t \geq 1} \sup_{h, b \in \mathcal{H}'_n} |b^{-q} M_{1n}| = o_p\left(\|\theta - \theta_0\| n^{-1/2} + \|\theta - \theta_0\|^2 + n^{-1}\right)$$

uniformly over θ in a $o(1)$ neighborhoods of θ_0 and θ_2 in a $O_p(n^{-1/2})$ neighborhood of θ_0 . \square

Proof of Corollary 8. Assumption E6, Lemma A.2 of Appendix A, and Lemma B.2 of Appendix B ensure that the class of functions $(x, u) \mapsto W_n^{-1/2}(x - hu) \nabla_{\theta} \tau(x - hu, \theta_0) f(x - hu)$ is Euclidean entrywise for a constant envelope, so that

$$\left\{ x \mapsto \int W_n^{-1/2}(x - hu) \nabla_{\theta} \tau(x - hu, \theta_0) f(x - hu) K(u) du : h, b \in [0, h_0] \right\}$$

is uniformly Euclidean for a constant envelope by Nolan and Pollard (1987, Lemma 20). Reason similarly for the functions $W_n^{-1/2}(\cdot) H_{\theta, \theta} \tau_n^{(k)}(\cdot, \theta_0) f(\cdot)$, $1 \leq k \leq r$ and $W_n^{-1/2}(\cdot) H(\cdot) f(\cdot)$. Use Lemmas 2 and 3 in Section 5.1 and Equation (7) to obtain an asymptotic representation similar to the one of Theorem 2.2. \square

□

Proof of Theorem 3.1. Under H_0 , $\tilde{\theta}_n^R = R(\tilde{\gamma}_n)$ where $\tilde{\gamma}_n = \arg \min_{\gamma} M_n(R(\gamma))$. Let $D = \nabla'_{\gamma} R(\gamma_0)$. From Theorem 2.2's proof, $\sqrt{n}(\tilde{\gamma}_n - \gamma_0) = -(V_n^R)^{-1} B_n + o_p(1)$, where $V_n^R = D' V_n D$ and $B_n = D' A_n$, and

$$\begin{aligned} M_n(\tilde{\theta}_n) - M_n(\theta_0) &= \frac{1}{2} (\tilde{\theta}_n - \theta_0)' V_n (\tilde{\theta}_n - \theta_0) + \frac{1}{\sqrt{n}} A_n' (\tilde{\theta}_n - \theta_0) + o_p(1/n) \\ &= -\frac{1}{2n} A_n' V_n^{-1} A_n + o_p(1/n), \\ M_n(R(\tilde{\gamma}_n)) - M_n(R(\gamma_0)) &= \frac{1}{2} (\tilde{\gamma}_n - \gamma_0)' V_n^R (\tilde{\gamma}_n - \gamma_0) + \frac{1}{\sqrt{n}} B_n' (\tilde{\gamma}_n - \gamma_0) + o_p(1/n), \\ &= -\frac{1}{2n} A_n' D (D' V_n D)^{-1} D' A_n + o_p(1/n) \\ \text{so that } DM_n &= A_n' V_n^{-1/2} \left[I_p - V_n^{1/2} D (D' V_n D)^{-1} D' V_n^{1/2} \right] V_n^{-1/2} A_n + o_p(1) \end{aligned}$$

uniformly in $h \in \mathcal{H}_n$ under H_0 . Our conclusion follows from the extended continuous mapping theorem, see van der Vaart and Wellner (1996, Theorem 1.11.1).

When H_0 does not hold, it follows from the arguments of Theorem 2.1's proof that $M_n(R(\tilde{\gamma}_n)) - M_n(\tilde{\theta}_n)$ converges in probability to a positive constant. □

Proof of Theorem 3.2. Consider $\{(Z_i, w_i)\}$ as the sample and reason as in the proofs of Theorem 2.1 and 2.2, using $\mathbb{E}w^4 < \infty$, to obtain that uniformly in $h \in \mathcal{H}_n$ and over $O_p(1/\sqrt{n})$ neighborhoods of θ_0 ,

$$M_n^*(\theta) - M_n^*(\theta_0) = \frac{1}{2} (\theta - \theta_0)' V_n (\theta - \theta_0) + \frac{1}{\sqrt{n}} A_n^{*'} (\theta - \theta_0) + o_p(1/n),$$

where $V_n = H_{\theta, \theta} \mathbb{E} M_n^*(\theta_0) = H_{\theta, \theta} \mathbb{E} M_n(\theta_0)$ and A_n^* is the centered empirical process based on

$$w g_n'(Z, \theta_0) \left[\int_{\mathbb{R}^q} \nabla'_{\theta} \tau_n(x, \theta_0) f(x) h^{-q} K((X-x)/h) dx \right].$$

Hence $\sqrt{n}(\tilde{\theta}_n^* - \theta_0) + V_n^{-1} A_n^* = o_p(1)$ and $\mathbb{P} \left[\sup_{h \in \mathcal{H}_n} \left| \sqrt{n}(\tilde{\theta}_n^* - \theta_0) + V_n^{-1} A_n^* \right| \geq \varepsilon | Z_1, \dots, Z_n \right] = o_p(1)$ by Markov inequality.

Now, $\sqrt{n}(\tilde{\theta}_n^* - \tilde{\theta}_n) = -V_n^{-1} (A_n^* - A_n) + o_p(1)$, where $A_n^* - A_n$ is the centered empirical process based on

$$(w-1) g_n'(Z, \theta_0) \left[\int_{\mathbb{R}^q} \nabla'_{\theta} \tau_n(x, \theta_0) f(x) h^{-q} K((X-x)/h) dx \right].$$

It is then clear that the process $A_n^* - A_n$ has asymptotically and conditionally upon the initial sample the same distribution as A_n uniformly in h , see e.g. Zhang (2001), so that $\sqrt{n}(\tilde{\theta}_n^* - \tilde{\theta}_n)$ has asymptotically and conditionally upon the initial sample the same distribution as $\sqrt{n}(\tilde{\theta}_n - \theta_0)$ uniformly in h .⁹ Therefore,

⁹Zhang (2001) assumes that w has an exponential distribution, but uses only moment conditions. It is easily seen that our assumptions on w are sufficient.

uniformly in $h \in \mathcal{H}_n$,

$$\begin{aligned}
M_n^*(\tilde{\theta}_n^*) - M_n^*(\theta_0) &= -\frac{1}{2} (\tilde{\theta}_n^* - \theta_0)' V_n (\tilde{\theta}_n^* - \theta_0) + o_p(1/n), \\
M_n^*(\tilde{\theta}_n) - M_n^*(\theta_0) &= \frac{1}{2} (\tilde{\theta}_n - \theta_0)' V_n (\tilde{\theta}_n - \theta_0) - (\tilde{\theta}_n^* - \theta_0)' V_n (\tilde{\theta}_n - \theta_0) + o_p(1/n), \\
\text{and } n [M_n^*(\tilde{\theta}_n^*) - M_n^*(\tilde{\theta}_n)] &= -\frac{1}{2} \sqrt{n} (\tilde{\theta}_n^* - \tilde{\theta}_n)' V_n \sqrt{n} (\tilde{\theta}_n^* - \tilde{\theta}_n) + o_p(1) \\
&= -\frac{1}{2} (A_n^* - A_n)' V_n^{-1} (A_n^* - A_n) + o_p(1).
\end{aligned}$$

As before, this expansion also holds conditionally. Therefore, the latter process has asymptotically and conditionally upon the initial sample the same distribution as $n [M_n(\tilde{\theta}_n) - M_n(\theta_0)]$. \square

Proof of Theorem 3.3. Theorem 3.2's proof deals with the unconstrained problem. A similar reasoning applies to the constrained problem. Proceed as in Theorem 3.1's proof to conclude that DM_n^* has asymptotically and conditionally upon the initial sample the same distribution as DM_n under H_0 uniformly in $h \in \mathcal{H}_n$.

When H_0 does not hold, it follows from Theorem 2.1's proof that $M_n^*(\tilde{\theta}_n^*) - M_n^*(\tilde{\theta}_n) = o_p(1)$ and similarly $M_n^*(R(\tilde{\gamma}_n^*)) - M_n^*(R(\tilde{\gamma}_n)) = o_p(1)$, so that $DM_n^* = o_p(n)$ uniformly in $h \in \mathcal{H}_n$. \square

ADDITIONAL REFERENCES

- HIGHAM, N.J. (2008). *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics: Philadelphia.
- SHERMAN, R.P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61(1)**, 123-137.
- SHERMAN, R.P. (1994b). U -processes in the analysis of a generalized semiparametric regression estimator. *Econometric Theory* **10(2)**, 372-395.
- VAN DE GEER, S. (2000). *Applications of Empirical Process Theory*. Cambridge University Press: New-York.
- ZHANG, D. (2001). Bayesian bootstrap for U -processes, hypothesis tests and convergence of Dirichlet U -processes. *Statist. Sinica* **11(2)**, 463-478.

Table 1: Descriptive statistics in Setup (4.9) $n = 100$

		Homoscedastic case			Heteroscedastic case			
		h	Mean	Median	St. dev.	Mean	Median	St. dev.
Intercept estimator								
TSLS			0.000	-0.001	0.101	0.000	-0.000	0.100
SMD	1		0.000	0.000	0.104	0.000	-0.001	0.090
Eff. IV	.2654		-0.000	0.000	2.186	-0.034	-0.004	2.877
	.4220		-0.003	0.000	0.541	0.076	-0.000	5.845
	.5308		0.041	0.001	3.382	-0.225	0.001	13.97
Eff. SMD	.2654		0.002	-0.001	0.103	-0.000	-0.001	0.090
	.4220		-0.000	-0.001	0.102	-0.000	-0.001	0.090
	.5308		-0.000	-0.001	0.102	-0.000	-0.001	0.090
Slope estimator								
TSLS			-0.001	0.001	0.051	-0.001	0.001	0.071
SMD	1		-0.007	-0.005	0.057	-0.007	-0.004	0.063
Eff. IV	.2654		-0.189	0.005	9.065	0.064	0.007	4.530
	.4220		0.043	0.003	2.296	0.014	0.003	16.75
	.5308		-0.041	0.003	14.14	-0.046	0.005	36.75
Eff. SMD	.2654		-0.004	-0.002	0.055	-0.006	-0.003	0.066
	.5308		-0.004	-0.002	0.054	-0.005	-0.002	0.064
	.5308		-0.004	-0.002	0.053	-0.005	-0.002	0.063

Table 2: Empirical levels in Setup (4.10) $n = 100$

		Case $\mu = 0$		Case $\mu = 1$		
		h	5% level	10% level	5% level	10% level
Wald NLS			7.00	13.00	6.18	11.72
Wald DL			5.48	10.72	4.92	10.08
SMD	2		4.32	10.18	4.70	9.24
			(6.16)	(12.20)	(5.12)	(10.10)
		1	5.98	11.42	5.46	11.08
(6.32)	(12.26)		(5.38)	(10.76)		
0.3		5.96	10.82	5.30	10.64	
	(7.00)	(12.66)	(6.02)	(11.44)		

Percentages using asymptotic χ^2 critical values into parentheses.

Table 3: Empirical levels in Setup (4.11) for $H_0 : \beta_2 = 1$

		$n = 50$			$n = 100$		
		h	5% level	10% level	h	5% level	10% level
FGLS			29.2	35.6		20.7	27.6
Eff. SMD	.1524		4.8	10.8	.1327	4.6	8.4
			(0.0)	(0.0)		(0.0)	(0.0)
		.3049	6.8	12.6	.2654	4.6	9.2
(0.0)	(0.0)			(0.0)	(0.0)		
.6097		9.4	14.0	.5308	5.8	9.8	
	(0.0)	(0.0)		(0.0)	(1.0)		

Percentages using asymptotic χ^2 critical values into parentheses.

Figure 1: Estimators' densities: $X \sim N(0, 1)$, $n = 100$

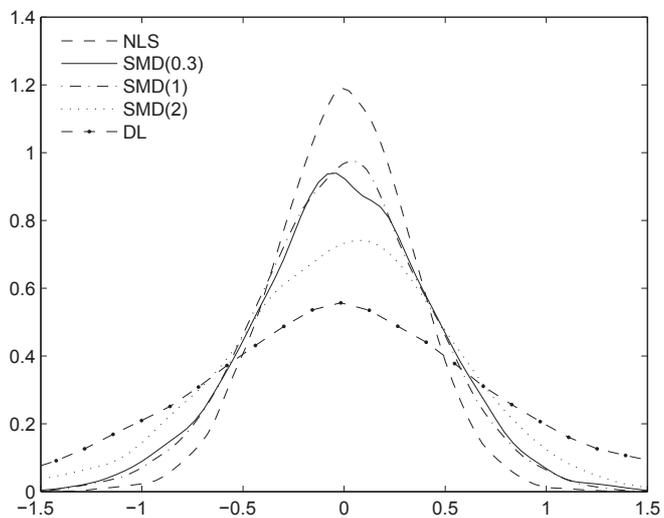


Figure 2: Estimators' densities: $X \sim N(1, 1)$, $n = 100$

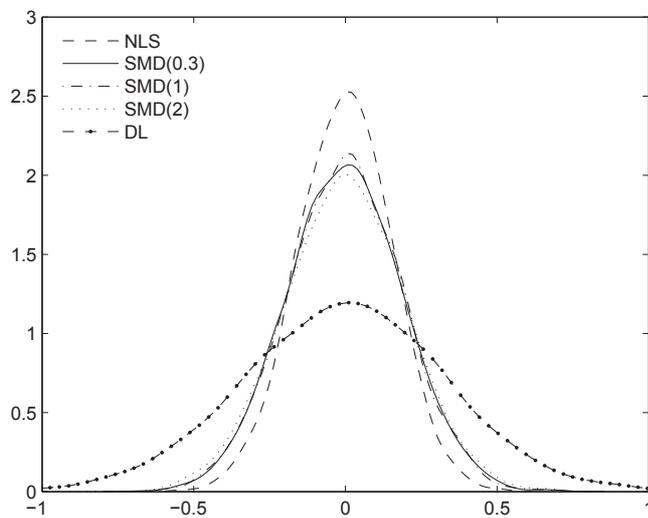


Figure 3: Empirical rejection of tests: $X \sim N(0, 1)$, $n = 100$

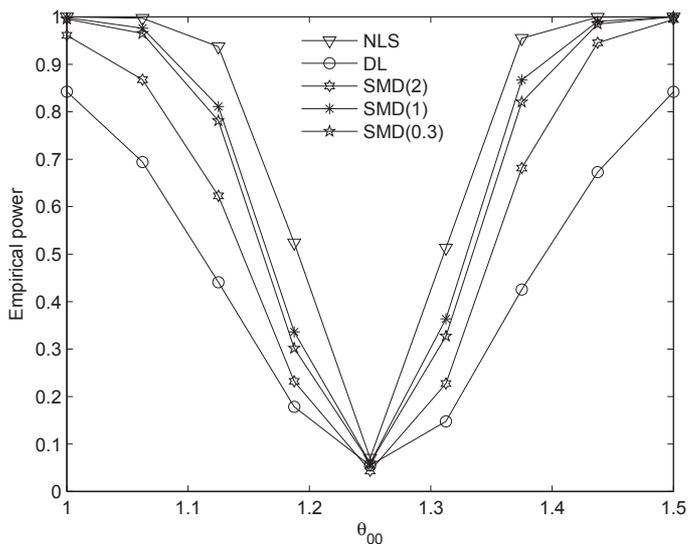


Figure 4: Empirical rejection of tests: $X \sim N(0, 1)$, $n = 100$

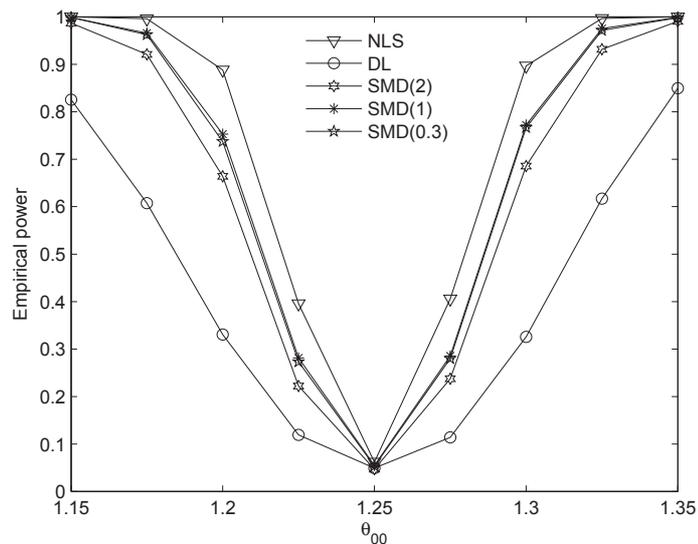


Figure 5: RMSE and bandwidth

