Smooth Minimum Distance Estimation and Testing in Conditional Moment Restrictions Models: Uniform in Bandwidth Theory

Pascal Lavergne, Simon Fraser University and CREST-ENSAI
Valentin Patilea, INSA-IRMAR and CREST-ENSAI

This version: October 2008

Abstract

We propose a new estimation method for models defined by conditional moment restrictions, that minimizes a distance criterion based on kernel smoothing. Whether the bandwidth parameter is fixed or decreases to zero with the sample size, our approach defines a whole class of estimators. 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 inference procedures based on a distance metric statistic for testing restrictions on parameters and we propose a new bootstrap technique. Our new methods apply to non-smooth problems, are simple to implement, and perform well in small samples.

Keywords: Conditional Moments, Smoothing Methods.

JEL classification: Primary C31; Secondary C13, C14.

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, Dept. of Economics, Simon Fraser University, 8888 University Drive, Burnaby BC, V5A 1S6 CANADA. Emails: pascal_lavergne@sfu.ca, patilea@ensai.fr

1 Introduction

Many econometric models involve conditional moment restrictions (CMR), or equivalently an infinite number of unconditional ones. Generalized Method of Moments (GMM), as introduced by Hansen (1982), only exploits a finite number of unconditional moment restrictions. Subsequent research has focused on accounting for CMR to gain efficiency. Provided a preliminary consistent estimator, Robinson (1987) and Newey (1993) show how to estimate optimal instruments by nonparametric methods to obtain a two-step efficient estimator. However, as Dominguez and Lobato (2004) recently point out, in nonlinear models an arbitrary finite number of instruments, and even the optimal ones, may fail to globally identify the parameters of interest, see Dominguez and Lobato (2007) for further examples. The identification issue is crucial in practice: since classical GMM relies on a finite number of unconditional moments, we can never be sure that the chosen ones identify the parameters of interest and the estimator may be inconsistent.

Recent work has focused on accounting for CMR at the outset. Carrasco and Florens (2000) propose an estimator based on an infinite (countable or uncountable) number of moments. Antoine, Bonnal, and Renault (2007) develop a Euclidean Empirical Likelihood (EL) approach. Donald, Imbens, and Newey (2003), Kitamura, Tripathi, and Ahn (2004), and Smith (2007a,b) focus on smoothed generalized EL methods that provide *one-step* efficient estimators, thus avoiding the need for a preliminary consistent estimator. All these methods rely on a user-chosen parameter, whether it is a regularization parameter, as in Carrasco and Florens (2000), a bandwidth parameter, as in Antoine, Bonnal and Renault (2007), Kitamura, Tripathi and Ahn (2004), and Smith (2007a,b), or the number of series functions, as in Donald, Imbens and Newey (2003). Consistency and efficiency is shown when the user-chosen parameter, or its inverse in the latter case, converges to zero as the sample size increases. However, as pointed out by Kitamura, Tripathi, and Ahn, choosing the bandwidth is a vexing problem in applications. Moreover, one can never in practice set the smoothing parameter arbitrarily close to zero. Dominguez and Lobato (2004) propose the first consistent estimator that does not require a user-chosen parameter, but still exploits all CMR. Efficiency however is not reached using their criterion.

In this work, we propose a framework for estimation of parameters in models defined by CMR that bridges the gap between Dominguez and Lobato's approach and competing ones.

Our new estimator optimizes a new minimum distance criterion based on kernel averaging. Our *smooth minimum distance* (SMD) approach defines a whole class of consistent estimators: when the bandwidth parameter is fixed, our estimator is similar to but different than Dominguez and Lobato's estimator, and our simulations show that it is less variable; when the bandwidth decreases to zero, our estimator is close in spirit, but still different, to other proposals.

We develop a theory for SMD estimation and testing that focuses on accounting for the influence of the bandwidth. This feature is crucial for practical application since this parameter is usually selected depending on the sample size and the features of the data. Also it is key if one aims at deriving an optimal data-driven choice, as recently entertained by Carrasco (2007). Though we follow a different route, our work is similar in aim to recent work on heteroscedasticity-autocorrelation robust variance estimators where the focus is to account for the influence of the truncation parameter, see Kiefer and Volgelsang (2005), Sun, Phillips, and Jin (2008), and the references therein. It is also related to recent work on uniform in bandwidth consistency of kernel estimators, see Einmahl and Mason (2005) and the references therein. Specifically, we show uniform in bandwidth consistency and we provide a $\sqrt{n}$-asymptotic representation of the SMD estimator as a process indexed by the bandwidth. To the best of our knowledge, our uniform in bandwidth results are the first of their kind for estimation methods in models defined by CMR and are not available for smoothed EL estimators.

Our results extend to misspecified models. The behavior of GMM under misspecification has recently attracted some attention, see Hall and Inoue (2003), Aguirre-Torres and Dominguez-Toribio (2004), and Dridi, Guay and Renault (2007). Schennach (2007) recently shows that under misspecification the standard EL estimator cannot be $\sqrt{n}$-consistent for a pseudo-true value whenever the functions entering the moment restrictions are unbounded. Little is known on the behavior of estimators based on CMR, but one should fear that such a phenomenon also occurs for smoothed EL estimators. As our results show, the SMD estimator enjoys similar properties under misspecification than when the model is correct.

Our estimator can attain the semiparametric efficiency bound when the smoothing parameter decreases to zero. The efficient estimator requires neither estimation of conditional expectation of derivatives nor differentiability of the functions entering the moment restrictions, but only estimation of the density-weighted conditional variance, which can be done easily by kernel methods. In general, an efficient two-step estimator obtains based on a preliminary SMD estimator, which is consistent irrespective to the bandwidth's choice, and a kernel estimator of the density-weighted conditional variance. When the conditional variance is known, as in conditional quantile models, the efficient estimator is one-step, as the ones recently proposed by Otsu (2008) and Komunjer and Vuong (2006). We establish the uniform in bandwidth efficiency of our general estimator within a large range of bandwidths that tend to zero. This is reassuring from a theory viewpoint, though by nature such a result ignores the influence of the bandwidth. From a practical viewpoint, the efficient SMD is easy to implement and more straightforward to compute than the smoothed EL estimator of Kitamura, Tripathi and Ahn (2004). We also show through simulations that it behaves comparatively well in small samples.

Testing restrictions on parameter can be entertained from a distance metric approach based on our SMD criterion. Indeed, twice the difference between the constrained and unconstrained optimized criteria behaves like a likelihood-ratio statistic. When considered as a process, the statistic is a quadratic form in a tight asymptotically Gaussian process. If one neglects the influence of the bandwidth and assumes an efficient estimator, a classical chi-square distribution obtains. But basing the testing procedure on the general distribution should yield more reliable inference. We then propose a simple bootstrap method to approximate the distribution of our estimator and of our distance metric test statistic. To the best of our knowledge, this is the first general bootstrap method proposed for inference in nonlinear models defined by CMR. It is based on perturbing the objective function and thus does not require resampling observations. It is similar to a method recently proposed by Jin, Ying and Wei (2001). We show that the test and the bootstrap method are valid uniformly in the bandwidth.

We first focus in Section 2 on obtaining general consistency and asymptotic normality results uniformly over a large range of bandwidths including fixed ones. In Section 3, we investigate a distance-metric inference procedure for testing restrictions on parameters and our proposed bootstrap method for inference. Section 4 focus on deriving an efficient form of the SMD estimator, and shows that our testing results extend to the efficient estimator. Section 5 reports the results of a simulation study. Section 6 concludes. Proofs are gathered in Section 7. Two Appendices discuss in detail some of our technical conditions.

2 SMD Estimation

For a matrix A, $\|A\|$ is the usual extension of the Euclidean norm, $\lambda_{min}(A)$ and $\lambda_{max}(A)$ denote the smallest and the largest eigenvalue of A. For a function $l(\cdot)$, $\mathcal{F}[l](\cdot)$ is its Fourier transform. The operators ∇_θ and $\mathrm{H}_{\theta,\theta}$ respectively denote the vector of first partial derivatives and the matrix of second derivatives with respect to θ.

2.1 The Estimator and its Consistency

Let $g(Z,\theta) = (g^{(1)}(Z,\theta),...,g^{(r)}(Z,\theta))'$ be a $r-$vector valued function, $r \geq 1$, with $Z = (Y',X')' \in \mathbb{R}^{d+q}$, $d \geq 1$, $q \geq 1$, and $\theta \in \Theta \subset \mathbb{R}^p$, $p \geq 1$. With at hand independent copies $\{Z_1,\ldots Z_n\}$ from Z, we aim at estimating a parameter defined through the CMR

$$\mathbb{E}\left[g(Z,\theta_0)|X\right] = 0 \quad \text{a.s.} \tag{2.1}$$

We make the following identifiability assumption of θ_0.

Assumption 1. *(i) The parameter space Θ is compact. (ii) θ_0 is the unique value in Θ satisfying (2.1), that is $\mathbb{E}\left[g(Z,\theta)|X\right] = 0$ a.s. $\Rightarrow \theta = \theta_0$.*

The SMD criterion in its simplest form writes

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

with a multivariate kernel $K(\cdot)$ and $h = h_n$ a sequence of bandwidth parameters. This is the statistic studied by Delgado, Dominguez, and Lavergne (2005), a generalization of the one introduced by Zheng (1996) and Li and Wang (1996) for specification testing of regression models. When h tends to zero, the criterion has limit

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

where $f(\cdot)$ is the density of X. Hence, provided a consistent estimator for θ_0, the statistic can be used for testing (2.1). Here we use the statistic for estimation purposes and we thus do not assume the existence of a preliminary consistent estimator. Minimizing our criterion with respect to θ minimizes the (density-weighted) distance of $\mathbb{E}\left[g(Z,\theta)|X\right]$ to zero, provided h tends to zero. This provides a first intuition for our label *smooth minimum* *distance*. Moreover, as we may want to combine and weigh the different components of $g(\cdot,\theta)$, we introduce a sequence of non-random p.d. weighting matrices $W_n(\cdot)$ and the criterion

$$M_{n,h}(\theta) = \frac{1}{2n(n-1)} \sum_{1\leq i\neq j\leq n} g'(Z_i,\theta) W_n^{-1/2}(X_i) W_n^{-1/2}(X_j) g(Z_j,\theta) K_{ij}\,. \tag{2.2}$$

Discrete covariates U with finite support could be handled by multiplying each term by $\mathbb{I}(U_i = U_j)$ and our proofs would easily adapt. For the sake of simplicity, we do not formally consider this possibility in what follows.[1] Our estimator is

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

It belongs to the class of MINPIN estimators as studied by Andrews (1994). However, to study our estimator as a process indexed by the bandwidth, we cannot use his results. When h is fixed, our criterion resembles the one proposed by Dominguez and Lobato (2004), which for an univariate $g(\cdot,\cdot)$ writes

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

By contrast to our criterion, the weight in the above double sum depends on all observations X_i and may vary from 1 to $1/n$. Our criterion has then less variability, and this in turn can reduce variability in parameter estimation, as illustrated by our simulations results.

To understand why our estimator is consistent even when h does not tend to zero, let $W_n(X)$ be the identity matrix for simplicity. Then

$$\begin{aligned}
\mathbb{E}M_{n,h}(\theta) &= \frac{1}{2}\mathbb{E}\left[g'(Z_1,\theta)g(Z_2,\theta)h^{-q}K\left((X_1-X_2)/h\right)\right] \qquad (2.3)\\
&= \frac{1}{2}(2\pi)^{-q/2}\,\mathbb{E}\left[g'(Z_1,\theta)g(Z_2,\theta)\int_{\mathbb{R}^q}\exp\left(it'(X_1-X_2)\right)\mathcal{F}[K](ht)\,dt\right]\\
&= \frac{1}{2}(2\pi)^{q/2}\sum_{k=1}^{r}\left\{\int_{\mathbb{R}^q}\left|\mathcal{F}\left[\mathbb{E}[g^{(k)}(Z,\theta)|X=\cdot]f(\cdot)\right](t)\right|^2\mathcal{F}[K](ht)\,dt\right\}\,,
\end{aligned}$$

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 the label *smooth minimum distance* criterion.

[1] Russell Davidson suggested to include equal indices in the double sum. Our proofs would easily adapt with this modification, but we do not pursue further this suggestion because unreported simulation results do not indicate any general advantage in favor of this modification.

Since the expectation of the criterion accounts for the Fourier transform of $\mathbb{E}\left[g(Z,\theta)|X\right]$ at all frequencies, it yields a consistent estimator independently of h. Indeed, if $\mathcal{F}\left[K\right](\cdot)$ is strictly positive on $\mathbb{R}^q$, then using the unicity of the Fourier transform and Assumption 1,

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

A necessary condition for consistency is then the strict positivity of the Fourier transform of $K(\cdot)$. It is fulfilled for instance by products of the triangular, normal, logistic (see Johnson, Kotz, and Balakrishnan, 1995, Section 23.3), Student (see Hurst, 1995), or Laplace densities. It is then clear that $\tilde{\theta}_{n,h}$ is consistent for θ_0 provided the convergence $M_{n,h}(\theta)$ of the process to its limit uniformly in θ and h. Let us introduce our basic assumptions.

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

Symmetry of the kernel is not strictly speaking necessary here, but leads to simpler proofs later on. 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.[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). Nolan and Pollard (1987), among others, provide some sufficient conditions which are fulfilled by our above examples.

Assumption 3. *For all n, $W_n(\cdot)$ is a $r\times r$ symmetric positive definite 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 positive definite matrix function $W(\cdot)$ such that $W_n(u)-W(u)=o(1)$ for all $u\in\mathbb{R}^q$ with $0<\inf_u\lambda_{min}(W(u))\leq\sup_u\lambda_{max}(W(u))<\infty$.*

Assumption 4. *(i) The function $\sup_\theta\|\,\mathbb{E}[g(Z,\theta)\mid X=x]\,\|f(x)$ 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 $\mathbb{E}G^2<\infty$.*

[2] In recent statistical work, Euclidean families are also called VC classes.

Assumption 3 ensures that $W_n^{-1/2}(x)$ is well-defined and the spectral radius of $W_n^{-1/2}(\cdot)$ is uniformly bounded. Assumption 4 as a whole does not require the continuity of the functions $\theta \mapsto g(z,\theta)$. Assumption 4-(i) is used to show 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

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

is uniformly Euclidean for a squared integrable envelope, see Lemma 2.14-(ii) of Pakes and Pollard (1989). Here, *uniformly* means that the constants in the definition of the Euclidean family are independent of n. This property allows us to derive the desired result, where we abstract from measurability issues, that is we assume that conditions are met for the measurability of $\tilde{\theta}_{n,h}$ for any h.

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

$$\sup_{h_0 \geq h > 0,\ nh^{2q} \geq \ln(n+1)} \|\tilde{\theta}_{n,h} - \theta_0\| = o_p(1)\,.$$

A few remarks are useful. First, the result easily extends to any approximate estimator such that $M_{n,h}(\tilde{\theta}_{n,h}) \leq \min_\Theta M_{n,h}(\theta) + o_p(1)$ uniformly in h. Second, consistency is obtained under more general conditions that the ones in Kitamura, Tripathi, and Ahn (2004), who impose smoothness of the functions in $g(\cdot,\cdot)$ and more stringent conditions on the bandwidth. Third, the strict positivity of $\mathcal{F}[K](\cdot)$ can be weakened to positivity if X has a bounded support. In that case, Equation (2.3) 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 the $\theta = \theta_0$ using Theorem 1 of Bierens (1982). This allows in particular for the use of higher-order kernels taking negative values, as for instance the normalized sinc kernel whose Fourier transform is a uniform density. Fourth, one could allow $W_n(\cdot)$ to depend on another parameter b, as the smoothing parameter of a nonparametric estimator of an unknown matrix $W(\cdot)$. We consider this possibility later on, for now we note that our results would carry over assuming that Assumption 3 holds uniformly in b and that the class of matrix-valued functions $W_n(\cdot;b)$ indexed by b is Euclidean entrywise for a constant envelope.

2.2 Asymptotic Normality

Let us make the following supplementary assumptions.

Assumption 1. *(iii)* θ_0 *belongs to the interior of* Θ.

Assumption 4. *(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,* $\mathbb{E}\left\|g(Z,\theta) - g(Z,\theta_0)\right\|^2 \le c\|\theta - \theta_0\|$.

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

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

(iii) The components of $\nabla_\theta \tau(\cdot,\theta_0) f(\cdot)$ *are in* $L^1 \cap L^2$.

If the components $g^{(k)}(z,\theta)$ are twice differentiable with respect to θ, Assumptions 5(i)-(ii) are implied by the following condition.

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* $\mathbb{E}\tilde{H}^4(Z) < \infty$ *and* $a \in (0,1]$ *such that*

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

Under Condition 1, $\mathbb{E}\left\|g(Z,\theta) - g(Z,\theta_0)\right\|^2 = O(\|\theta - \theta_0\|^2)$, 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. Equation (A.11) in Zheng (1998). By Assumption 3, $g_n(Z,\theta) = W_n^{-1/2}(X)g(Z,\theta)$ also satisfies 4-(iv), and $\tau_n(X,\theta) = W_n^{-1/2}(X)\tau(X,\theta)$ inherits the smoothness properties of $\tau(X,\theta)$.

Let $\mathcal{F}_n = \{\phi_{n,h}(\cdot) : h \in [0,h_0]\}$ be the family of functions

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

and $\phi_{n,0}(z) = \nabla'_\theta \tau(x,\theta_0) W_n^{-1}(x) g\left(z,\theta_0\right) f(x)$. Define similarly $\phi_h(\cdot)$ for $h \in [0,h_0]$ with $W(\cdot)$ in place of $W_n(\cdot)$. We denote by $\{\mathbb{G}_n\phi_{n,h} : h \in [0,h_0]\}$ the sequence of centered empirical processes indexed by the families $\mathcal{F}_n$, that is $\mathbb{G}_n\phi_{n,h} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \phi_{n,h}(Z_i) - \mathbb{E}\phi_{n,h}(Z_i)$.

Under our following Assumption 6, the empirical process $\{\mathbb{G}_n \phi_{n,h} : h \in [0, h_0]\}$ converges in distribution to a tight zero-mean Gaussian process with covariance function $\Delta_{h_1,h_2} = \mathbb{E}[\phi_{h_1}(Z)\phi_{h_2}(Z)] - \mathbb{E}\phi_{h_1}(Z)\mathbb{E}\phi_{h_2}(Z)$, which is finite by Assumption 3 and 4-(i).

We now introduce a general condition that allows to analyze the above process. We say that a sequence of real-valued functions ψ_n satisfies Condition (E) with kernel $K(\cdot)$ for an envelope $\Psi(\cdot)$ (independent of n) if for each $n \geq 1$ the class of functions

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

is uniformly Euclidean for the envelope $\Psi(\cdot)$. Sufficient mild conditions on $\psi_n(\cdot)$ and $K(\cdot)$ that guarantee Condition (E) are provided in Appendix A. In particular, it is sufficient that the $\psi_n(\cdot)$ belong to some Sobolev space of functions, or are Hölder continuous on their support.[3]

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 $\mathbb{E}\Phi_1^a(X) < \infty$ 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 $\mathbb{E}\Phi_2^a(X) < \infty$ for some $a \geq 4/3$.*

Let us define the non-random matrices

$$V_{n,h} = \mathrm{H}_{\theta,\theta}\mathbb{E}M_{n,h}(\theta_0) = \mathbb{E}\left[\nabla'_\theta \tau_n(X_1, \theta_0)\nabla_\theta \tau_n(X_2, \theta_0)h^{-q}K\left((X_1 - X_2)/h\right)\right] \text{ for } h \in (0, h_0],$$

and $V_{n,0} = \lim_{h\downarrow 0} V_{n,h} = \mathbb{E}[\nabla'_\theta \tau_n(X, \theta_0)\nabla_\theta \tau_n(X, \theta_0)f(X)]$, which follows from Assumption 5-(iii) and arguments similar to those in Equation (2.3). The matrices V_h and V_0 are similarly obtained replacing $W_n(\cdot)$ by $W(\cdot)$. As is common, we first derive $\sqrt{n}$-consistency before establishing a $\sqrt{n}$-asymptotic representation for our estimator $\tilde{\theta}_{n,h}$ as a process.

Theorem 2.2. *For arbitrary constants, $h_0 > 0$, $C > 0$, and $0 < \alpha < 1$, let $\tilde{\mathcal{H}}_n = \left\{h_0 \geq h > 0 : nh^{4q/\alpha} \geq C\right\}$. For an i.i.d. sample and under Assumptions 1–6,*

i. $\sqrt{n}\left(\theta_{n,h} - \theta_0\right) = O_p(1)$ uniformly in $h \in \mathcal{H}_n$.

[3] 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). However, since we need to impose Euclidean conditions to investigate the rates of different degenerate U-processes, we use such conditions throughout.

ii. if $V_{n,h}$, $h \in [0, h_0]$, are 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 constants c_{min} and c_{max} independent of n and h,

$$\sqrt{n}\left(\tilde{\theta}_{n,h} - \theta_0\right) + V_{n,h}^{-1}\mathbb{G}_n\phi_{n,h} = o_p(1) \quad \textit{uniformly in } h \in \mathcal{H}_n \, ,$$

where $\{\mathbb{G}_n\phi_{n,h} : h \in [0, h_0]\}$ is a zero mean tight asymptotically Gaussian process.

Our theorem readily yields that $\sqrt{n}\left(\tilde{\theta}_{n,h} - \theta_0\right)$ is a tight asymptotically Gaussian process with asymptotic covariance $V_{h_1}^{-1}\Delta_{h_1,h_2}V_{h_2}^{-1}$, where

$$V_h = \mathbb{E}\left[\nabla_\theta \mathbb{E}\left[g(Z_1,\theta_0)|X_1\right] W^{-1/2}(X_1)W^{-1/2}(X_2)\nabla_\theta' \mathbb{E}\left[g(Z_2,\theta_0)|X_2\right] h^{-q} K\left((X_1 - X_2)/h\right)\right] .$$

For most purposes, our interest lies on its asymptotic variance, that is $V_h^{-1}\Delta_{h,h}V_h^{-1}$, where

$$\begin{aligned}\Delta_{h,h} \;=\; & \mathbb{E}\left[\nabla_\theta \mathbb{E}\left[g(Z_1,\theta_0)|X_1\right] W^{-1/2}(X_1)W^{-1/2}(X_2)\mathrm{Var}\left[g(Z_2,\theta_0)|X_2\right] W^{-1/2}(X_2)\right. \\ & \left. W^{-1/2}(X_3)\nabla_\theta' \mathbb{E}\left[g(Z_3,\theta_0)|X_3\right] h^{-2q} K\left((X_1 - X_2)/h\right) K\left((X_2 - X_3)/h\right)\right] .\end{aligned}$$

As also shown in our proofs section, a similar uniform-in-bandwidth result obtains for $\tilde{\mathcal{H}}_n = \left\{h_0 \geq h > 0 : nh^{2q/\alpha} \geq C\right\}$ if Condition 1 holds. This is almost as good as Andrews' general condition for MINPIN that the preliminary nonparametric estimator should converge at rate $n^{-1/4}$ at least.[4]

2.3 Study Under Misspecification

We now study our estimator under misspecification. As previously argued, this is useful at least as a "robustness" check. As we now show, the behavior of the SMD estimator is very similar whether misspecification exists or not, and specifically is always $\sqrt{n}$-consistent. While no formal result has established the properties under misspecification of alternative estimators methods referred to in the Introduction, Schennach (2007) shows that the excellent properties of EL estimator degrades enormously under the slightest misspecification, causing the loss of $\sqrt{n}$-consistency, and provides an in-depth discussion. In particular, she argues that under misspecification the implied EL probabilities place large weight on a few extreme observations to satisfy the moment restrictions. By contrast, SMD estimation does not impose the CMR, but aim at matching them at best.

[4] Indeed, that $\sqrt{nh^q}$ is at least of the same order than $n^{1/4}$ is equivalent to the requirement that nh^{2q} is bounded away from zero. Compare this to our condition that $nh^{2q/\alpha}$ is bounded away from zero for $\alpha < 1$.

Define the pseudo-true value $\bar{\theta}_{n,h}(W_n) = \bar{\theta}_{n,h} = \arg\min_\Theta \mathbb{E}M_{n,h}(\theta)$, which we assume to be uniquely defined and interior to the parameter space Θ for all h.[5] Note that for each n the criterion $\mathbb{E}M_{n,h}(\theta)$ is continuous as a function of θ and h so that $\bar{\theta}_{n,h}$ can be considered as continuous in h and we can extend its definition by continuity as

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

Let $\bar{\mathcal{F}}_n = \{\bar{\phi}_{n,h}(\cdot) : h \in [0,h_0]\}$, where

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

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

$$\begin{aligned}\bar{V}_{n,h} &= \mathrm{H}_{\theta,\theta}\mathbb{E}M_n(\bar{\theta}_{n,h}) = \mathbb{E}\left[\nabla'_\theta \tau_n(X_1,\bar{\theta}_{n,h})\nabla_\theta \tau_n(X_2,\bar{\theta}_{n,h}) h^{-q} K\left((X_1-X_2)/h\right)\right] \\ &\quad + \sum_{k=1}^{r} \mathbb{E}\left[\mathrm{H}_{\theta,\theta}\tau_n^{(k)}(X_1,\bar{\theta}_{n,h}) g_n(X_2,\bar{\theta}_{n,h}) h^{-q} K\left((X_1-X_2)/h\right)\right], \end{aligned} \tag{2.4}$$

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

Theorem 2.3. *Assume that $\nabla_\theta \tau_n(x,\theta) f(x)$ is continuous as a function of x and θ. For an i.i.d. sample, under Assumptions 1–6, where θ_0 is replaced by $\bar{\theta}_{n,h}$, the neighborhood of θ_0 is replaced by balls $B(\bar{\theta}_{n,h}, r)$ with r independent on h and n, and the constants and the uniform bounds involved in the assumptions are independent on h and n,*

i. $\sqrt{n}\left(\tilde{\theta}_{n,h} - \bar{\theta}_{n,h}\right) = O_p(1)$ *uniformly in $h \in \mathcal{H}_n$, where $\mathcal{H}_n$ is as in Theorem 2.2.*

ii. *if $V_{n,h}$, $h \in [0,h_0]$, are positive definite matrices such that $0 < c_{min} \le \lambda_{min}(V_{n,h}) \le \lambda_{max}(V_{n,h}) \le c_{max} < \infty$ for some constants c_{min} and c_{max} independent of n and h,*

$$\sqrt{n}\left(\tilde{\theta}_{n,h} - \bar{\theta}_{n,h}\right) + V_{n,h}^{-1}\mathbb{G}_n\bar{\phi}_{n,h} = o_p(1) \quad \textit{uniformly in } h \in \mathcal{H}_n ,$$

where $\left\{\mathbb{G}_n\bar{\phi}_{n,h} : h \in [0,h_0]\right\}$ is a zero mean tight asymptotically Gaussian process.

[5] When W_n does not depend on n, $\bar{\theta}_{n,h}$ depends only on h.

3 SMD-Based Testing for Parameter Restrictions

3.1 Asymptotics

Suppose we want to test the parametric restriction in explicit form

$$H_0 : \theta_0 = R(\gamma_0)\,, \tag{3.5}$$

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. Alternatively, one could look at a null hypothesis in implicit form, but the explicit formulation is as general.

Assumption 8. *(i) $R(\cdot)$ is twice continuously differentiable. (ii) Either $\nabla_\gamma R(\gamma_0)$ 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

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

This is analog to the test statistic used in a classical GMM context. For smoothed EL, a similar statistic is studied by Kitamura, Tripathi an Ahn (2004) in the differentiable case, and Otsu (2008) for conditional quantile models. One could alternatively consider tests of the Wald, Score or Lagrange Multiplier type. A theoretical advantage of the distance metric test is that it is automatically invariant to the formulation of the null hypothesis.

For $h \in [0, h_0]$, define

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

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 and Assumption 8*

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

The process $(\mathbb{G}_n\phi_{n,h})' \Lambda_{n.h} (\mathbb{G}_n\phi_{n,h})$ is tight and for each h it has an asymptotic weighted sum of chi-squares distribution $M_{p-s}(\cdot, \lambda)$, see e.g. Vuong (1989) for the definition of this distribution, where λ is the vector of positive eigenvalues of

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

when $\bar{r} = s$, and $\Lambda_h = V_h^{-1/2} \Delta_{h,h} V_h^{-1/2}$ when $\bar{r} = 0$. Hence we label this process a *tight weighted sum of chi-squares* process. The distribution of our distance-metric statistic is thus in general non-pivotal. Our result looks familiar: in a maximum-likelihood context, the likelihood-ratio test is asymptotically a weighted sum of chi-squares, as established by Vuong (1989); a similar result obtains in a GMM context, see Marcellino and Rossi (2008). The classical chi-square distribution reappears when the information matrix equality or its analog holds. As we show in Section 4, we recover a χ^2_{p-s} 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. We note that the previous theorem extends to misspecified models using our results and conditions in Section 2.2, though we do not formally consider such a generalization.

Determining critical values requires estimation of the eigenvalues λ and then of the matrix Λ_h. When Condition 1 holds, one can respectively estimate V_h and $\Delta_{h,h}$ by

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

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

where $\widehat{\text{Var}}\left[g(Z_k, \theta)|X_k\right]$ is a nonparametric consistent estimator of $\text{Var}\left[g(Z_k, \theta)|X_k\right]$, see for instance (4.7) below. If $g(\cdot, \cdot)$ is not differentiable, one can use numerical methods similar to the ones proposed by Pakes and Pollard (1989). In what follows, we shall propose another route.

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. In particular, wild bootstrap is widely used for hypothesis testing in regression models, see e.g. Mammen (1992) and the references therein. In CMR models, application of wild bootstrap requires generating resamples with the same values for the exogenous variables, but new observations for the endogenous variables. To yield valid critical values for hypothesis testing, the bootstrap samples should in addition mimic the behavior of the data under the null hypothesis. This can be done in simple cases, e.g. in regression models, and has been shown to give reliable approximations. In general however, generating bootstrap samples may be difficult or even infeasible: for instance if the model is nonlinear in the endogenous variables, a reduced form may not be available or unique.

We now propose a simple method that allows to circumvent these difficulties if they appear, that applies generally and is easy to implement. Instead of resampling observations, we perturb the objective function and recompute our test statistic using this perturbed objective function. A similar method has been proposed by Jin, Ying and Wei (2001); when the objective function is the sample mean of moments, Parzen and Lipsitz (2007) note that the latter is equivalent to the Bayesian bootstrap of Rubin (1981).[6] Jin, Ying and Wei (2001) show the validity of their method under conditions that do not apply in our context. More crucially, they do not investigate the use of their method for testing.

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 perturbed criterion as

$$M^*_{n,h}(\theta) = \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} w_i w_j g'(Z_i,\theta) W_n^{-1/2}(X_i) W_n^{-1/2}(X_j) g(Z_j,\theta) K_{ij}\,. \tag{3.6}$$

With this new criterion, we repeat the estimation process, that is we compute

$$\tilde{\theta}^*_{n,h} = \arg\min_{\theta} M^*_{n,h}(\theta)\,.$$

We first note that $\mathbb{E}\left(M^*_{n,h}(\theta)|Z_1,\ldots Z_n\right) = M_{n,h}(\theta)$ and since $\tilde{\theta}_{n,h}$ minimizes this quantity, $\tilde{\theta}^*_{n,h}$ is expected to tend to $\tilde{\theta}_{n,h}$ conditionally on the sample. Now, as we show in the proofs section, the perturbed and the original criterion have a similar "good" quadratic expansion. Therefore, the distribution of the distribution of $M^*_{n,h}(\tilde{\theta}^*_n) - M^*_{n,h}(\tilde{\theta}_{n,h})$ is close to the one of $M_{n,h}(\tilde{\theta}_{n,h}) - M_{n,h}(\theta_0)$, and similarly for the distributions of $\tilde{\theta}^*_{n,h} - \tilde{\theta}_{n,h}$ and $\tilde{\theta}_{n,h} - \theta_0$. Our method consists in generating a large number of sample draws from the same distribution w so as to determine precisely enough the distribution of the above statistics. In what follows, we show the uniform in bandwidth validity of this method.

[6]Our method is also related to the Bayesian bootstrap for models defined by *unconditional* moments restrictions as studied by Florens and Rolin (1994).

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

ii. $n\left(M^*_{n,h}(\tilde{\theta}^*_{n,h}) - M^*_{n,h}(\tilde{\theta}_{n,h})\right)$ *has asymptotically the same distribution as* $n\left(M_{n,h}(\tilde{\theta}_{n,h}) - M_{n,h}(\theta_0)\right)$.

The first result allows the use of our bootstrap method to approximate the distribution of $\tilde{\theta}_n$, and in particular can be used to determine confidence intervals for a single parameter. The second result will be the basis for critical value's determination in hypothesis testing. To understand how it can be done, 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. In that aim, we repeat the estimation process with and without the constraint (3.5), that is we compute

$$\tilde{\theta}^*_{n,h} = \arg\min_{\theta} M^*_{n,h}(\theta) \qquad \text{and} \qquad \tilde{\theta}^{R*}_{n,h} = \arg\min_{\theta,\theta=R(\gamma)} M^*_{n,h}(\theta) .$$

The bootstrap distance metric test statistic is then defined 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 can use Theorem 2.3 to show that DM^*_n is bounded in probability irrespective to whether H_0 holds.

4 Efficient SMD Estimation

We now turn to the possibility of rendering our estimator efficient: this is desirable from a theoretical viewpoint and suggests that the SMD estimator can compare well to competitors in practice. Our Theorem 2.2 readily gives the optimal weighting matrix $W(\cdot)$ that yields a semiparametric efficient estimator as characterized by Chamberlain (1987).

Corollary 4.1. *Under the Assumptions of Theorem 2.2, $\tilde{\theta}_{n,h}$ is semiparametrically efficient uniformly over $h \in \mathcal{H}'_n = \left\{1/\ln(n+1) \geq h > 0 : nh^{4q/\alpha} \geq C\right\}$ for arbitrary $C > 0$ and $0 < \alpha < 1$ if $W(X) = \operatorname{Var}\left[g(Z,\theta_0)|X\right] f(X)$.*

By contrast to GMM, the optimal weighting matrix does not involve $\nabla_\theta \mathbb{E}\left[g(Z,\theta_0)|X\right]$, and then makes the efficient SMD we shall propose easy to apply even if $g(\cdot,\cdot)$ is not differentiable. Let $\check{\theta}_n$ be a $\sqrt{n}$-consistent SMD estimate of θ_0, computed for instance by choosing $W_n(\cdot) = I$ and $h = 1$. Consider the nonparametric estimator of the optimal weight matrix-valued function $\operatorname{Var}[g(Z,\theta_0) \mid X = x]f(x)$ defined as

$$\widehat{W}_n(x,\theta) = \frac{1}{nb^q} \sum_{1\leq k\leq n} g(Z_k,\theta)g'(Z_k,\theta)L((x-X_k)/b) \tag{4.7}$$

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 (under mild conditions on $L(\cdot)$ the estimator is semi-definite positive). However, the probability of this event vanishes when n grows under our subsequent assumptions.

Our estimator is $\widehat{\theta}_{n,h,b} = \arg\min_\Theta \widehat{M}_{n,h,b}(\theta)$, where

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

It is thus in general a two-step estimator. But a preliminary estimator for θ_0 may not be necessary. Consider for instance the case of nonlinear quantile restrictions where $g(Z,\theta) = \mathbb{I}\left[Y - \mu(X,\theta) \leq 0\right] - \rho$ for known ρ, e.g. $\rho = 1/2$ for median restrictions. Then $W(x) = \rho(1-\rho)f(x)$, no preliminary estimator is needed, and a one-step efficient estimator obtains under our following assumptions, as to the ones recently proposed by Otsu (2008) and Komunjer and Vuong (2006).

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 $(x,\bar{x}) \mapsto L((x-\bar{x})/h)$, $x,\bar{x} \in \mathbb{R}^q$, with $h > 0$, 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* $0 < \inf_u \lambda_{\min}(W(u)) < \sup_u \lambda_{\max}(W(u)) < \infty$*. (iii)* $W(\cdot)$ *is Hölder continuous on* D*. (iv) Let* $W(\cdot, \theta) = \mathbb{E}[g(Z, \theta)g'(Z, \theta) \mid X = \cdot]f(\cdot)$*. For* θ *in a neighborhood of* θ_0*, some* $d > 2/3$*, and* $c > 0$*,* $\|W(x, \theta) - W(x)\| \leq c\|\theta - \theta_0\|^d$ *for all* x*.*

Assumption E4 is needed to apply a result from Einmahl and Mason (2005). Assumption E7 corresponds to supplementary restrictions with respect to the previous sections. Part (i) allows for a flexible form of the support of X. Allowing for a vanishing density would involve introducing some trimming into the objective function, as done by Kitamura, Tripathi and Ahn (2004), but this is outside the scope of this work. They also note that trimming does not affect their estimator in practice and in view of our simulations results we feel confident that the same applies for efficient SMD. 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 $\widehat{W}_n(\cdot, \check{\theta})$. Under our assumptions, it is easy to show that $\widehat{\theta}_{n,h,b}$ is consistent by adapting the proof of Theorem 2.1. Focusing on efficiency matters, we consider that h goes to zero and that the bandwidth b is in the same range than h. No relationship between the two bandwidths is required, though in practice they can be chosen related or even equal.

Theorem 4.2. *For an i.i.d. sample, under Assumptions 1, 2, E2, E4, 5, and E7, then*

$$\sup_{h,b\in\mathcal{H}'_n} \left|\widehat{M}_{n,h,b}(\theta) - M_{n,h,b}(\theta)\right| = o_p\left(n^{-1} + \|\theta - \theta_0\|/\sqrt{n} + \|\theta - \theta_0\|^2\right) \tag{4.8}$$

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

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 an equivalent of Assumption 6 holds that accounts for the dependence of the weighting matrix on b. We here provide some primitive conditions that together with Assumption E7 ensure such an assumption, though they are likely not the only or weakest possible.

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

Corollary 4.3. *Under the assumptions of Theorem 4.2 and E6, the results of Sections 2.1, 2.2, and 3 holds for* $\widehat{\theta}_{n,h,b}$ *uniformly in* $h, b \in \mathcal{H}'_n$*, and in particular*

$$\sqrt{n}\left(\widehat{\theta}_{n,h,b} - \theta_0\right) + \mathbb{E}_{n,h,b} = o_p(1) \quad \textit{uniformly in } h, b \in \mathcal{H}'{}_n\,,$$

where $\{\mathbb{E}_{n,h,b} : h, b \in [0, h_0]\}$ *is a zero mean tight asymptotically Gaussian process whose asymptotic covariance for* $h, b \in \mathcal{H}'_n$ *is identically equal to the efficiency bound.*

5 Small sample study

The first setup is the one considered by Dominguez and Lobato (2004), where

$$Y = \theta_0^2 X + \theta_0 X^2 + \varepsilon, \tag{5.9}$$

with $\theta_0 = 5/4$, $X \sim N(\mu, 1)$, and $\varepsilon \sim N(0, 1)$ independently of X. The unknown parameter is not globally identified whenever $\mu \neq 0$. Dominguez and Lobato (2004, hereafter DL) illustrate theoretically and through simulations the consequences of lack of global identification on nonlinear least-squares (NLS). Here we abstract from this issue and considered as our benchmark the NLS estimator optimized locally around the true value of the parameter, which is the infeasible efficient estimator based on the knowledge that the model is homoscedastic. We consider three versions of SMD: (i) $W = I$ and $h = 1$ (ii) $W = I$ and $h = 0.3$ (iii) the efficient version with $h = b = 0.3$. For implementation, we use a Gaussian kernel. We considered the cases $\mu = 0$ and $\mu = 1$ and two sample sizes, $n = 50$ and $n = 200$. All results are based on 5000 replications.

Figures 1 to 4 compare the densities of the different estimators centered and scaled by $\sqrt{n}$. Table 1 reports the ratios of root mean squared error (RMSE) and mean absolute deviation (MAE) of each estimator with respect to the one of the locally optimized NLS. DL's estimator is more variable than versions (i) and (ii) of SMD. Increasing the sample size does not significantly affect the performances of the latter with respect to NLS, and changing the bandwidth has little effect. The efficient version performs very well compared to NLS, and its accuracy improves when the sample size increases, even though the bandwidth does not adapt to the sample size.

The second setup is the one of Cragg (1983), Newey (1993), and Kitamura, Tripathi and Ahn (2004, herefater KTA), where

$$Y = \beta_1 + \beta_2 X + \varepsilon, \quad \mathbb{E}(\varepsilon|X) = 0, \quad \mathrm{Var}(\varepsilon|X) = .1 + .2X + .3X^2\,, \tag{5.10}$$

with $\beta_1 = \beta_2 = 1$, $\ln X \sim N(0,1)$, and ε is normally distributed. KTA (2004) concluded that in this setup the Smoothed Empirical Likelihood (SEL) works best among various estimators. As a benchmark, we considered the generalized least squares estimator based on the true variance function, and we also compute the feasible version based on the knowledge of the variance functional form. We consider only efficient SMD with varying bandwidth and a Gaussian kernel. Results for SMD are based on 5000 replications, while results for SEL are based on 500 replications as reported by KTA.

Table 2 reports the ratios of root mean squared error (RMSE) and mean absolute deviation (MAE) of each estimator with respect to the infeasible GLS. The efficient SMD performs well compared to the feasible GLS, though the latter relies on the parametric form of the variance. To gain further insight, Figure 5 reports the ratio of RMSE as a function of the bandwidth. The RMSE of the slope parameter varies much less than the one of the intercept. The relative performances of SEL and SMD vary depending on the bandwidth choice. However, the variation with respect to the bandwidth is strikingly different from what is reported for SEL by KTA: there is a clear minimum, especially for β_2, and the corresponding "optimal" bandwidth decreases with the sample size, while surprisingly it does not seem to for SEL (see Figure 1 in KTA).

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 run 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 because it has third central moment equal to one. We expect that this helps better mimicking the behavior of the statistic as is the case in simpler setups, see e.g. Mammen (1992).

Table 3 reports empirical levels of the test. In all cases, the level accuracy increases when the sample size increases. For Model (5.9), the empirical level accuracy is reasonable for $n =$

50, while somewhat away for $X \sim N(1,1)$, and very close to the nominal one for $n = 100$. By comparison, the Wald test based on the locally optimized NLS estimator over rejects. For Model (5.10), the range of bandwidths was selected in the grid $n^{-1/5} \times (2/3, 5/3, 8/3)$ according to KTA to investigate the effect of the bandwidth. This should not be taken as a recommendation: a bandwidth greater than one is quite large, as compared for instance with the interquartile range of X, which is 1.45. Our results follow a general pattern similar to the previous experiment: for relatively large bandwidths, the test over rejects, but this phenomenon fades out with increasing sample size. Tests based on FGLS (Wald and LR tests yield identical results) are severely oversized and are then not reliable. We also checked that using asymptotic critical values from the chi-square distribution with one degree of freedom for our test yields rejection percentages between 0 and 1.6% depending on the bandwidth and sample size, and thus does not constitute a credible alternative in small samples.

To sum up, our SMD estimator performs well in our simulation experiments and is competitive compared to the Smoothed Empirical Likelihood estimator. Our bootstrap technique yields reliable test level. Both are quite robust to the bandwidth choice in small samples, though too large a bandwidth yields less accurate estimates and critical values.

6 Conclusion

We have proposed a smooth minimum distance estimation method for finite-dimensional parameters in models defined by conditional moment restrictions. Our SMD estimator depends on a smoothing parameter but is $\sqrt{n}$-consistent independently of this parameter within a wide range allowing for a fixed one. In our theory, we consider this estimator as a process indexed by the bandwidth(s) and we establish a uniform in bandwidth asymptotic representation. Our results are derived under weaker smoothness conditions than the ones available for competing estimators, so that they readily apply to many models, as conditional quantile restrictions models. We have developed a testing procedure based on a distance-metric statistic. Since the smoothing parameter cannot in practice be chosen arbitrarily close to zero, and thus the behavior of our estimator and test can be badly approximated by asymptotics, we have proposed a new bootstrap method. We have also shown how to obtain an efficient version of the SMD estimator when the bandwidth converges to zero. In practice, both the estimator and the bootstrap method are simple to implement and are found to perform reasonably well in our simulations.

The higher-order properties of the estimator, the influence of the bandwidth and the optimal bandwidth choice should be investigated. An overidentification testing procedure based on our optimized criterion needs to be developed. Generalizations to situations where a functional nuisance parameter is present and to time-series contexts also require further study.

7 Proofs

7.1 Preliminary lemmas

In what follows we adopt the notations of Sherman (1993, 1994a) concerning U-statistics. Following his use, we say that $H_n(\theta) = o_p(1)$, respectively $O_p(1)$, uniformly over $o_p(1)$ neighborhoods of θ_0 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 $\sup_{h\in\mathcal{H}_n} \sup_{\|\theta-\theta_0\|\leq r_n} |H_n(\theta)| \leq b_n$. The following is a extension of Corollary 8 of Sherman (1994a).

Lemma 7.1. *Let* $\mathcal{F} = \{f(\cdot,\theta,h) : \theta \in \Theta, h > 0\}$ *be a class of degenerate functions on* $\mathcal{S}^k$, $k \geq 1$, *where* $f(\cdot,\theta_0,\cdot) \equiv 0$. *If*

i. $\mathcal{F}$ *is Euclidean for an envelope* F *satisfying* $\mathbb{E}F^4 < \infty$,

ii. There exists a neighborhood $\mathcal{N}$ *of* θ_0 *and two positive constants* a *and* c *such that* $\mathbb{E}f^2(\cdot,\theta,h) \leq c\|\theta-\theta_0\|^a$ *for all* θ *in that neighborhood and* $h > 0$,

then uniformly over $\mathcal{N}$ *and over* $h > 0$, *and for any* $0 < \alpha < 1$

$$n^{k/2} U_n^k f(\cdot,\theta,h) = \|\theta-\theta_0\|^{a\alpha/2} O_p(1) + O_p\left(n^{-\alpha/4}\right).$$

If we assume further that $f^2(\cdot,\theta,h) \leq \Phi(\cdot)\|\theta-\theta_0\|^a$ *with* $\mathbb{E}\Phi < \infty$, *then* $n^{k/2}U_n^k f(\cdot,\theta,h) = \|\theta-\theta_0\|^{a\alpha/2} O_p(1)$.

Proof. Following the proof of Sherman (1994a, Corollary 8),

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

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

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

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

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 the proof in Sherman (1993).

Lemma 7.2. *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. Assume that*

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 $h \in \mathcal{H}_n$

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

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 7.3. *Let $\theta_{n,h}$ be as in Lemma 7.2. 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 $\bar\theta_{n,h}$,*

$$M_{n,h}(\theta) = M_{n,h}(\bar\theta_{n,h}) + \frac{1}{2}\left(\theta - \bar\theta_{n,h}\right)' V_{n,h}\left(\theta - \bar\theta_{n,h}\right) + \frac{1}{\sqrt{n}} A'_{n,h}\left(\theta - \bar\theta_{n,h}\right) + o_p(1/n) \qquad (7.11)$$

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

7.2 Main proofs

In the main proofs, we use for simplicity a single index n in place of the double indices n and h, so that for instance 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 (2.3) yields

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

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 (2.3), 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,

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

so that $\inf_{\|\theta-\theta_0\|\geq\varepsilon}\sup_{h\in\bar{\mathcal{H}}_n}\left[M_n(\theta) - M_n(\theta_0)\right] \geq \mu/2$ 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) and (iii) of Theorem 2.3's proof, setting $\bar{\theta}_n = \theta_0$ and accounting for (2.1). □

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: We have

$$\begin{aligned}&\mathbb{E}M_n(\theta) - \mathbb{E}M_n(\bar{\theta}_n) \\ &= (\theta-\bar{\theta}_n)'\nabla_\theta\mathbb{E}M_n(\bar{\theta}_n) + \frac{1}{2}(\theta-\bar{\theta}_n)'\mathrm{H}_{\theta,\theta}\mathbb{E}M_n(\bar{\theta}_n)(\theta-\bar{\theta}_n) + o(\|\theta-\bar{\theta}_n\|^2) \\ &= \frac{1}{2}(\theta-\bar{\theta}_n)'\bar{V}_n(\theta-\bar{\theta}_n) + o(\|\theta-\bar{\theta}_n\|^2) \geq (\kappa + o_p(1))\|\theta-\bar{\theta}_n\|^2\ ,\end{aligned}$$

uniformly in $h \in \mathcal{H}_n$, since $\nabla_\theta\mathbb{E}M_n(\bar{\theta}_n) = 0$.
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 $\tilde{l}_\theta(Z_i) = \mathbb{E}[l_\theta(Z_i,Z_j)\mid Z_i] + \mathbb{E}[l_\theta(Z_i,Z_j)\mid Z_j] - 2\mathbb{E}[l_\theta(Z_i,Z_j)]$,

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

By Assumption 5,

$$
\begin{aligned}
2\mathbb{E}[l_{1\theta}(Z_i,Z_j)\mid Z_i] &= g_n'(Z_i,\bar\theta_n)\mathbb{E}\left[\left(g_n(Z,\theta)-g_n(Z,\bar\theta_n)\right)h^{-q}K\left(\left(X_i-X\right)/h\right)|Z_i\right] \\
&= g_n'(Z_i,\bar\theta_n)\left[\int_{\mathbb{R}^q}\nabla_\theta'\tau_n(x,\bar\theta_n)f(x)h^{-q}K\left(\left(X_i-x\right)/h\right)\,dx\right]\left(\theta-\bar\theta_n\right) \quad (7.12)\\
&\quad+\frac{1}{2}g_n'(Z_i,\bar\theta_n)\sum_{k,l=1}^{p}\left(\theta^{(k)}-\bar\theta_n^{(k)}\right)\left(\theta^{(l)}-\bar\theta_n^{(l)}\right)\\
&\quad\left[\int_{\mathbb{R}^q}\mathrm{H}_{\theta^{(k)}\theta^{(l)}}\tau_n(x,\bar\theta_n)f(x)h^{-q}K\left(\left(X_i-x\right)/h\right)\,dx\right]+R_{1n}(Z_i,\theta) \quad (7.13)
\end{aligned}
$$

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

and $H_n(\cdot)=W_n^{-1/2}(\cdot)H(\cdot)$. By Assumption 6-(i), the functions $\nabla_\theta\tau_n^{(k)}(\cdot,\bar\theta_n)f(\cdot)$, $n\ge1$ satisfy Condition (E) for an envelope Φ with $\mathbb{E}\Phi^a(X)<\infty$ for some $a\ge4$. Use Assumption 4 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 (7.12) is uniformly Euclidean for a squared-integrable envelope. Let $A_n=\bar{\mathbb{G}}_n\widetilde{\phi}_{n,h}$, then $A_n=O_p(1)$ uniformly in θ and $h\in[0,h_0]$ by Corollary 4 of Sherman (1994a). Similarly, the family of functions in (7.13) 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 an $O_p(1)$ uniformly in $h\in[0,h_0]$. By symmetry, the same reasoning applies for $l_{2\theta}$. Finally a similar expansion for $l_{3\theta}$ yields

$$
\begin{aligned}
&2\mathbb{E}[l_{3\theta}(Z_i,Z_j)\mid Z_i]\\
&= \left(g_n(Z_i,\theta)-g_n(Z_i,\bar\theta_n)\right)'\mathbb{E}\left[\left(g_n(Z,\theta)-g_n(Z,\bar\theta_n)\right)h^{-q}K\left(\left(X_i-X\right)/h\right)|Z_i\right]\\
&= \left(g_n(Z_i,\theta)-g_n(Z_i,\bar\theta_n)\right)'\\
&\quad\left[\int_{\mathbb{R}^q}\nabla_\theta'\tau_n(x,\bar\theta_n)f(x)h^{-q}K\left(\left(X_i-x\right)/h\right)\,dx\right]\left(\theta-\bar\theta_n\right) \quad (7.14)\\
&\quad+\frac{1}{2}\left(g_n(Z_i,\theta)-g_n(Z_i,\bar\theta_n)\right)'\sum_{k,l=1}^{p}\left(\theta^{(k)}-\bar\theta_n^{(k)}\right)\left(\theta^{(l)}-\bar\theta_n^{(l)}\right)\\
&\quad\left[\int_{\mathbb{R}^q}\mathrm{H}_{\theta^{(k)}\theta^{(l)}}\tau_n(x,\bar\theta_n)f(x)h^{-q}K\left(\left(X_i-x\right)/h\right)\,dx\right]+R_{3n}(Z_i,\theta)\,.
\end{aligned}
$$

Since the function in (7.14) is such that

$$
\mathbb{E}\left|\left(g_n(Z_i,\theta)-g_n(Z_i,\bar\theta_n)\right)'\left[\int_{\mathbb{R}^q}\nabla_\theta'\tau_n(x,\bar\theta_n)f(x)h^{-q}K\left(\left(X_i-x\right)/h\right)\,dx\right]\right|\to0
$$

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

$$\mathbb{P}_n \tilde{l}_\theta = \frac{1}{\sqrt{n}} A_n' \left(\theta - \bar\theta_n\right) + \|\theta-\bar\theta_n\|^2 o_p(1)\,, \tag{7.15}$$

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

Consider now the second order degenerate 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 $a_1, ..., a_r \in \mathbb{R}$, $(a_1 + ... + a_r)^2 \leq r(a_1^2 + ... + a_r^2)$, deduce that $\mathbb{E} h^{2q} l_\theta^2(Z_i, Z_j) = O(\|\theta - \bar\theta_n\|)$. From Assumption 4–(iii), $h^q l_\theta(Z_i, Z_j)$ is Euclidean for an integrable envelope with fourth moment. Use Lemma 7.1 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 θ_0, which yields

$$\sup_{h \in \mathcal{H}_n} |U_n \bar{l}_\theta| = \|\theta-\bar\theta_n\|^{\alpha/2} O_p(\sup_{h\in\mathcal{H}_n} n^{-1} h^{-q}) + O_p(\sup_{h\in\mathcal{H}_n} n^{-1-\alpha/4} h^{-q})\,. \tag{7.16}$$

Choose α such that $nh^{4q/\alpha} \geq C$ for all $h \in \mathcal{H}_n$ 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 $O_p(\varepsilon_{0,n})$ where $\varepsilon_{0,n} = o(\sup_{h\in\mathcal{H}_n} n^{-1}h^{-q})$. Use Equations (7.15) and (7.16) in conjunction with Lemma 7.2 to obtain $\|\tilde\theta_n - \bar\theta_n\| = O_p(\varepsilon_{0n}^{1/2})$. Plug in this result in (7.16), 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 $4/(4-\alpha)$. Hence, after m iterations with m finite large enough, the first term in Equation (7.16) is a $O_p(n^{-1})$ if one can choose α such that $\varepsilon_{0,n}^{4/(4-\alpha)} = o(n^{-1})$. This is possible since there is some $\alpha < 1$ such that $nh^{4q/\alpha} \geq C$ for all $h \in \mathcal{H}_n$. Apply then again Lemma 7.2 to conclude that $\|\tilde\theta_n - \bar\theta_n\| = O_p(n^{-1/2})$.

Remark that under Condition 1, Equation 7.16 becomes

$$\sup_h |U_n \bar{l}_\theta| = \|\theta - \bar\theta_n\|^{\alpha} O_p(\sup_h n^{-1} h^{-q})\,.$$

Reason as above to obtain that $\sup_h |U_n \bar{l}_\theta| = O_p(n^{-1})$ as there is some $\alpha < 1$ such that $nh^{\frac{2q}{\alpha}} \geq C$ for all h.

(iii) Asymptotic representation: From the proof of Part ii and (7.15),

$$M_n(\theta) = M_n(\bar\theta_n) + \frac{1}{2}\left(\theta-\bar\theta_n\right)' \bar{V}_n \left(\theta - \bar\theta_n\right) + \frac{1}{\sqrt{n}} A_n' \left(\theta - \bar\theta_n\right) + o_p(1/n)\,, \tag{7.17}$$

uniformly over $O_p(1/\sqrt{n})$ neighborhoods of $\bar\theta_n$ and in $h \in \mathcal{H}_n$. Conclude from Lemma 7.3 that $\sqrt{n}\left(\tilde\theta_n - \bar\theta_n\right) + \bar V_n^{-1} A_n = o_p(1)$.

(iv) Behavior of $\mathbb{G}_n \phi_{n,h}$: We consider the case $r = 1$, the multivariate case follows similarly at the cost of more cumbersome algebra. The result follows from Theorem 19.28 of van der Vaart (1998) if we 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 $\delta \to 0$. Note that the Lindeberg condition follows from our Assumption 4 and 6. Let $\omega_n^2(X) = \text{Var}\left[g_n(Z,\bar\theta_n)|X\right]$. Proceed as in the consistency proof to show that

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

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

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

Use the uniform continuity of $\mathcal{F}[K](\cdot)$, the properties of the Fourier transform, and the Lebesgue dominated convergence theorem to conclude. □

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

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

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 to a positive constant, so that DM_n diverges. □

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 $h \in \mathcal{H}_n$ (i) $\tilde\theta_n^* - \theta_0 = o_p(1)$, (ii)

$\tilde{\theta}_n^* - \theta_0 = O_p(1/\sqrt{n})$, (iii) uniformly over $O_p(1/\sqrt{n})$ neighborhoods of θ_0,

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

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

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

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

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

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

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.[7] On the other hand,

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

As before, this expansion also holds conditionally. Hence, 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.

When H_0 does not hold, it follows from the arguments of 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 $DM_n^* = o_p(n)$. □

[7] 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.

Proof of Corollary 4.1. Under the conditions of the corollary $\tilde{\theta}_{n,h}$ is asymptotically a tight Gaussian process with covariance identically equal to $V_0^{-1}\Delta_{0,0}V_0^{-1}$, with

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

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

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

Proof of Theorem 4.2. 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 7.4. *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 $\eta_{\varphi,n,b}(x) = (nb^q)^{-1}\sum_{1\leq i\leq n} \varphi(Z_i)L((x-X_i)/b)$, $\varphi \in \Phi$ and $\|\cdot\|_\infty$ be the supremum norm. There exists $c > 0$ such that, with probability 1

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

Now notice that from our Assumption E7,

$$\|\mathbb{E}\left[(W(X,\theta) - W(X,\theta_0))\, b^{-q}L((X-x)/b)\right]\| \leq c\|\theta-\theta_0\|^d \|\mathbb{E}\left[b^{-q}L((X-x)/b)\right]\| \leq C\|\theta-\theta_0\|^d$$

for some finite constant $C > 0$ and $d > 2/3$. Hence uniformly for θ in a $O(n^{-1/2})$ neighborhood of θ_0 and in b, the above difference is a $o(n^{-1/3})$.

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 (2007), for any positive definite matrix A

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

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

$$\begin{aligned} (A+B)^{-1} &= A^{-1} - A^{-1}\left(I + BA^{-1}\right)^{-1} BA^{-1}\,, \qquad (7.18)\\ \text{and} \quad \left[I + t^2 D\left(t^2 A + I\right)^{-1}\right]^{-1} &= I - t^2 D\left(t^2 A + I\right)^{-1} + R\,,\\ \text{with} \qquad \|R\| \le \sqrt{q}\,\|R\|_2 &\le \frac{\sqrt{q}\left\|t^2 D\left(t^2 A + I\right)^{-1}\right\|_2^2}{1 - \left\|t^2 D\left(t^2 A + I\right)^{-1}\right\|_2}\\ &\le \sqrt{q}\,\|D\|_2^2\left[\frac{t^2}{1 + t^2\lambda_{min}(A)}\right]^2\left[1 - \frac{t^2\,\|D\|_2}{1 + t^2\lambda_{min}(A)}\right]^{-1}\\ &\le k(c)\,\|D\|_2^2 \le k(c)\,\|D\|^2\,. \end{aligned}$$

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

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

Use the integral representation for $(A+D)^{-1/2}$ and $A^{-1/2}$ and apply (7.18) with A replaced by $t^2 A + I$ and $B = t^2 D$ to deduce that

$$\begin{aligned} (A+D)^{-1/2} - A^{-1/2} = &-\frac{2}{\pi}\int_0^\infty t^2\left(t^2 A + I\right)^{-1} D\left(t^2 A + I\right)^{-1} dt\\ &+\frac{2}{\pi}\int_0^\infty t^4\left(t^2 A + I\right)^{-1} D\left(t^2 A + I\right)^{-1} D\left(t^2 A + I\right)^{-1} dt\\ &-\frac{2}{\pi}\int_0^\infty t^2\left(t^2 A + I\right)^{-1} RD\left(t^2 A + I\right)^{-1} dt\,, \qquad (7.19) \end{aligned}$$

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

This implies that for some constant C the last integral in (7.19) is bounded by

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

Step 3 consists in applying Identity (7.19) 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 $0 < \lambda \le \inf_x \lambda_{min}(W_n(x))$ for some fixed $\rho > 0$, which exists by our Assumption E7. For any $\theta \in \Theta$ and θ_2 in a $O(n^{-1/2})$

neighborhood of θ_0,

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

where the last term is an $o_p(n^{-1})$ by Step 1 uniformly in $b\in\mathcal{H}_n'$. In the above equation,

$$\begin{aligned}M_{1n}(t) &= M_{1n}(t,\theta,\theta_2,h,b) \\ &= \frac{t^{-4}\left(1+t^2\lambda\right)^2}{2n(n-1)}\sum_{i\neq j}g(Z_i,\theta)[W_n(X_i)+t^{-2}I]^{-1}D_{n,i}(\theta_2) \\ &\qquad\times[W_n(X_i)+t^{-2}I]^{-1}W_n^{-1/2}(X_j)g(Z_j,\theta)K_{ij}\,, \\ M_{2n}(t) &= M_{2n}(t,\theta,\theta_2,h,b) \\ &= \frac{t^{-6}\left(1+t^2\lambda\right)^3}{2n(n-1)}\sum_{i\neq j}g(Z_i,\theta)[W_n(X_i)+t^{-2}I]^{-1}D_{n,i}(\theta_2)[W_n(X_i)+t^{-2}I]^{-1} \\ &\qquad\times D_{n,i}(\theta_2)[W_n(X_i)+t^{-2}I]^{-1}W_n^{-1/2}(X_j)g(Z_j,\theta)K_{ij}\,,\end{aligned}$$

$$\begin{aligned}&M_{3n}(t,s) = M_{3n}(t,s,\theta,\theta_2,h,b) \\ &\quad= \frac{(1+t^2\lambda)^2(1+s^2\lambda)^2}{t^4s^42n(n-1)}\sum_{i\neq j}g(Z_i,\theta)[W_n(X_i)+t^{-2}I]^{-1}D_{n,i}(\theta_2)[W_n(X_i)+t^{-2}I]^{-1} \\ &\qquad\qquad\times[W_n(X_j)+s^{-2}I]^{-1}D_{n,j}(\theta_2)[W_n(X_j)+s^{-2}I]^{-1}g(Z_j,\theta)K_{ij}\,.\end{aligned}$$

Strictly speaking, we should separate the integrals on $[0,1)$ and $[1,\infty)$. 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

$$\sup_{t,s\ge1}\sup_{b,h\in\mathcal{H}_n}\left\{\|M_{1n}\|+\|M_{2n}\|+\|M_{3n}\|\right\} = o_p\left(n^{-1}+\|\theta-\theta_0\|/\sqrt{n}+\|\theta-\theta_0\|^2\right), \qquad (7.20)$$

which implies (4.8). The terms M_{1n}, M_{2n} and M_{3n} involve the family of matrix-valued functions

$$\left\{\left[W_n(\cdot)+t^{-2}I\right]^{-1} : b\in\mathcal{H}_n,\ t\ge1\right\}\quad\text{and}\quad\left\{W_n^{-1/2}(\cdot):b\in\mathcal{H}_n\right\}.$$

For $t \in [0,1)$, the first family has to be replaced by $\left\{\left[t^2 W_n(\cdot)+I\right]^{-1} : b \in \mathcal{H}_n,\, t \in [0,1)\right\}$. We here focus on the case $t \geq 1$, the arguments for the other case being similar. Lemma 7.8 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 (7.20) 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

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

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

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

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

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

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

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

By convention, for $j = 0, \dots 3$, we denote by M_{11j} the U-process based on m_{11j}. The term $h^q M_{110}$ 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 $h^q M_{111}$ 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

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

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 $h^q M_{112}$. For $h^q M_{113}$, the different terms in Hoeffding's decomposition are the third order degenerate U-process, the two degenerate second order U-processes based on $\mathbb{E}[h^q m_{113} \mid Z_j, Z_k] - \mathbb{E}[h^q m_{113} \mid Z_k]$ and $\mathbb{E}[h^q m_{113} \mid Z_i, Z_k] - \mathbb{E}[h^q m_{113} \mid Z_k]$, and the empirical process based on $\mathbb{E}[h^q m_{113} \mid Z_k]$. 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

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

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

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

Use Hoeffding's decomposition and the last statement of Lemma 7.1 to deduce that $h^q M_{10}$ 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 7.1 to show that each of M_{1j}, $j = 1, 2$, 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 $M_{13} = O_p\left(\|\theta-\theta_0\|^2 n^{-2\alpha/3}\right)$ for any $\alpha < 1$. Gathering these facts and using $n^{-1}\{\inf \mathcal{H}_n\}^{-4q/\alpha} = o_p(1)$ for some $\alpha < 1$,

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

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 4.3. Assumption E6, Lemma 7.6 of Appendix A, and Lemma 7.8 of Appendix B ensure that the class of functions $(x,u)\mapsto W_n^{-1/2}(x-hu;b)\nabla_\theta\tau(\cdot,\theta_0)f(\cdot)$ is Euclidean entrywise for a constant envelope, so that

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

is uniformly Euclidean for a constant envelope by Nolan and Pollard (1987, Lemma 20). Reason similarly for the functions $W_n^{-1/2}(\cdot;b)\mathrm{H}_{\theta,\theta}\tau_n^{(k)}(\cdot,\theta_0)f(\cdot)$, $1\le k\le r$ and $W_n^{-1/2}(\cdot;b)H(\cdot)f(\cdot)$. Use Lemmas 7.2 and 7.3 in Section 7.1 and Equation (4.8) to obtain an asymptotic representation similar to the one of Theorem 2.2. □

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 7.5. *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*

$$\sup_n \int \|t\|^a \left|\mathcal{F}[\phi_n](t)\right| dt < \infty. \tag{7.21}$$

Proof. For any ϕ_n, write

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

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

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

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 (7.21) is fulfilled when $\phi_n(\cdot)$ belongs to $W^{m,1}$, the subspace of functions of L^1 such that their weak partial derivatives belongs to L^1 up to integer order $m \geq 3$, see e.g. Malliavin (1995, Section III.3). Another possible space is the Sobolev space of functions H^s. Indeed,

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

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

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

Condition (7.21) then holds for a sequence $\phi_n(\cdot)$ from the Sobolev space of functions H^s with $s > 3/2$ endowed with the norm

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

For any integer $s \geq 1$, H^s is isomorph to $W^{s,2}$ endowed with the norm $\|\phi\|^2_{W^{s,2}} = \sum_{0\leq|\alpha|\leq s} \|D^\alpha \phi\|^2_{L^2}$, 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 $m \geq 3$, their product belongs to $W^{m,1}$ and thus also fulfills Condition (E).
Different sufficient conditions are provided in the next lemma.

Lemma 7.6. *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 7.6-(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 7.6-(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 7.6-(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 4 to hold.

Lemma 7.7. *Let $\omega(x; b)$, $b \in [0, h_0]$, be positive definite matrix-valued functions defined 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 $\{(x, u) \mapsto \omega^{-s}(x - uh; b) : h, b \in [0, h_0]\}$, $s = 1/2$ or 1, is Euclidean for a constant envelope.*

Proof. We treat the case $s = 1/2$, the other case similarly follows. For any positive definite 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 $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$,

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

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 7.8. *If $\omega(x)$ defined on D has eigenvalues uniformly bounded away from zero and infinity on D and is Hölder continuous with exponent a 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 $U(x)$ with positive diagonal entries such that $\omega(x) = U(x)U'(x)$. Since the eigenvalues of $\omega(\cdot)$ are uniformly bounded away from zero and infinity, the same holds for their square roots, that is the diagonal entries of $U(\cdot)$. Moreover, the entries of $U(\cdot)$ are Hölder continuous functions with exponent a since they obtain recursively from the entries of $\omega(\cdot)$ through the equations

$$U_{i,i}^2(x) = \omega_{i,i}(x) - \sum_{k=1}^{i-1} U_{i,k}^2(x),\ U_{i,j}(x) = U_{i,i}^{-1}(x)\left(\omega_{i,j}(x) - \sum_{k=1}^{i-1} U_{i,k}(x)U_{j,k}(x)\right),\ 1 \le i \le n,\ i > j\,.$$

By Theorem 3.3 and Remark 3.4 of Le Gruyer and Archer (1998), each entry $U_{i,j}(x)$ can be extended to $\mathbb{R}^q$ such that its extension is Hölder continuous with the same exponent and remains between $\inf_{x\in D} U_{i,j}(x)$ and $\sup_{x\in D} U_{i,j}(x)$. The corresponding matrix extension $\tilde{U}(\cdot)$ yield an extension $\tilde{\omega}(\cdot) = \tilde{U}(\cdot)\tilde{U}'(\cdot)$ of $\omega(\cdot)$ on $\mathbb{R}^q$ which is positive definite with eigenvalues uniformly bounded away from zero and infinity and Hölder continuous. By Lemma 2.13 of Pakes and Pollard (1989) and the fact that multiplication preserves Euclideanity, deduce that the class of functions $(x,u,v) \mapsto \tilde{\omega}(x-uh-vb)\mathbb{I}(x-hu-vb \in D) = \omega(x-uh-vb)$, $x,u,v \in \mathbb{R}^q$, $h,b \in [0,h_0]$, is Euclidean for a constant envelope. The result follows since Euclideanity is preserved by integration. $\square$

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

Lemma 7.9. *$\{(x,u) \mapsto \bar{\omega}^{-s}(x-hu;b)\mathbb{I}(x-hu \in D) : h,b \in [0,h_0]\}$, $s = 1/2$ or 1, is Euclidean for a constant envelope under the assumptions of Lemma 7.8.*

Proof. Lemma 7.6 and the fact that Euclideanity is preserved by addition yield that the class of functions $\{(x,u) \mapsto \tilde{\bar{\omega}}(x-uh;b) = \mathbb{I}(x-hu \in D^c)\mathrm{I}_q + \bar{\omega}(x-hu;b) : h,b \in [0,h_0]\}$ is Euclidean for a constant envelope. By definition, the eigenvalues of $\tilde{\bar{\omega}}(x-uh;b)$ stay away from zero and infinity and $\tilde{\bar{\omega}}(x-uh;b) = \bar{\omega}(x-uh;b)$ whenever $x-uh \in D$. By Lemma 7.7, the class $\left\{(x,u) \mapsto \tilde{\bar{\omega}}^{-1/2}(x-uh;b) : h,b \in [0,h_0]\right\}$ is then Euclidean for a constant envelope, and by Lemma 7.6-(i), so is the class $\left\{(x,u) \mapsto \tilde{\bar{\omega}}^{-1/2}(x-uh;b)\mathbb{I}(x-hu \in D) : h,b \in [0,h_0]\right\}$. A similar reasonning applies when $s=1$. $\square$

REFERENCES

Adams, R.A., and Fournier, J.J.F (2003). *Sobolev Spaces (2end edition).* Academic Press, Elsevier Science: Oxford.

Aguirre-Torres, V., and Dominguez Toribio, M. (2004). Efficient method of moments in misspecified i.i.d. models. *Econometric Theory* **20(3)**, 513-534.

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.

BROWN, D., AND WEGKAMP, M. (2002). Weighted minimum mean-square distance from independence estimation. *Econometrica* **70(5)**, 2035-2051.

CARRASCO, M. (2007). A regularization approach to the many instruments problem. Université de Montréal.

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.

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

DELGADO, M.A., DOMINGUEZ, M.A., AND LAVERGNE, P. (2006). Consistent tests of conditional moment restrictions. *Ann. Econom. Statist.* **81(1)**, 33–67.

DOMINGUEZ, M.A., AND LOBATO, I.N. (2004). Consistent estimation of models defined by conditional moment restrictions. *Econometrica* **72(5)**, 1601-1615.

DOMINGUEZ, M.A., AND LOBATO, I.N. (2007). A consistent specification test for models defined by conditional moment restrictions. ITAM.

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.

DRIDI, R., GUAY, A., AND RENAULT, E. (2007). Indirect inference and calibration of dynamic stochastic general equilibrium models. *J. Econometrics* **136(2)**, 397-430.

EINMAHL, U., AND MASON, D.M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, **33(3)**, 1380–1403.

FLORENS, J.-P., AND ROLIN, A. (1994). Bayes, bootstrap, moments. University of Toulouse.

HALL, A.R., AND INOUE, A. (2003). The large sample behavior of the generalized method of moments estimator in misspecified models. *J. Econometrics* **114(2)**, 361-394.

HANSEN, L.P. (1982). Large sample properties of Generalized-Method of Moments estimators. *Econometrica* **50(4)**, 1029-1054.

HIGHAM, N.J. (2007). Functions of matrices. Chapter 11 in *Handbook of Linear Algebra*, L. Hogben ed. Chapman & Hall/CRC: .

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*, second edition. Wiley: New-York.

Kiefer, N.M., and Vogelsang, T.J. (2005). A new asymptotic theory for heteroscedasticity-autocorrelation robust tests. *Econometric Theory* **21(6)**, 1130-1164.

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.

Li, Q., and Wang, S. (1998). A simple consistent bootstrap test for a parametric regression function. *J. Econometrics* **87(1)**, 145–165.

Major, P. (2006). An estimate on the supremum of a nice class of stochastic integrals and U-statistics. *Probab. Theory Related Fields* **134(3)**, 489-537.

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.

Marcellino, M., and Rossi, B. (2008). Model selection for nested and overlapping non-linear, dynamic and possibly misspecified models. Duke University.

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.

Parzen, M., and Lipsitz, S.R. (2007). Perturbing the minimand resampling with Gamma(1,1) random variables as an extension of the Bayesian Bootstrap. *Statist. and Prob. Letters* **77(6)**, 654-657.

Robinson, P.M. (1987). Asymptotically efficient estimation in the presence of heteroscedasticity of unknown form. *Econometrica* **55(4)**, 875-891.

Rubin,D. (1981). The Bayesian bootstrap. *Ann. Statist.* **9(1)**, 130-134.

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. published as 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.

Sun, Y., Phillips, P.C.B., and Jin, S. (2008). Optimal bandwidth selection in heteroscedasticity-autocorrelation robust testing. *Econometrica* **76(1)**, 175-194.

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.

Vuong, Q.H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57(2)**, 301-333.

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

Zheng, J.X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75(2)**, 263–289.

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

Table 1: Results for Dominguez and Lobato's setup

	$n = 50$			$n = 200$		
Estimator	h	Ratio RMSE	Ratio MAE	h	Ratio RMSE	Ratio MAE
			$X \sim N(0, 1)$			
NLS		0.0504	0.0390		0.0236	0.0186
DL		2.3590	2.2527		2.4795	2.4181
SMD	1	1.2828	1.2858	1	1.2626	1.2590
SMD	0.3	1.3298	1.3332	0.3	1.3110	1.3101
Eff. SMD	0.3	1.2160	1.2057	0.3	1.0952	1.0895
			$X \sim N(1, 1)$			
NLS		0.0226	0.0178		0.0109	0.0087
DL		2.1284	2.1362		2.2066	2.2157
SMD	1	1.2348	1.2363	1	1.2257	1.2313
SMD	0.3	1.2513	1.2522	0.3	1.2274	1.2319
Eff. SMD	0.3	1.1353	1.1299	0.3	1.0581	1.0603

The levels of RMSE and MAE, not their ratio, are reported for NLS.

Table 2: Results for Kitamura and al.'s setup

	$n = 50$			$n = 200$		
Estimator	h	Ratio RMSE	Ratio MAE	h	Ratio RMSE	Ratio MAE
GLS		0.1342	0.1066		0.0657	0.0523
		0.1623	0.1285		0.0795	0.0632
FGLS		1.2757	1.2345		1.3037	1.3944
		1.4323	1.3854		1.2347	1.3397
SEL	.3049	1.4266	1.3241	.2310	1.2894	1.1589
		1.2938	1.2279		1.1797	1.1077
	.7622	1.3015	1.2056	.5776	1.1608	1.0359
		1.1886	1.1522		1.0982	1.0917
	1.2195	1.3681	1.2166	.9242	1.1561	1.1047
		1.2170	1.1574		1.0940	1.1035
Eff. SMD	.3049	1.1660	1.1627	.2310	1.0967	1.0914
		1.2077	1.2073		1.1190	1.1150
	.7622	1.2968	1.2829	.5776	1.2128	1.2143
		1.2159	1.2070		1.1552	1.1524
	1.2195	1.4417	1.4338	.9242	1.3449	1.3499
		1.2719	1.2645		1.2170	1.2162

The levels of RMSE and MAE, not their ratio, are reported for GLS.

Table 3: Rejection percentages of bootstrap test

	$n = 50$			$n = 100$		
	h	5% level	10% level	h	5% level	10% level
Model (5.9) $X \sim N(0,1)$ $H_0: \ \theta_0 = 5/4$						
NLS		10.3	16.3		7.0	13.0
SMD	1	4.4	11.8	1	4.8	10.4
	.3	5.0	12.0	.3	5.0	9.4
Model (5.9) $X \sim N(1,1)$ $H_0: \ \theta_0 = 5/4$						
NLS		8.1	14.3		6.2	11.7
SMD	1	7.0	13.8	1	5.4	11.0
	.3	8.0	13.4	.3	5.6	10.2
Model (5.10) $H_0: \ \beta_2 = 1$						
FGLS		29.2	35.6		20.7	27.6
Eff. SMD	.3049	6.8	12.6	.2654	4.6	9.2
	.7622	9.8	15.2	.6635	6.2	11.2
	1.2195	11.8	18.6	1.0616	7.8	12.4

Figure 1: Estimators' densities: $X \sim N(0,1)$, $n = 50$

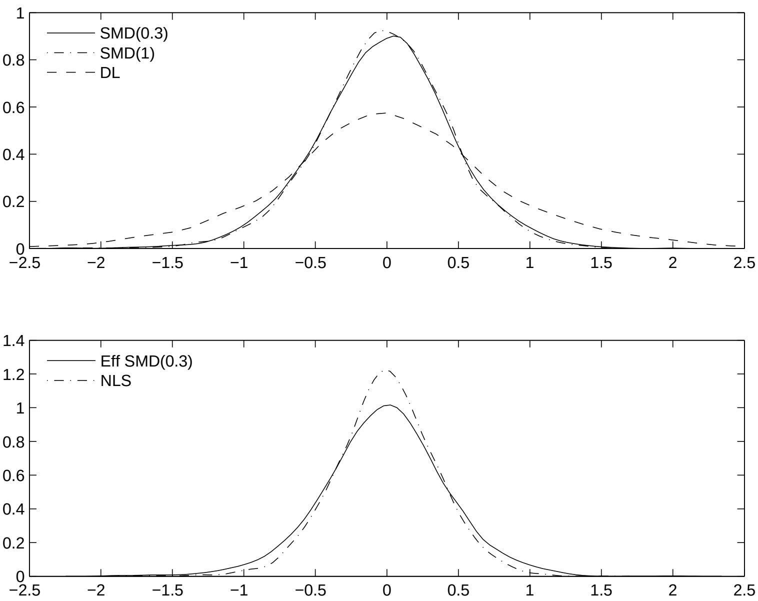

Figure 2: Estimators' densities: $X \sim N(0,1)$, $n = 200$

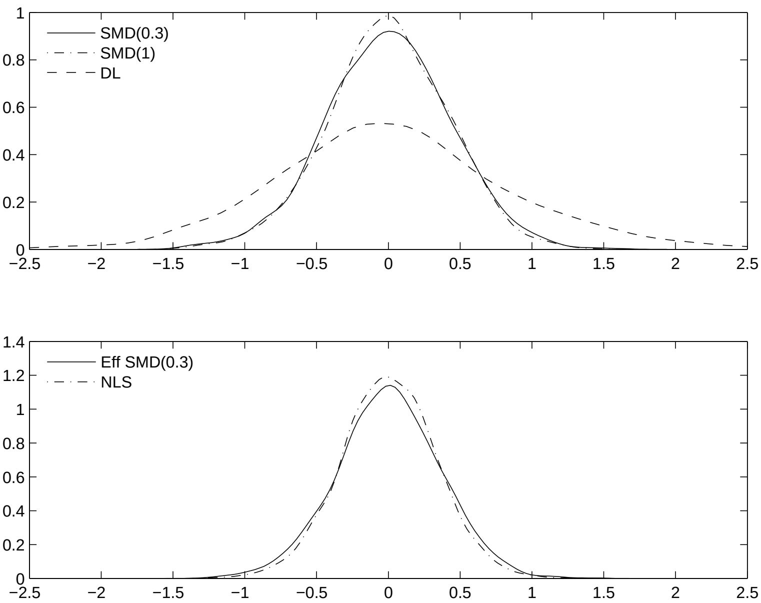

Figure 3: Estimators' densities: $X \sim N(1,1)$, $n = 50$

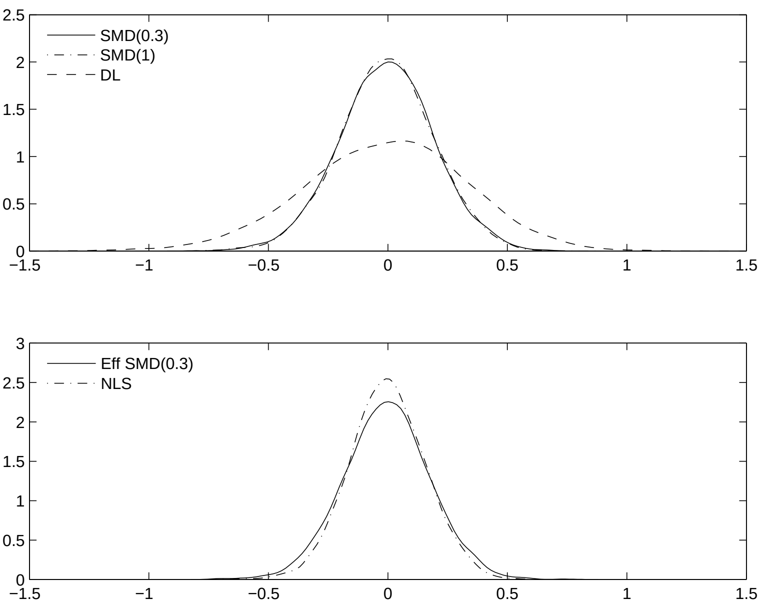

Figure 4: Estimators' densities: $X \sim N(1,1)$, $n = 200$

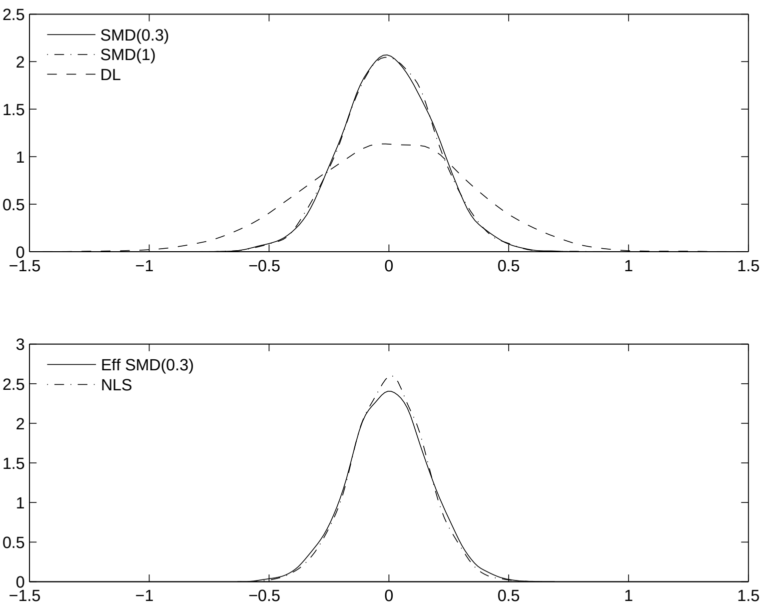

Figure 6: RMSE and bandwidth