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

Pascal Lavergne, Toulouse School of Economics (GREMAQ)
Valentin Patilea, INSA-IRMAR and CREST-ENSAI

This version: June 2010

Abstract

We propose a new class of estimators for parameters defined by conditional estimating equations. Our generic estimator minimizes a distance criterion based on kernel smoothing. We develop a theory that focuses on uniformity in bandwidth. We establish a $\sqrt{n}$-asymptotic representation of our estimator as a process depending on the bandwidth within a wide range including fixed bandwidths and that applies to misspecified models. We also study an efficient version of our estimator. 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.
MSC classification: Primary 62E20; Secondary 62F40.

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 Allees de Brienne, 31000 TOULOUSE, FRANCE. Emails: pascal_lavergne@univ-tlse1.fr, patilea@ensai.fr

1 Introduction

We focus on estimation and testing of parameters identified by a set of conditional estimating equations such as

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

Here $g(Z,\theta)$ 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 $\theta \in \Theta \subset \mathbb{R}^p$. The problem is semiparametric in nature, since we aim at estimating a finite dimensional parameter θ_0 without specifying entirely the distribution of the variables of interest. Common models that fit into this framework are (i) regression models, where $g(Z,\theta) = Y - \mu(X,\theta)$; (ii) conditional quantile models, where $g(Z,\theta) = \mathbb{I}\left(Y - \mu(X,\theta) \leq 0\right) - \rho$ for a quantile of order ρ; (iii) regression models with a nonlinear transform of the dependent variable, e.g. the Box-Cox transform, where $g(Z,\theta) = h(Y,\lambda) - X'\beta$; (iv) linear or nonlinear simultaneous equations; (v) econometric models of optimizing agents, such as the consumption model of Hansen and Singleton (1982), where $g(Z,\theta) = \beta Y_1 Y_2^{\gamma} - 1$.

Inference may exploit only a finite number of unconditional estimating equations

$$\mathbb{E}\left[A(X)g(Z,\theta_0)\right] = 0 \quad \text{a.s.} \tag{1.2}$$

where $A(X)$ is a user-selected matrix function that may depend on the unknown θ_0. Qin and Lawless (1994) develop an Empirical Likelihood (EL) estimator that is semiparametrically efficient. Hansen (1982) proposes the Generalized Method of Moments (GMM), which minimizes a weighted quadratic form in the empirical analog of the moment conditions, see also Chen, Linton, and Van Keilegom (2003) for extensions. Considering (1.2) instead of (1.1) potentially raises an identification issue. Global identification requires that (1.2) holds uniquely for θ_0, but this is ensured in general only when the estimating equations are linear in parameters. Dominguez and Lobato (2004) provide some examples of nonlinear regression models where a finite number of estimating equations do not identify the parameter for infinitely many distributions of X. A second related issue in considering (1.2) instead of (1.1) is the loss of information. Recent work has thus focused on accounting for (1.1) at the outset for efficiency matters. Some methods rely on increasing the number of considered estimating equations with the sample size, as the minimum distance approach of Ai and Chen (2003), or in generalizations of GMM and EL by Donald, Imbens and Newey (2003) and Hjort, McKeague and Van Keilegom (2009). Other EL-type estimators are based on (1.1) and use nonparametric smoothing, such as Antoine, Bonnal and Renault (2007), Kitamura, Tripathi and Ahn (2004), and Smith (2007a,b). The generalization of GMM to a continuum of estimating equations by Carrasco and Florens (2000) uses a regularization parameter. In every case, global identification and semiparametric efficiency are ensured by assuming that the user-chosen parameter behaves adequatly as the sample size increases (diverges to infinity or converges to zero depending on the methods). Another estimator proposed by Dominguez and Lobato (2004) is based on a Cramer-von-Mises type criterion and does not require any user-chosen parameter, but is in general not semiparametrically efficient.

We propose a new estimator, labeled *Smooth Minimum Distance* (SMD), for parameters characterized by (1.1). Our estimation method is not a generalization of any existing one. It is based on the estimation of the following contrast

$$\mathbb{E}\left[g'(Z_1,\theta)g(Z_2,\theta)h^{-q}K((X_1-X_2)/h)\right] , \qquad (1.3)$$

where $K(\cdot)$ is a kernel function, h a bandwidth parameter, and $g'(\cdot,\theta)$ denotes the transpose of $g(\cdot,\theta)$. As we will show the above contrast is zero if and only if $\theta=\theta_0$, so that the minimizer of an empirical equivalent of (1.3) is a consistent estimator. Our SMD estimator has several theoretical and practical features that makes it valuable compared to previously proposed ones. From a practical viewpoint, our method is versatile, as it applies to all the aforementioned models. In particular, our theory allows for non-differentiable functions $g(\cdot,\theta)$, as is the case for quantile regression models. Our estimator is also easy to implement. EL-type estimators based on (1.1) require nonlinear optimization over many parameters whose number increases with the sample size. In this respect, it is noteworthy that for the EL-type estimators referred to above, only Kitamura et al. (2004) report the results of a simulation study. The SMD estimator does not involve an optimization problem of increasing complexity and thus is more practical. From a theoretical viewpoint, we study the properties of the SMD estimator as a process 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 it is the first of its kind for a semiparametric estimator, for which previous studies adopts simultaneous asymptotics, where the smoothing parameter changes along the sample size. First, we show that SMD is $\sqrt{n}$-consistent and asymptotically normal uniformly in h for a large range of bandwidths including fixed ones, that is bandwidths that are independent of the sample size. Such a result has appealing practical consequences, since it implies that the practical bandwidth's choice does not affect $\sqrt{n}$-consistency and asymptotic normality. Moreover, it directly allows for the use of a data-dependent sequence of bandwidths. Second, while theoretical analysis of previous estimators considers that the bandwidth vanishes and then does not influence the asymptotic estimator's precision, our uniform in bandwidth theory indicates that it can be misleading to ignore the bandwidth's influence even at first-order. Indeed, the bandwidth's choice influence the $\sqrt{n}$-asymptotic variance of the estimator, so that small sample inference based on the assumption that the bandwidth vanishes could be misleading. To account for the bandwidth's influence and thus obtain more reliable inference, we extend a bootstrap method recently proposed by Jin, Ying and Wei (2001) and Bose and Chatterjee (2003) and we show that a bootstrap testing procedure based on SMD is valid uniformly in the bandwidth. Third, our basic estimator can be tailored to obtain a semiparametrically efficient estimator. Specifically, such an estimator obtains when the bandwidth h vanishes and $g(Z, \theta)$ is replaced in the above contrast by $W^{-1/2}(X, \theta_0) g(Z, \theta)$, where $W(X, \theta_0)$ is the density-weighted conditional variance matrix of $g(Z, \theta_0)$. Our efficient estimator thus follows in general from a two-step procedure, where the first step obtains a preliminary SMD estimator, which is consistent irrespective to the bandwidth's choice, and where the second step uses a vanishing h and a kernel estimator of $W(\cdot, \theta_0)$, which involves a second vanishing bandwidth. We establish the efficiency of this estimator uniformly within a large range for the two bandwidths involved. Fourth, the properties of our SMD estimator are robust to potential misspecification, that is when no parameter value is such that (1.1) holds. For instance a nonlinear regression model may be misspecified but still provides valuable information on the best global approximation of the true unknown regression function, see e.g. White (1981). Schennach (2007) shows that the excellent properties of 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.

Our paper is structured as follows. In Section 2, we present our estimation method and develop our uniform-in-bandwidth results concerning consistency and the $\sqrt{n}$-asymptotic representation of the SMD estimator. We then show that these results extend to possibly misspecified models. We also develop a semiparametrically efficient version of our estimator. In Section 3, we investigate a distance-metric procedure for testing restrictions on parameters and the validity of a bootstrap method to approximate the distribution of our test statistic uniformly in the bandwidth. In Section 4, we compare via simulations the SMD estimator to the estimator of Dominguez and Lobato (2004) and the smoothed EL estimator of Kitamura et al. (2004). Our estimator outperforms the first estimator in all our simulation experiments, and performs well relatively to the second. The bootstrap testing approach is found to yield reasonably accurate levels in moderate samples. Proofs are gathered in Section 5. Two Appendices provides some sufficient conditions for our technical assumptions to hold.

2 SMD Estimation

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 $l(\cdot)$, $\mathcal{F}[l](\cdot)$ is its Fourier transform, $\nabla_\theta l(\cdot)$ and $\mathrm{H}_{\theta,\theta} l(\cdot)$ 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.

2.1 The Estimator and its Consistency

The parameter to estimate is uniquely defined through (1.1), that is we assume throughout that $\mathbb{E}[g(Z,\theta)|X] = 0$ a.s. $\Rightarrow \theta = \theta_0$. With at hand independent copies $\{Z_1, \ldots Z_n\}$ from Z, our estimator is $\tilde{\theta}_{n,h} = \arg\min_\Theta M_{n,h}(\theta)$, where

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

with a multivariate kernel $K(\cdot)$ and $h = h_n$ a sequence of bandwidth parameters. Discrete covariates U with finite support could be handled by multiplying each term by the indicator function $\mathbb{I}(U_i = U_j)$ and our proofs would easily adapt, but for the sake of simplicity, we do not formally consider this possibility in what follows.

Our criterion estimates a weighted distance of $\mathbb{E}[g(Z,\theta)|X]$ to zero when h tends to zero, thus providing a first justification for the label *smooth minimum distance*. To understand why our estimator is consistent even when h does not tend to zero, note that

$$\begin{aligned}\mathbb{E}M_{n,h}(\theta) &= \frac{1}{2}\mathbb{E}\left[g'(Z_1,\theta)g(Z_2,\theta)h^{-q}K\left((X_1-X_2)/h\right)\right] \\ &= \frac{1}{2}(2\pi)^{-q/2}\,\mathbb{E}\left[g'(Z_1,\theta)g(Z_2,\theta)\int_{\mathbb{R}^q}\exp\left(it'(X_1-X_2)\right)\mathcal{F}[K](ht)\,dt\right] \\ &= \frac{1}{2}(2\pi)^{q/2}\sum_{k=1}^{r}\left\{\int_{\mathbb{R}^q}\left|\mathcal{F}\left[\mathbb{E}[g^{(k)}(Z,\theta)|X=\cdot]f(\cdot)\right](t)\right|^2\mathcal{F}[K](ht)\,dt\right\},\end{aligned} \tag{2.4}$$

This equation shows that the criterion estimates a weighted L^2-distance of the Fourier transform of $\mathbb{E}[g(Z,\theta)|X=\cdot]f(\cdot)$ to zero, thus providing a second justification for its label. If $\mathcal{F}[K](\cdot)$ is strictly positive on $\mathbb{R}^q$, the criterion's expectation accounts for the Fourier transform of $\mathbb{E}[g(Z,\theta)|X]$ at all frequencies and thus is uniquely minimized at θ_0 independently of h. Indeed, using the unicity of the Fourier transform and Assumption 1,

$$\begin{aligned}\mathbb{E}M_{n,h}(\theta)=0 \;&\Leftrightarrow\; \mathcal{F}\left[\mathbb{E}[g^{(k)}(Z,\theta)|X=\cdot]f(\cdot)\right](t)=0 \quad \forall\, t\in\mathbb{R}^q,\; k=1,\ldots r \\ &\Leftrightarrow\; \mathbb{E}[g(Z,\theta)|X]f(X)=0 \quad \text{a.s.} \Leftrightarrow \theta=\theta_0\,.\end{aligned}$$

A necessary condition for consistency is then the strict positivity of $\mathcal{F}[K](\cdot)$. This is true for instance for products of the triangular, normal, logistic (see Johnson, Kotz, and Balakrishnan, 1995, Section 23.3), Student (including Cauchy, see Hurst, 1995), or Laplace densities. Consistency is thus ensured if $M_{n,h}(\theta)-\mathbb{E}M_{n,h}(\theta)$ converges to zero uniformly in θ and h. We assume that 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, and that the families $\mathcal{G}_k=\{g^{(k)}(\cdot,\theta):\theta\in\Theta\}$, $1\le k\le r$, are Euclidean for a squared integrable envelope. The Euclidean property is a mild one for parametric families of functions. We refer to Nolan and Pollard (1987), Pakes and Pollard (1989), and Sherman (1994a) for the definition and properties of Euclidean families. In particular, our above examples of kernels all satisfy the Euclidean property, see Nolan and Pollard (1987) for sufficient conditions.

For further use, we consider in our proofs a more general version of our estimator, where

$$\begin{aligned}M_{n,h}(\theta) &= \frac{1}{2n(n-1)}\sum_{1\le i\ne j\le n}g'(Z_i,\theta)W_n^{-1/2}(X_i)W_n^{-1/2}(X_j)g(Z_j,\theta)K_{ij}\,, \\ K_{ij} &= \frac{1}{h^q}K\left(\frac{X_i-X_j}{h}\right), \quad 1\le i\ne j\le n\,,\end{aligned} \tag{2.5}$$

and $W_n(\cdot)$ is a sequence of non-random positive definite (p.d.) weighting matrices. Provided the spectral radius of $W_n^{-1/2}(\cdot)$ is uniformly bounded and that $W_n(\cdot)$ converges pointwise to some p.d. matrix $W(\cdot)$, this does not affect our previous reasonning. To avoid technicalities, only our central assumptions are discussed in the text, but the detailed assumptions are spelled out in Section 5.2.

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* $h_0 > 0$.[1]

A few remarks are in order. First, consistency is obtained under more general conditions that the ones imposed for EL-type 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. Second, consistency obtains irrespective of the sequence of matrices $W_n(\cdot)$, assuming it is well-behaved. This will prove useful in Section 2.4, where we allow it to depend on θ_0 and another smoothing parameter b. Third, the strict positivity of $\mathcal{F}[K](\cdot)$ can be weakened to positivity if X has a bounded support. In that case, Equation (2.4) yields that $\mathbb{E}M_{n,h}(\theta) = 0$ iff $\mathcal{F}\left[\mathbb{E}[g^{(k)}(Z,\theta)|X=\cdot]f(\cdot)\right](t) = 0$ for all t in a neighborhood of 0, and this yields $\theta = \theta_0$, using for instance Theorem 1 of Bierens (1982). This allows in particular for higher-order kernels taking negative values, as for instance the normalized sinc kernel whose Fourier transform is a uniform density.

2.2 Uniform in Bandwidth Asymptotic Normality

This section contains our central result. We assume smoothness of the functions $\tau(x,\theta) = \mathbb{E}\left[g(Z,\theta)|X=x\right]$, specifically we assume that all second partial derivatives with respect to the components of θ exist in a neighborhood $\mathcal{N}$ of θ_0 and that $\|\mathrm{H}_{\theta,\theta}\tau^{(k)}(X,\theta) - \mathrm{H}_{\theta,\theta}\tau^{(k)}(X,\theta_0)\| \leq H(X)\|\theta-\theta_0\|^a$, $\forall\, \theta \in \mathcal{N},\ k = 1,\ldots r,$ for some $a \in (0,1]$ and $\tilde{H}(\cdot)$ with $\mathbb{E}\tilde{H}^4 < \infty$. 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

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

and $\phi_{n,0}(z) = \nabla_\theta \tau(x,\theta_0) W_n^{-1}(x) g\left(z,\theta_0\right) f(x)$. Denote by $\{\mathbb{G}_n\phi_{n,h} : h \in [0,h_0]\}$ the sequence of (centered) empirical processes indexed by $\mathcal{F}_n$, that is $\mathbb{G}_n\phi_{n,h} = n^{-1/2}\sum_{i=1}^n \phi_{n,h}(Z_i)$. The

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

process $\{\mathbb{G}_n \phi_{n,h} : h \in [0, h_0]\}$ is expected to weakly converge to a tight zero-mean Gaussian process with covariance function

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

To ensure this convergence, we assume that $\psi_n(\cdot) = \nabla_\theta \tau(\cdot, \theta_0) W_n^{-1/2}(\cdot) f(\cdot)$ satisfies Condition (E) with kernel $K(\cdot)$ for an envelope $\Psi(\cdot)$, which states 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. Sufficient mild conditions on these functions and $K(\cdot)$ 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.[2] Let us finally define $V_{n,h} = \mathrm{H}_{\theta,\theta} \mathbb{E} M_{n,h}(\theta_0)$ and

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

We assume that $\inf_{n,h} \lambda_{\min}(V_{n,h}) > 0$, see Lemma 5.5 for a proof that this holds under very mild conditons.

Theorem 2.2. *Let $h \in \mathcal{H}_n = \left\{h_0 \geq h > 0 : nh^{4q/\alpha} \geq C\right\}$ for arbitrary constants h_0, $C > 0$, and $0 < \alpha < 1$. For an i.i.d. sample, under Assumptions 1–6 and $\inf_{n,h} \lambda_{\min}(V_{n,h}) > 0$, $\sqrt{n}\left(\tilde{\theta}_{n,h} - \theta_0\right) = -V_{n,h}^{-1}\ \mathbb{G}_n \phi_{n,h} + o_p(1)$ 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 $V_{h_1}^{-1} \Delta_{h_1,h_2} V_{h_2}^{-1}$.*

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 does affect its variance. By contrast, existing results on semiparametric estimation rely on a deterministic sequence of bandwidths and assume away the bandwidth's influence by considering that it vanishes as the sample size grows, see the references in the Introduction. As

[2] Condition (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). As we impose Euclidean conditions to investigate the rate of different first and second-order degenerate U-process, we use such conditions throughout.

shown in our proofs section, if one assumes differentiability of $g(Z,\theta)$, a similar result obtains uniformly for h in $\left\{h_0 \geq h > 0 : nh^{2q/\alpha} \geq C\right\}$. 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 moment conditions, and a stronger one is required for the asymptotics of their EL estimator.

The rest of this section provides generalizations of our main Theorem 2.2 in different directions, related to misspecification and efficiency.

2.3 Study Under Misspecification

We here study our estimator under misspecification, that is when there is no θ_0 such that (1.1) holds. As previously argued, this is useful at least as a "robustness" check. Schennach (2007) shows that under misspecification the EL estimator of Qin and Lawless (1994) cannot be $\sqrt{n}$-consistent for its probability limit whenever the functions entering the estimating equations are unbounded, and provides an in-depth discussion. While no study exists on the properties under misspecification of EL-type estimators based on (1.1), one may fear that they exhibit a similar behavior. As we now show, the behavior of the SMD estimator is very similar whether misspecification exists or not.

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. Since for each n the criterion $\mathbb{E}M_{n,h}(\theta)$ is continuous as a function of θ and h under our assumptions, $\bar{\theta}_{n,h}$ can be extended by continuity to

$$\bar{\theta}_{n,0} = \arg\min_{\Theta} \mathbb{E}\left\{\mathbb{E}\left[g'(Z,\theta)|X\right] W_n^{-1}(X) \mathbb{E}\left[g(Z,\theta)|X\right] f(X)\right\} .$$

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 $\bar{\theta}_h = \lim_{n\uparrow\infty} \bar{\theta}_{n,h}$,

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

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 $\inf_{n,h}\lambda_{\min}(\bar{V}_{n,h}) > 0$, $\sqrt{n}\left(\tilde{\theta}_{n,h} - \bar{\theta}_{n,h}\right)$ converges in distribution to a tight random process whose marginal distributions are zero-mean normal with covariance function $\bar{V}_{h_1}^{-1}\bar{\Delta}_{h_1,h_2}\bar{V}_{h_2}^{-1}$.*

2.4 Efficient SMD Estimation

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

Corollary 2.4. *Under the Assumptions of Theorem 2.2, if $W(X) = \mathrm{Var}\left[g(Z,\theta_0)|X\right] f(X)$, then uniformly over $h \in \mathcal{H}'_n = \left\{1/\ln(n+1) \geq h > 0 : nh^{4q/\alpha} \geq C\right\}$, $C > 0$ and $0 < \alpha < 1$ arbitrary, $\sqrt{n}\left(\tilde{\theta}_{n,h} - \theta_0\right)$ is asymptotically $N(0,\Sigma^{-1})$ with*

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

The asymptotic variance Σ^{-1} is the semiparametric efficiency bound characterized by Chamberlain (1987). In classical efficient GMM, see e.g. Newey (1993), the optimal weighting matrix involves not only $\mathrm{Var}\left[g(Z,\theta_0)|X\right]$, but also $\mathbb{E}\left[\nabla_\theta g(Z,\theta_0)|X\right]$, which introduces a supplementary source of variability for its estimation. By contrast, our efficient SMD only requires estimation of $W(X)$. Let $\check{\theta}_n$ be a $\sqrt{n}$-consistent SMD estimate of θ_0, computed for instance by choosing $W_n(\cdot) = I$ and $h = h_0$. Consider the nonparametric estimator of the optimal weight matrix-valued function $\mathrm{Var}[g(Z,\theta_0) \mid X = x]f(x)$ defined as

$$\widehat{W}_n(x,\theta) = \frac{1}{nb^q}\sum_{1\leq k\leq n} g(Z_k,\theta)g'(Z_k,\theta)L((x-X_k)/b) \tag{2.6}$$

where $L(x)$ is a kernel and b is a vanishing bandwidth. By convention, $\widehat{W}_n(x,\theta) = I$ when the right-hand side of the last display 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 will use a generalization of a result from Einmahl and Mason (2005) on kernel estimators to control the behavior of the variance estimator. 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. If $g(Z,\theta)$ is differentiable, a single quasi-Newton step around the preliminary estimator is sufficient. A preliminary estimator for θ_0 may not even be necessary. Consider for instance the case of nonlinear quantile restrictions where $g(Z,\theta) = \mathbb{I}\left[Y - \mu(X,\theta) \leq 0\right] - \rho$. Then $W(x) = \rho(1-\rho)f(x)$, no preliminary estimator is needed, and a one-step efficient estimator obtains, as the ones recently proposed by Otsu (2008) and Komunjer and Vuong (2006).

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. Since in general 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 related or even equal. The main supplementary assumptions needed for our next results are a bounded support for X and some smoothness of $\mathbb{E}\left[g(Z,\theta)g(Z,\theta)|X=x\right]$ 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 applies to our estimator.

Theorem 2.5. *For an i.i.d. sample, under Assumptions 1, 2, E2, E4, 5, and E7,*

$$\sup_{h,b\in\mathcal{H}'_n} \left|\widehat{M}_{n,h,b}(\theta) - M_{n,h,b}(\theta)\right| = o_p\left(n^{-1} + \|\theta-\theta_0\|/\sqrt{n} + \|\theta-\theta_0\|^2\right) \qquad (2.7)$$

uniformly over θ in $o(1)$ neighborhoods of θ_0, where $M_{n,h,b}(\theta)$ is defined as in (2.5) with $W_n(x,\theta_0) = \mathbb{E}\left[\widehat{W}_n(x,\theta_0)\right]$.

This result ensures the equivalence of $\widehat{\theta}_{n,h,b}$ and the estimator $\tilde{\theta}_{n,h}$ with weighting matrix $W_n(\cdot) = \mathbb{E}\left[\widehat{W}_n(\cdot,\theta_0)\right]$. Now we can apply Theorem 2.2 provided we account for the dependence of the weighting matrix on b. Note that the SMD estimator is not affected by boundary effects in the estimation of $W(X)$, since only pointwise convergence of $W_n(x)$ for x in the interior of the support of X is necessary for our Theorem 2.2 to apply.

Corollary 2.6. *Under the assumptions of Theorem 2.5 and E6, $\sqrt{n}\left(\widehat{\theta}_{n,h,b} - \theta_0\right)$ is asymptotically $N(0, \Sigma^{-1})$ uniformly in $h, b \in \mathcal{H}'_n$.*

3 SMD-Based Testing for Parameter Restrictions

In this section, we focus on inference on parameters through testing. Our primary aim is to account for the bandwidth's influence in testing. Hence we do not assume that the SMD estimator is efficient, 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 case where all parameters values are completely determined under H_0. The constrained SMD estimator is $\tilde{\theta}^R_{n,h} = \arg\min_{\theta\in\Theta,\theta=R(\gamma)} M_{n,h}(\theta)$. A distance metric statistic for testing H_0 is

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

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

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

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

Theorem 3.1. *Under the assumptions of Theorem 2.2* *i. under* H_0, $DM_{n,h} - (\mathbb{G}_n\phi_{n,h})' \Lambda_{n.h} (\mathbb{G}_n\phi_{n,h}) = o_p(1)$ *uniformly in* $h \in \mathcal{H}_n$.

ii. if H_0 *does not hold* $\mathbb{P}[n^{-1}DM_{n,h} > c] \to 1$ *uniformly in* $h \in \mathcal{H}_n$ *for some* $c > 0$.

The process $(\mathbb{G}_n\phi_{n,h})' \Lambda_{n.h} (\mathbb{G}_n\phi_{n,h})$ is asymptotically tight and for each h has an asymptotic weighted sum of chi-squares distribution, see Johnson, Kotz, and Balakrishnan (1995), where the weights λ_h are the (positive) eigenvalues of $\Lambda_h = \lim_{n\uparrow\infty} \Lambda_{n,h}$. Hence we label this process an *asymptotically tight weighted sum of chi-squares* process. The distribution of our distance-metric statistic is thus in general non-pivotal. The usual chi-square distribution reappears when we use an efficient estimator, that is for the optimal weighting matrix and h tending to zero. However the general distribution obtained without imposing this restriction likely provides a more accurate approximation because it accounts for the influence of the smoothing parameter. Determining critical values requires estimation of Λ_h. Under differentiability of $g(Z,\theta)$, we can estimate V_h and $\Delta_{h,h}$ respectively by

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

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

where $\widehat{W}_n(X_k,\tilde{\theta}_{n,h})$ is the nonparametric estimator defined by (2.6). Consistency of these 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). In what follows, we shall propose another route based on the bootstrap.

3.2 Bootstrapping SMD

Bootstrapping is popular to approximate the distribution of statistics when asymptotics may not reflect accurately their behavior in small or moderate samples. For testing restrictions on parameters defined by (1.1), application of bootstrap would require to generate resamples with the same values of X, but new observations for Y. In addition, the bootstrap samples should mimic the behavior of the data under the null hypothesis. This can be done easily in simple cases, e.g. wild bootstrap in regression models, and has been shown to give reliable approximations in many situations. In general however, generating bootstrap samples may be difficult or even infeasible: in simultaneous equations systems that are nonlinear in the variables Y, a reduced form may not be available or unique. We here propose a simple method that allows to circumvent these difficulties if they appear, that applies generally and is easy to implement. This method has been proposed by Jin, Ying and Wei (2001) and Bose and Chatterjee (2003), see also Chatterjee and Bose (2005) for a similar method applied to Z-estimators and Chen and Pouzo (2009) for sieve minimum distance estimators. However, they impose conditions that do not hold in our context. More crucially, they do not investigate the use of this method for testing.

Instead of resampling observations, we perturb the objective function and recompute our test statistic using this perturbed objective function. Consider n independent identical copies $w_i, i = 1, \ldots n$, of a known positive random variable w with $\mathbb{E}(w) = \mathrm{Var}(w) = 1$ and $\mathbb{E}w^4 < \infty$. Define the new perturbed criterion as

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

We can then compute $\tilde{\theta}^*_{n,h} = \arg\min_\theta M^*_{n,h}(\theta)$. The method consists in generating a large number of sample draws from the same distribution w. 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*

$\sup_{h\in\mathcal{H}_n}\sup_u \left|\mathbb{P}\left[\sqrt{n}\left(\tilde{\theta}^*_{n,h} - \tilde{\theta}_{n,h}\right) \leq u | Z_1, \ldots Z_n\right] - \mathbb{P}\left[\sqrt{n}\left(\tilde{\theta}_{n,h} - \theta_0\right) \leq u\right]\right| = o_p(1).$

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 $\mathbb{E}\left(M^*_{n,h}(\theta)|Z_1, \ldots Z_n\right) = M_{n,h}(\theta)$ 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 section, the perturbed and the original criterion have a similar quadratic expansion in θ. Therefore, the distribution of $n\left(M^*_{n,h}(\tilde{\theta}^*_n) - M^*_{n,h}(\tilde{\theta}_{n,h})\right)$ 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 bootstrap method to approximate the distribution of $\tilde{\theta}_{n,h}$, and in particular can be used to determine confidence intervals for a single parameter. It is likely that a studentized version would yield a more accurate approximation, but such an investigation is beyond the scope of this paper. Since confidence regions are sets of values that are not rejected by a test, our following procedure can be used to construct such regions.

The determination of bootstrap critical values for hypothesis testing is based on Part (ii) of the above result. Consider the decomposition

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

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. Hence we repeat estimation under the constraint (3.8), that is $\tilde{\theta}^{R*}_{n,h} = \arg\min_{\theta,\theta=R(\gamma)} M^*_{n,h}(\theta)$, and we define the bootstrap distance metric test statistic as

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

Theorem 3.3. *Under the Assumptions of Theorem 2.2, then conditionally on the sample and uniformly over* $h \in \mathcal{H}_n$

i. Under H_0, $DM^*_{n,h}$ *has asymptotically the same distribution as* $DM_{n,h}$,

ii. When H_0 *does not hold,* $DM^*_{n,h} = o_p(n)$.

The last part suffices to obtain a consistent test, since $DM_{n,h}$ diverges at rate n from Theorem 3.1. However, under suitable assumptions, one could 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 similar local power than the asymptotic one.

4 Small sample study

The first set of experiments aimed at evaluating the small sample behavior of SMD for a bandwidth that does not change with the sample size. We used the setup considered by

Dominguez and Lobato (2004, hereafter DL), where

$$Y = \theta_0^2 X + \theta_0 X^2 + \varepsilon, \tag{4.9}$$

with $\theta_0 = 5/4$, $X \sim N(\mu, 1)$, and $\varepsilon \sim N(0,1)$ independently of X. DL's estimator minimizes the criterion

$$\frac{1}{n^3}\sum_{k=1}^n \left[\sum_{i=1}^n g(Z_i,\theta)\mathbb{I}(X_i \le X_k)\right]^2 = \frac{1}{n^2}\sum_{i,j=1}^n g(Z_i,\theta)g(Z_j,\theta)\left[\frac{1}{n}\sum_{k=1}^n \mathbb{I}(X_i \le X_k)\mathbb{I}(X_j \le X_k)\right], \tag{4.10}$$

where $g(Z,\theta) = Y - \theta_2 X - \theta X^2$. Since it does not depend on a bandwidth, we compared it to SMD when h does not vary with n. For SMD, we used a Gaussian kernel and $W_n = I$ with different bandwidths. All results are based on 5000 replications. Figures 1 to 4 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 already accurate for the small sample sizes considered here. Our SMD estimator outperforms DL's one for the range of considered bandwidths. Since the weight in (4.10) depends on all observations X_i and can vary from 1 to $1/n$, this yields more variability in the criterion and in turn more variability in estimation. Moreoever, with increasing sample size, the performances of DL's estimator do not improve. We also implemented the efficient SMD with $h = b = 0.3$ and compared it Nonlinear Least-Squares (NLS), which is efficient given that the error term is homoscedastic. Though SMD does not use this knowledge, it compares well with NLS and improves when n increases, even though the bandwidth does not adapt to the sample size.

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

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

with $\beta_1 = \beta_2 = 1$, $\ln X \sim N(0,1)$, and ε is normally distributed. KTA (2004) 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 also computed the feasible GLS estimator based on the knowledge of the variance functional form, for details see Newey (1993). We considered the efficient SMD with $h = b$, and we now adapt the bandwidth to the sample size as in KTA. 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.[3] 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. 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.

We then investigated the behavior of our bootstrap distance-metric statistic under the null hypothesis. We did not explore the power properties of our test, such a study is left for future research. We ran 500 replications for sample sizes $n = 50$ and 100, and for each replication 99 bootstrapped statistics were computed to determine the critical value. For bootstrapping, we used the two-point distribution defined through

$$\mathbb{P}\left[w = \frac{3-\sqrt{5}}{2}\right] = \frac{5+\sqrt{5}}{10} \qquad \text{and } \mathbb{P}\left[w = \frac{3+\sqrt{5}}{2}\right] = \frac{5-\sqrt{5}}{10}\,.$$

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). Table 1 reports empirical levels of the tests. In all cases, the level accuracy increases when the sample size increases. For Model (4.10), the empirical level accuracy is reasonable for $n = 50$, but for $X \sim N(1,1)$ where the test overrejects, and close to the nominal one for $n = 100$, while the Wald test based on the NLS estimator always over rejects. This indicates that our bootstrap method does a relatively good job in approximating the non-pivotal distribution of the test statistic. For Model (4.11), we focus on small to medium

[3]RMSE figures for SEL were kindly provided by Yuichi Kitamura.

Table 1: Rejection percentages of bootstrap test

	$n=50$			$n=100$		
	h	5% level	10% level	h	5% level	10% level
Model (4.10) $X \sim N(0,1)$ $H_0: \theta_0 = 5/4$						
NLS		10.3	16.3		7.0	13.0
SMD	2	4.0	8.2	2	4.8	9.4
	1	4.4	11.8	1	4.8	10.4
	.3	5.0	12.0	.3	5.0	9.4
Model (4.10) $X \sim N(1,1)$ $H_0: \theta_0 = 5/4$						
NLS		8.1	14.3		6.2	11.7
SMD	2	6.8	13.0	2	6.0	9.8
	1	7.0	13.8	1	5.4	11.0
	.3	8.0	13.4	.3	5.6	10.2
Model (4.11) $H_0: \beta_2 = 1$						
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.

bandwidths, that is $h = cn^{-1/5}$ with $c = 1/3, 2/3, 4/3$. The results exhibits a similar pattern. For a medium bandwidth, the test over rejects, but this phenomenon fades out with increasing sample size. Using asymptotic critical values from the chi-square distribution with one degree of freedom for our test yields rejection percentages between 0 and 1% and thus does not constitute a credible alternative in small samples. Tests based on FGLS (Wald and LR tests yield identical results) are severely oversized and are then not reliable either. To sum up, our SMD estimator performs well in our simulation experiments, is competitive with SEL while it exhibits a different behavior with respect to the bandwidth, and our bootstrap technique yields reliable test levels for moderate sample sizes.

5 Proofs

5.1 Assumptions

5.1.1 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 $\mathbb{E}\left[g(Z,\theta)|X\right] = 0$ a.s. $\Rightarrow \theta = \theta_0$. (iii) θ_0 belongs to the interior of Θ.*

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 $L^1 \cap L^2$. 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 $\sup_\theta \| \mathbb{E}[g(Z,\theta) \mid X = \cdot] \| f(\cdot)$ is in $L^1 \cap L^2$. For all x, the map $\theta \mapsto \mathbb{E}[g(Z,\theta) \mid X = x]$ 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 $\mathrm{Var}\,[g(Z,\theta_0)|X = \cdot]\, f(\cdot)$ 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, see Lemma 2.14-(ii) of Pakes and Pollard (1989).

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*

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

Assumption 5 is implied by the following Condition 1.

Condition 1. *(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*

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

Under Condition 1, $\mathbb{E}\,\|g(Z,\theta) - g(Z,\theta_0)\|^2 = O(\|\theta - \theta_0\|^2)$, 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 1 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 $\mathrm{H}_{\theta,\theta}\tau_n^{(k)}(\cdot,\theta_0) f(\cdot)$, $1 \leq k \leq r$ 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 $a \geq 4/3$.*

5.1.2 Assumptions for 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

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

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

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

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

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

Assumption M5. *(i) For any x, all second partial derivatives of $\tau(x, \cdot) = \mathbb{E}\left[g(Z, \cdot) | X = x\right]$ 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*

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

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 $a \ge 4$. (ii) The components of $\mathrm{H}_{\theta,\theta} \tau_n^{(k)}(\cdot, \cdot) f(\cdot)$, $1 \le k \le r$, 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 $a \ge 4/3$.*

5.1.3 Assumptions for Theorem 2.5

Assumption E2. *(i) $L(\cdot)$ is a density of bounded variation with bounded support and is strictly positive around the origin. (ii) The class of functions $(x, \bar{x}) \mapsto L((x - \bar{x})/h)$, $x, \bar{x} \in \mathbb{R}^q$, $h > 0$, is Euclidean for a constant envelope.*

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

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) $W(\cdot) = \mathbb{E}[g(Z, \theta_0)g'(Z, \theta_0) \mid X = \cdot]f(\cdot)$ 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 $\omega^2(\cdot, \theta) = \mathbb{E}[g(Z, \theta)g'(Z, \theta) \mid X = \cdot]$. 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 $W_n(\cdot) = \mathbb{E}\left[\widehat{W}_n(\cdot, \theta_0)\right]$ 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 $\nabla_\theta \tau(\cdot, \theta_0)f(\cdot)$, $\mathrm{H}_{\theta,\theta}\tau^{(k)}(\cdot, \theta_0)f(\cdot)$, $1 \leq k \leq r$ and $H(\cdot)f(\cdot)$ is Hölder continuous on D, with possibly different exponents.*

5.2 Lemmas

In what follows we adopt the notations of Sherman (1993, 1994a) 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 (1994a).

Lemma 5.1. *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 $h > 0$, $n^{k/2}U_n^k f(\cdot, \theta, h) = \|\theta - \theta_{n,h}\|^{a\alpha/2}O_p(1)$ for any $0 < \alpha < 1$.

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

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 5.2. *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 $h \in \mathcal{H}_n$

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

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 5.3. *Let $\theta_{n,h}$ be as in Lemma 5.2. 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 $\bar{\theta}_{n,h}$,*

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

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 $\sqrt{n}\left(\theta_{n,h} - \bar{\theta}_{n,h}\right) + V_{n,h}^{-1} A_{n,h} = o_p(1)$ uniformly in $h \in \mathcal{H}_n$.

Lemma 5.4. *Under Assumptions 3 and 4(v), $\sup_{n,h} \lambda_{\max}(V_{n,h}) < \infty$.*

Lemma 5.5. *Under Assumptions 3 and 4(v), if* $\mathcal{F}[K](ht) \geq \mathcal{F}[K](h_0 t)$ $\forall\ t \in \mathbb{R}^q$, $\forall h \in [0, h_0]$, $\mathrm{H}_{\theta,\theta}\mathbb{E}[\tau'(X,\theta_0)\tau(X,\theta_0)]$ *positive definite implies* $\liminf_n \inf_h \lambda_{\min}(V_{n,h}) > 0$.

Note that if we knew the form of $\tau(X,\theta)$, we would minimize $\mathbb{E}[\tau'(X,\theta)\tau(X,\theta)]$ to obtain θ_0. The positive definiteness of the Hessian at θ_0 is thus a quite natural condition. The assumption on the kernel is fulfilled for instance by products of normal, logistic, Laplace, and Student densities.

Proof of Lemmas 5.4 and 5.5. For any n, h, and $a \in \mathbb{R}^p$,

$$\begin{aligned} a'V_{n,h}a &= \mathbb{E}\left[a'\nabla_\theta \tau_n(X_1,\theta_0)\nabla'_\theta \tau_n(X_2,\theta_0)a\ h^{-q}K\left(\frac{X_1 - X_2}{h}\right)\right] \\ &= (2\pi)^{q/2}\left\{\int_{\mathbb{R}^q}\sum_{k=1}^{r}\left|\mathcal{F}\left[a'\nabla_\theta \tau_n^{(k)}(\cdot,\theta_0)f(\cdot)\right](t)\right|^2 \mathcal{F}[K](ht)\,dt\right\}, \end{aligned} \tag{5.12}$$

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

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

If $\mathcal{F}[K](ht) \geq \mathcal{F}[K](h_0 t)$ 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 (5.12). 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 (5.12) and the unicity of the Fourier transform,

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

But $a'\mathrm{H}_{\theta,\theta}\mathbb{E}[\tau'(X,\theta_0)\tau(X,\theta_0)]\,a = 2\mathbb{E}[a'\nabla_\theta \tau(X,\theta_0)\nabla'_\theta \tau(X,\theta_0)a] = 0$ iff $a = 0$. Thus $\lambda_{\min}(V_{h_0}) > 0$. □

5.3 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 (2.4) yields

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

as $W_n(X)$ is positive definite. Since $\mathbb{E}M_n(\theta)$ is continuous in θ from Assumption 4-(ii) as well as in h, see (2.4), we have that $\forall \varepsilon > 0$, $\exists \mu > 0$ such that $\inf_{\|\theta-\theta_0\|\geq\varepsilon, 0\leq h\leq h_0} \mathbb{E}M_n(\theta) \geq \mu$. The family of functions $\{g'(Z_1,\theta)W_n^{-1/2}(X_1)W_n^{-1/2}(X_2)g(Z_2,\theta)K\left((X_1 - X_2)/h\right) : \theta \in \Theta, h > 0\}$ 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 (1994a), $\sup_{\theta\in\Theta,h>0} |h^q M_n(\theta) - \mathbb{E}h^q M_n(\theta)| = O_{\mathbb{P}}(n^{-1/2})$. 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,

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

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 $M_n(\tilde\theta_n) \leq M_n(\theta_0)$, it follows that $\sup_{h\in\bar{\mathcal{H}}_n} \|\tilde\theta_n - \theta_0\| < \varepsilon$ with probability tending to one. □

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). □

Proof of Theorem 2.3. (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 $\sup_{h\in\bar{\mathcal{H}}_n} \|\tilde\theta_n - \bar\theta_n\| = o_p(1)$.
(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 $h \in \mathcal{H}_n$

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

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 $\tilde l_\theta(Z_i) = \mathbb{E}[l_\theta(Z_i,Z_j) \mid Z_i] + \mathbb{E}[l_\theta(Z_i,Z_j) \mid Z_j] - 2\mathbb{E}[l_\theta(Z_i,Z_j)]$,

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

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,

$$
\begin{aligned}
2\mathbb{E}[l_{1\theta}(Z_i, Z_j) \mid Z_i] &= g_n'(Z_i, \bar{\theta}_n)\mathbb{E}\left[\left(g_n(Z, \theta) - g_n(Z, \bar{\theta}_n)\right) h^{-q} K\left((X_i - X)/h\right)|Z_i\right] \\
&= g_n'(Z_i, \bar{\theta}_n)\left[\int_{\mathbb{R}^q} \nabla_\theta' \tau_n(x, \bar{\theta}_n) f(x) h^{-q} K\left((X_i - x)/h\right) dx\right]\left(\theta - \bar{\theta}_n\right) \quad (5.13)\\
&+ \frac{1}{2} g_n'(Z_i, \bar{\theta}_n) \sum_{k,l=1}^{p} \left(\theta^{(k)} - \bar{\theta}_n^{(k)}\right)\left(\theta^{(l)} - \bar{\theta}_n^{(l)}\right) \\
&\left[\int_{\mathbb{R}^q} \mathrm{H}_{\theta^{(k)}\theta^{(l)}} \tau_n(x, \bar{\theta}_n) f(x) h^{-q} K\left((X_i - x)/h\right) dx\right] + R_{1n}(Z_i, \theta) \quad (5.14)
\end{aligned}
$$

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

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 $\widetilde{\phi}_{n,h}'(z)$ indexed by h in (5.13) is uniformly Euclidean for a squared-integrable envelope. Hence $A_n' = \bar{\mathbb{G}}_n \widetilde{\phi}_{n,h}' = O_p(1)$ uniformly in θ and $h \in [0, h_0]$ by Corollary 4 of Sherman (1994a). Similarly, the family of functions in (5.14) 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

$$
\begin{aligned}
2\mathbb{E}[l_{3\theta}(Z_i, Z_j) \mid Z_i] &= \left(g_n(Z_i, \theta) - g_n(Z_i, \bar{\theta}_n)\right)' \mathbb{E}\left[\left(g_n(Z, \theta) - g_n(Z, \bar{\theta}_n)\right) h^{-q} K\left((X_i - X)/h\right)|Z_i\right] \\
&= \left(g_n(Z_i, \theta) - g_n(Z_i, \bar{\theta}_n)\right)' \\
&\left[\int_{\mathbb{R}^q} \nabla_\theta' \tau_n(x, \bar{\theta}_n) f(x) h^{-q} K\left((X_i - x)/h\right) dx\right]\left(\theta - \bar{\theta}_n\right) \quad (5.15)\\
&+ \frac{1}{2}\left(g_n(Z_i, \theta) - g_n(Z_i, \bar{\theta}_n)\right)' \sum_{k,l=1}^{p}\left(\theta^{(k)} - \bar{\theta}_n^{(k)}\right)\left(\theta^{(l)} - \bar{\theta}_n^{(l)}\right) \\
&\left[\int_{\mathbb{R}^q} \mathrm{H}_{\theta^{(k)}\theta^{(l)}} \tau_n(x, \bar{\theta}_n) f(x) h^{-q} K\left((X_i - x)/h\right) dx\right] + R_{3n}(Z_i, \theta).
\end{aligned}
$$

Since the function in (5.15) is such that

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

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

$$\mathbb{P}_n \tilde{l}_\theta = \frac{1}{\sqrt{n}} A_n' \left(\theta - \bar{\theta}_n\right) + \|\theta - \bar{\theta}_n\|^2 o_p(1) \,, \tag{5.16}$$

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 $\theta \in \mathcal{N}$, $\mathbb{E} h^{2q} l_\theta^2(Z_i, Z_j) = \mathbb{E}\left[\left(g_n'(Z_i,\theta) g_n(Z_j,\theta) - g_n'(Z_i,\bar{\theta}_n) g_n(Z_j,\bar{\theta}_n)\right) K\left(\left(X_i - X_j\right)/h\right)\right]^2$. Since $K(\cdot)$ is bounded, the Z_i are independent, and for any $a_1, ..., a_r \in \mathbb{R}$, $(a_1 + ... + a_r)^2 \le r(a_1^2 + ... + a_r^2)$, 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 5.1 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

$$\sup_{h \in \mathcal{H}_n} |U_n \bar{l}_\theta| = \|\theta - \bar{\theta}_n\|^{\alpha/2} O_p(\sup_{h \in \mathcal{H}_n} n^{-1} h^{-q}) + O_p(\sup_{h \in \mathcal{H}_n} n^{-1-\alpha/4} h^{-q}) \,. \tag{5.17}$$

Choose $\alpha < 1$ such that $n h^{4q/\alpha} \ge C$ 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 (5.16) and (5.17) in conjunction with Lemma 5.2 to obtain $\|\tilde{\theta}_n - \bar{\theta}_n\| = O_p(\varepsilon_{0n}^{1/2})$. Plug in this result in (5.17), 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 (5.17) is a $O_p(n^{-1})$. Apply then again Lemma 5.2 to conclude that $\|\tilde{\theta}_n - \bar{\theta}_n\| = O_p(n^{-1/2})$.

Remark that under Condition 1, Equation (5.17) 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 $n h^{\frac{2q}{\alpha}} \ge C$ 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 $\alpha \le \alpha' < 1$, 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 (5.16) to obtain

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

uniformly over $O_p(1/\sqrt{n})$ neighborhoods of $\bar{\theta}_n$ and in $h \in \mathcal{H}_n$. Conclude from Lemma 5.3 that $\sqrt{n}\left(\tilde{\theta}_n - \bar{\theta}_n\right) + \bar{V}_n^{-1} A_n = o_p(1)$.

(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 $\sup_{|h_1-h_2|<\delta}\mathbb{E}\left\|\phi_{n,h_1}(Z)-\phi_{n,h_2}(Z)\right\|^2\to 0$ whenever $\delta\to 0$. Let $\omega_n^2(X,\theta_0)=\mathbb{E}\left[g_n^2(Z,\theta_0)|X\right]$. Proceed as in the consistency proof to show that

$$\begin{aligned}\mathbb{E}\left[\phi'_{n,h_1}(Z)\phi_{n,h_2}(Z)\right] &= (2\pi)^{q/2}\int_{\mathbb{R}^q}\int_{\mathbb{R}^q}\mathcal{F}\left[\nabla'_\theta\tau_n(\cdot,\theta_0))f(\cdot)\right](-t)\mathcal{F}\left[\omega_n^2(\cdot,\theta_0)f(\cdot)\right](t-u)\\ &\qquad \mathcal{F}\left[\nabla_\theta\tau_n(\cdot,\theta_0))f(\cdot)\right](u)\mathcal{F}\left[K\right](h_1t)\mathcal{F}\left[K\right](h_2u)\,dt\,du\,.\end{aligned}$$

Hence, $\mathbb{E}\left\|\phi_{n,h_1}(Z)-\phi_{n,h_2}(Z)\right\|^2$

$$\begin{aligned}&= (2\pi)^{q/2}\int_{\mathbb{R}^q}\int_{\mathbb{R}^q}\mathcal{F}\left[\nabla'_\theta\tau_n(\cdot,\theta_0))f(\cdot)\right](-t)\mathcal{F}\left[\omega_n^2(\cdot,\theta_0)f(\cdot)\right](t-u)\mathcal{F}\left[\nabla_\theta\tau_n(\cdot,\theta_0)f(\cdot)\right](u)\\ &\qquad \left[\mathcal{F}\left[K\right](h_1t)\mathcal{F}\left[K\right](h_1u)-2\mathcal{F}\left[K\right](h_1t)\mathcal{F}\left[K\right](h_2u)+\mathcal{F}\left[K\right](h_2t)\mathcal{F}\left[K\right](h_2u)\right]\,dt\,du\,.\end{aligned}$$

Use the uniform continuity of $\mathcal{F}\left[K\right](\cdot)$, 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 $\sup_{|h_1-h_2|<\delta,\|\theta_1-\theta_2\|<\delta}\mathbb{E}\left\|\bar{\phi}_{n,h_1}(Z,\theta_1)-\bar{\phi}_{n,h_2}(Z,\theta_2)\right\|^2\to 0$ whenever $\delta\to 0$. When only h varies in this expression, we can apply our previous reasoning, provided we use $\omega_n^2(X,\theta_1,\theta_2)=\mathbb{E}\left[g_n(Z,\theta_1)g_n(Z,\theta_2)|X\right]$ 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. □

Proof of Corollary 2.4. 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

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

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

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

Proof of Theorem 2.5. 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 5.6. *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*

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

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

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

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

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

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

$$\begin{aligned}
(A+B)^{-1} &= A^{-1} - A^{-1}\left(I + BA^{-1}\right)^{-1} BA^{-1}\,, \qquad (5.18)\\
\text{and} \quad \left[I + t^2D\left(t^2A+I\right)^{-1}\right]^{-1} &= I - t^2D\left(t^2A+I\right)^{-1} + R\,,\\
\text{with} \quad \|R\| \leq \sqrt{q}\,\|R\|_2 &\leq \frac{\sqrt{q}\left\|t^2D\left(t^2A+I\right)^{-1}\right\|_2^2}{1 - \left\|t^2D\left(t^2A+I\right)^{-1}\right\|_2}\\
&\leq \sqrt{q}\,\|D\|_2^2\left[\frac{t^2}{1+t^2\lambda_{min}(A)}\right]^2\left[1 - \frac{t^2\,\|D\|_2}{1+t^2\lambda_{min}(A)}\right]^{-1}\\
&\leq k(c)\,\|D\|_2^2 \leq k(c)\,\|D\|^2\,.
\end{aligned}$$

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 (5.18) with A replaced by t^2A+I and $B=t^2D$ to deduce that

$$\begin{aligned}(A+D)^{-1/2}-A^{-1/2} &= -\frac{2}{\pi}\int_0^\infty t^2\left(t^2A+I\right)^{-1}D\left(t^2A+I\right)^{-1}dt \\ &\quad+\frac{2}{\pi}\int_0^\infty t^4\left(t^2A+I\right)^{-1}D\left(t^2A+I\right)^{-1}D\left(t^2A+I\right)^{-1}dt \\ &\quad-\frac{2}{\pi}\int_0^\infty t^2\left(t^2A+I\right)^{-1}RD\left(t^2A+I\right)^{-1}dt\,, \qquad (5.19)\end{aligned}$$

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

This implies that for some constant C the last integral in (5.19) is bounded by

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

Step 3 consists in applying Identity (5.19) 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 $0<\lambda\le\inf_{x,n}\lambda_{\min}(W_n(x))$ 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,

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

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 $\nu>2/3$, $\|\mathbb{E}\left[\left(\omega^2(X,\theta_2)-\omega^2(X,\theta_0)\right)b^{-q}L((X-x)/b)\right]\|\le c\|\theta_2-\theta_0\|^\nu\|\mathbb{E}\left[b^{-q}L((X-x)/b)\right]\|\le C\|\theta_2-\theta_0\|^\nu=o(n^{-1/3})$ uniformly in θ_2 in a $O(n^{-1/2})$ neighborhood of θ_0. In the last display,

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

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

Strictly speaking, we should separate the integrals on $[0,1)$ and $[1,\infty)$ in the following. Specifically, for $t\in[0,1)$, the terms $\left[W_n(\cdot)+t^{-2}I\right]^{-1}$ should be replaced by $\left[t^2W_n(\cdot)+I\right]^{-1}$, 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

$$
\sup_{t,s\geq 1}\sup_{b,h\in\mathcal{H}_n}\left\{\|M_{1n}\|+\|M_{2n}\|+\|M_{3n}\|\right\} = o_p\left(n^{-1}+\|\theta-\theta_0\|/\sqrt{n}+\|\theta-\theta_0\|^2\right)\,, \tag{5.20}
$$

which implies (2.7). The terms M_{1n}, M_{2n} and M_{3n} involve the family of matrix-valued functions

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

For $t\in[0,1)$, the first family has to be replaced by $\left\{\left[t^2W_n(\cdot)+I\right]^{-1}: b\in\mathcal{H}_n',\, t\in[0,1)\right\}$. 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 (5.20) 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

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

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

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

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

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

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

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 $j=0,\ldots 3$, 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 $\mathbb{E}[m_{111}\mid Z_i,Z_j]=\mathbb{E}[m_{111}\mid Z_j,Z_k]=0$. The third order degenerate U-process in that decomposition is a uniform $o_p(n^{-3/2})$ by

Corollary 8 of Sherman (1994a). The remaining term to be studied is the degenerate second order U-process defined by the family of functions

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

By a Taylor expansion of $\mathbb{E}\left[g(Z_j,\theta_0)|X_j\right]$ 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^qM_{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 $\mathbb{E}[m_{113}\mid Z_j,Z_k]-\mathbb{E}[m_{113}\mid Z_k]$ and $\mathbb{E}[m_{113}\mid Z_i,Z_k]-\mathbb{E}[m_{113}\mid Z_k]$, and the empirical process based on $\mathbb{E}[m_{113}\mid Z_k]$. 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

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

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

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

Use Hoeffding's decomposition and the last statement of Lemma 5.1 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 5.1 to show that each of $\tilde{M}_{1jn}$, $j=1,2$, is a $O_p\left(\|\theta-\theta_0\|n^{-1/2-2\alpha/3}\right)$ for any $\alpha<1$. Use similar arguments to show that $\tilde{M}_{13n} = O_p\left(\|\theta-\theta_0\|^2 n^{-2\alpha/3}\right)$ for any $\alpha < 1$. Gathering these facts and using $n^{-1}\{\inf \mathcal{H}'_n\}^{-4q/\alpha} = o_p(1)$ for some $\alpha < 1$,

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

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. □

Proof of Corollary 2.6. 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

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

is uniformly Euclidean for a constant envelope by Nolan and Pollard (1987, Lemma 20). Reason similarly for the functions $W_n^{-1/2}(\cdot)\mathrm{H}_{\theta,\theta}\tau_n^{(k)}(\cdot,\theta_0)f(\cdot)$, $1\leq k\leq r$ and $W_n^{-1/2}(\cdot H(\cdot)f(\cdot)$. Use Lemmas 5.2 and 5.3 in Section 5.1 and Equation (2.7) to obtain an asymptotic representation similar to the one of Theorem 2.2. □

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, $\sqrt{n}\,(\tilde{\gamma}_n-\gamma_0) = -\left(V_n^R\right)^{-1}B_n + o_p(1)$, where $V_n^R = D'V_nD$ and $B_n = D'A_n$, and

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

uniformly in $h\in\mathcal{H}_n$ under H_0. Our conclusions follows form 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,

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

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

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

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

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

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

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 $\sqrt{n}\left(\tilde{\theta}_n^* - \tilde{\theta}_n\right)$ has asymptotically and conditionally upon the initial sample the same distribution as $\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right)$ uniformly in h.[4] Therefore, uniformly in $h \in \mathcal{H}_n$,

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

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

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 $M_n^*\left(R(\tilde{\gamma}_n^*)\right) - M_n^*\left(R(\tilde{\gamma}_n)\right) = o_p(1)$, so that $DM_n^* = o_p(n)$ uniformly in $h \in \mathcal{H}_n$. □

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

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.

Cragg, J. (1983). More efficient estimation in the presence of heteroscedasticity of unknown form. *Econometrica* **51(3)**, 751-764.

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.

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.

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.

Higham, N.J. (2008). *Functions of Matrices: Theory and Computation.* Society for Industrial and Applied Mathematics: Philadelphia.

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 perturing 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. (2006). Efficient conditional quantile estimation: the time-series case. UC San Diego.

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.

McShane E.J. (1934). Extension of range of functions. *Bull. Amer. Math. Soc.* **40**, 837-842.

Newey, 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.

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 qunatile 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.

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.

Schennach, S.M. (2007). Point estimation with exponentially tilted empirical likelihood. *Ann. Statist.* **35(2)**, 634-672.

Sherman, R.P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61(1)**, 123-137.

Sherman, R.P. (1994a). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist.* **22(1)**, 439-459.

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.

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.

Zhang, D. (2001). Bayesian bootstrap for U-processes, hypothesis tests and convergence of Dirichlet U-processes. *Statist. Sinica* **11(2)**, 463–478.

Zheng, J.X. (1998). A consistent nonparametric test of parametric regression models under conditional quantile restriction. *Econometric Theory* **14(1)**, 123-138.

Appendix A

We focus here on providing sets of sufficient conditions that guarantee Condition (E). We note that since $\int \phi_n(x-uh)K(u)du$ is the expectation of a kernel estimator, our subsequent results 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 $\phi_n : \mathbb{R}^q \to \mathbb{R}$, $n \geq 1$ have integrable envelope $\Phi(\cdot)$, they satisfy Condition (E) with kernel $K(\cdot)$ for an envelope $\Phi(\cdot) + C$, $C > 0$, whenever*

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

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 $|\mathcal{F}[K](t_1)-\mathcal{F}[K](t_2)|\le c\|t_1-t_2\|^a$,

$$\begin{aligned}\left|\int \phi_n(x-h_1u)K(u)du-\int \phi_n(x-h_2u)K(u)du\right| &\le \int |\mathcal{F}[\phi_n](t)|\,|\mathcal{F}[K](h_1t)-\mathcal{F}[K](h_2t)|\,dt \\ &\le c|h_1-h_2|^a\int \|t\|^a\,|\mathcal{F}[\phi_n](t)|\,dt\,.\end{aligned}$$

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

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 $m\ge 3$, see e.g. Malliavin (1995, Section III.3). Another possible space is the Sobolev space of functions H^s. Indeed,

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

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

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

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

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 functions $\phi_n(\cdot)$ 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 $\{(x,u) \mapsto subgraph(\phi_n(x - uh)) : h \in [0,1]\}$ 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. □

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

We provide here useful lemmas for proving that the primitive assumptions on the conditional variance of $g(Z,\theta_0)$ are sufficient for our results from Section 2.4 to hold.

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 $\|\cdot\|_2$,

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

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

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

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. □

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

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

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 $(x, u, v) \mapsto \widetilde{\omega}(x - uh - vb)\mathbb{I}(x - hu - vb \in D) = \omega(x - uh - vb)$, $x, u, v \in \mathbb{R}^q$, $h, b \in [0, h_0]$, is Euclidean for a constant envelope. The result follows since Euclideanity is preserved by integration. □

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 $\{(x,u) \mapsto \widetilde{\bar{\omega}}(x-uh;b) = \mathbb{I}(x-hu \in D^c)I + \bar{\omega}(x-hu;b) : h,b \in [0,h_0]\}$ is Euclidean for a constant envelope. By definition, the eigenvalues of $\widetilde{\bar{\omega}}(x-uh;b)$ stay away from zero and infinity and $\widetilde{\bar{\omega}}(x-uh;b) = \bar{\omega}(x-uh;b)$ whenever $x-uh \in D$.

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