UNIVERSITY OF TORONTO
DEPARTMENT OF ECONOMICS

Working Paper No. 274

Bandwidth Selection for Semiparametric Estimators Using the m-out-of-n Bootstrap

Chuan Goh
Department of Economics
University of Toronto
150 Saint George Street
Toronto, Ontario Canada
M5S 3G7

January 2007

Bandwidth Selection for Semiparametric Estimators Using the m-out-of-n Bootstrap*

Chuan Goh†

11 March 2002
Last update: 30 November 2006

Abstract

This paper considers a class of semiparametric estimators that take the form of density-weighted averages. These arise naturally in a consideration of semiparametric methods for the estimation of index and sample-selection models involving preliminary kernel density estimates. The question considered in this paper is that of selecting the degree of smoothing to be used in computing the preliminary density estimate. This paper proposes a bootstrap method for estimating the mean squared error and associated optimal bandwidth. The particular bootstrap method suggested here involves using a resample of smaller size than the original sample. This method of bandwidth selection is presented with specific reference to the case of estimators of average densities, of density-weighted average derivatives and of density-weighted conditional covariances.

JEL Classification: C14
KEYWORDS: Bandwidth selection, density-weighted averages, bootstrap, m-out-of-n bootstrap, kernel density estimation

*The research presented here is based on the second chapter of the author's doctoral dissertation, and as such, could not have been initiated without the advice and encouragement of Jim Powell, who served as his supervisor. The author also acknowledges financial support provided by the Graduate Division of the University of California, Berkeley and by the Connaught Fund of the University of Toronto. The usual disclaimer applies.

†Department of Economics, University of Toronto, 150 Saint George Street, Toronto, Ontario, Canada M5S 3G7. Telephone: +1 (416) 978-4964. Facsimile: +1 (416) 978-6713. goh@economics.utoronto.ca; http://www.chuangoh.org.

1 Introduction

This paper is concerned with the issue of smoothing parameter selection for nonparametric estimators that are used as components of a semiparametric estimator. In this case the relative importance of bias and variance from the perspective of bandwidth selection is different than it is when nonparametric estimators are considered on their own. In particular, semiparametric estimates of a Euclidean parameter that incorporate nonparametric kernel estimators will typically involve *asymptotic undersmoothing* of the kernel estimates—to guarantee the $\sqrt{n}$-consistency of the Euclidean parameter estimate, the bandwidth used to implement the nonparametric "ingredient" must converge more rapidly to zero than would be optimal for estimates of the corresponding function evaluated at points of interest in its domain.[1]

The class of semiparametric estimator considered in this paper involves the relatively simple case of procedures designed to estimate density-weighted expectations. Generally known as *density-weighted averages*, the implementation of these estimators involves the use of smoothing via embedded kernel functions. Despite their relative simplicity, an investigation of estimators of this sort is interesting because of the wide range of econometric scenarios in which these estimators can be applied. In particular, estimators in this class arise naturally in the consideration of semiparametric methods for the estimation of single-index and sample selection models involving preliminary kernel density estimates. It should also be noted that apart from the case of semiparametric estimation of density-weighted average derivatives, there is at this point still a paucity of research on how best to choose smoothing parameters in this setting.

For density-weighted averages involving kernel smoothing, the selection of bandwidths used to implement the preliminary kernel estimates is complicated by the fact that the asymptotic distribution of the normalized semiparametric estimator does not actually depend on the bandwidth used. Asymptotic approaches to bandwidth selection in this setting will therefore depend on the use of higher-order distributional approximations, as used for example in the case of density-weighted average derivatives by Härdle and Tsybakov (1993). In particular, Härdle and Tsybakov (1993) used a higher-order approximation to the distribution of the nor-

[1]The necessity for asymptotic undersmoothing of preliminary nonparametric estimates embedded in semiparametric estimates of Euclidean parameters was noted in the unifying theory elaborated by Goldstein and Messer (1992). For estimators exhibiting this feature in an econometric setting, cf. among others, the papers of Robinson (1988); Härdle and Stoker (1989) and Powell et al. (1989).

malized and centred average derivative estimator to construct a bandwidth minimizing an asymptotic approximation of the estimator's mean squared error. This approach to the construction of an asymptotically optimal bandwidth was also taken in the more general context considered by Powell and Stoker (1996). Both Härdle and Tsybakov (1993) and Powell and Stoker (1996, Proposition 4.1) show that the asymptotically optimal bandwidth for the estimation problems they consider has the form $h = kn^{-r}$, where n denotes the sample size and r is positive and depends on the order of the kernel function and the dimension of the conditioning variables involved. Powell and Stoker (1996, §4.4) describe a "plug-in" method for estimating the leading constant k for a class of estimators including that considered by Härdle and Tsybakov (1993).

This paper considers the specific estimation context adopted by Powell and Stoker (1996) and proposes a new method of estimating the asymptotically optimal bandwidth in applications. The approach taken in this paper was inspired by a suggestion of Horowitz (1998, §2.8) and involves the use of resampling fewer observations than are present in the original sample—the so-called "m-out-of-n" or "m-bootstrap".

The approach taken in this paper also complements existing methods based on resampling as many observations as exist in the original sample coupled with an explicit method of bias correction. The "manual" bias correction called for in this case arises out of the inability of the full-sample bootstrap to generate adequate approximations of the bias of the semiparametric estimator.[2] The approach taken in this paper avoids any need to engage in the sort of case-specific explicit bias correction required by approaches involving the full-sample bootstrap.

The remainder of this paper proceeds as follows. The following section presents a discussion of the specific estimation problem considered in this paper. Examples are presented in Section 2.1. Section 3 presents the main results of this paper demonstrating the efficacy of the m-bootstrap method in estimating mean squared error and estimating the asymptotically optimal bandwidth. The first part of Section 3 presents the regularity conditions presumed to underlie the structure of the mean squared errors of the estimators considered in this paper. The final portion of Section 3 deals with the important practical issue of how to select the resample size m. Section 4 contains the proof of the major theorem of Section 3, while Section 5 presents the results of a simulation experiment comparing the performance in a small sample of the bootstrap method of bandwidth selection proposed here

[2]Cf. e.g., the approach taken by Nishiyama and Robinson (2005) for the case of semiparametric estimators of density-weighted average derivatives.

with the plug-in method suggested by Powell and Stoker (1996, §4.4). Section 6 concludes. Proofs of certain lemmas not presented in the main text appear in the appendix.

A word on notation

In what follows, $K : \mathbb{R}^d \to \mathbb{R}$ will generally be referred to as a "smoothing kernel". K will be assumed to satisfy $K(u) = K(-u)$ and $\int K(u)du = 1$. The symbol h will always denote a positive scalar-valued function of an integer n such that $h \equiv h(n) \to 0$ as $n \to \infty$. In this connection, $\hat{f}_n(\cdot, h)$ will always denote a "leave-out" kernel estimate based on a random sample $X_1, \ldots, X_n$. In particular,

$$\hat{f}_n(X_i, h) \equiv \frac{1}{n-1} \sum_{\{j: j \neq i\}} \frac{1}{h^d} K\left(\frac{X_i - X_j}{h}\right). \tag{1}$$

The symbol $\|\cdot\|$ will denote the usual Euclidean norm of a vector.[3]

2 Density-weighted averages

Suppose the data represent an iid sample of observations $X_1, \ldots, X_n$, where X_i, for $i \in \{1, \ldots, n\}$, is a d-vector of response and conditioning variables. The estimators considered in this paper take the form of second-order U-statistics with kernel functions depending on a smoothing parameter h. In particular, we consider estimators in the form

$$\hat{\theta}_n(h) \equiv \binom{n}{2}^{-1} \sum_{i<j} g(X_i, X_j, h), \tag{2}$$

where $g(\cdot, \cdot, h)$ is a function symmetric in pairs of observations[4] and h is a (scalar) smoothing parameter such that $h \equiv h(n) \to 0$ as $n \to \infty$.

Remark 1. *In what follows, the function $g(\cdot, \cdot, h)$ will be taken to be scalar-valued when it is convenient to do so. In applications involving a vector-valued U-statistic kernel, it is possible to extend the derivations for the scalar-valued case by applying them to single components of $g(\cdot, \cdot, h)$ and of $\hat{\theta}_n(h)$, and then subsequently deducing desired results for arbitrary linear combinations $\lambda^T g(\cdot, \cdot, h)$ and $\lambda^T \hat{\theta}_n(h)$.*

[3] Cf. e.g., Hall and Marron (1987), Jones and Sheather (1991) for discussion of issues related to the omission of the "$i = j$" terms in the kernel estimate (1).

[4] i.e., $g(X_i, X_j, h) = g(X_j, X_i, h)$. The estimator $\hat{\theta}_n(h)$ can be described as a second-order U-statistic with kernel $g(\cdot, \cdot, h)$.

Denote the expectation of $\hat{\theta}_n(h)$ as

$$\theta(h) \equiv E\left[\hat{\theta}_n(h)\right] = E[g(X_1, X_2, h)], \tag{3}$$

and the object of estimation as

$$\theta_0 \equiv \lim_{h \to 0} \theta(h). \tag{4}$$

The characterization of the estimator $\hat{\theta}_n(h)$ as a *density-weighted average* arises from the fact that the second-order U-statistic structure given above in (2) arises frequently when the estimand θ_0 has the form of a density-weighted expectation—this is indeed the case in each of the examples presented below in Section 2.1.

2.1 Examples

This section reproduces the three motivating examples of Powell and Stoker (1996). Although the material in this section has primarily been included to make this paper more self-contained than it otherwise would be, it does serve the purpose of linking the notation of (2)–(4) to concrete examples.

Example 1 (Average densities)**.** *Suppose $X_i \in \mathbb{R}^d$ is a continuous random vector with density f. The objective is to estimate*

$$\theta_0 = \int f(x_1)^2 dx_1 = E[f(X_1)]$$

with

$$\hat{\theta}_n(h) = \frac{1}{n} \sum_{i=1}^{n} \hat{f}_n(X_i, h),$$

where $\hat{f}_n(\cdot, h)$ is a kernel estimate given by

$$\hat{f}_n(X_i, h) \equiv \frac{1}{n-1} \sum_{\{j: j \neq i\}} \frac{1}{h^d} K\left(\frac{X_i - X_j}{h}\right),$$

which indicates a special case of the estimator given above in (2) with scalar-valued U-statistic kernel

$$g(X_i, X_j, h) = \frac{1}{h^d} K\left(\frac{X_i - X_j}{h}\right).$$

Remark 2. *Although it generally seems to be of little more than paedagogical interest, it should be noted at this point that the problem of estimating the average density, as presented in Example 1, arises in the measurement of the variance of rank estimators.*[5] *On the other hand, Examples 2 and 3 below describe estimators that seem to have become widely accepted in econometrics. Further discussion, along with examples of their application, can be found in Powell (1994) and Horowitz (1998).*

Example 2 (Density-weighted average derivatives)**.** *Suppose* $X_i = (Y_i, Z_i^T)^T$, $Y_i \in \mathbb{R}$, $Z_i \in \mathbb{R}^d$. *Assume that* Z_i *is an absolutely continuous random vector with density* f. *Let* $m(Z_i) \equiv E[Y_i|Z_i]$. *We want to estimate*

$$\theta_0 = E\left[f(Z_1)\frac{\partial m(Z_1)}{\partial Z_1}\right] = -2E\left[\frac{\partial f(Z_1)}{\partial Z_1}Y_1\right],$$

where we assume $f(z)m(z) \to 0$ *as* $\|z\| \to \infty$ *and that all derivatives and moments exist.* θ_0 *is important because it is proportional to the coefficients of a semiparametric index model—i.e., if* $m(Z_i) \equiv M(Z_i^T\beta)$, *then* θ_0 *is proportional to* β. *If* $\hat{f}_n(Z_i, h)$ *is a kernel estimate of the density of* Z_i, *the estimator* $\hat{\theta}_n(h)$ *is given by*

$$\hat{\theta}_n(h) \equiv \frac{1}{n}\sum_{i=1}^{n} -2Y_i\frac{\partial \hat{f}_n(Z_i, h)}{\partial Z_i},$$

which in turn yields a special case of the estimator given above in (2) with U-statistic kernel

$$g(X_i, X_j, h) = -\frac{1}{h^{d+1}}K'\left(\frac{Z_i - Z_j}{h}\right)(Y_i - Y_j),$$

where $K'(\cdot)$ *denotes the derivative of a smoothing kernel* $K(\cdot)$, *where* $K(\cdot)$ *has the generic properties assumed above in the discussion appearing at the end of the Introduction.*

Example 3 (Density-weighted conditional covariances)**.** *Suppose*

$$X_i = (Y_i, Z_i^T, W_i^T)^T,$$

$Y_i \in \mathbb{R}$, $Z_i \in \mathbb{R}^l$, *and* $W_i \in \mathbb{R}^d$. *Assume* W_i *is an absolutely continuous random vector with density* f. *The objective here is to estimate the density-weighted conditional covariances*

$$\theta_0^y = E[f(W_1)(Z_1 - E[Z_1|W_1])(Y_1 - E[Y_1|W_1])]$$

[5]Cf. e.g., Jurečková (1971); Jaeckel (1972).

and

$$\theta_0^z = E[f(W_1)(Z_1 - E[Z_1|W_1])(Z_1 - E[Z_1|W_1])^T].$$

These are relevant for example in the estimation of the partially linear model

$$Y_i = Z_i^T\beta + \upsilon(W_i) + U_i,$$

where $E[U_i|Z_i, W_i] = 0$. Assuming that θ_0^z is nonsingular and $\upsilon(\cdot)$ is sufficiently smooth we can write

$$\beta = [\theta_0^z]^{-1}\theta_0^y.$$

θ_0^y and θ_0^z are estimated by estimators in the form given above in (2) with U-statistic kernels given by

$$g^y(X_i, X_j, h) = \frac{1}{2h^d}K\left(\frac{W_i - W_j}{h}\right)(Z_i - Z_j)(Y_i - Y_j) \tag{5}$$

and

$$g^z(X_i, X_j, h) = \frac{1}{2h^d}K\left(\frac{W_i - W_j}{h}\right)(Z_i - Z_j)(Z_i - Z_j)^T, \tag{6}$$

respectively.

Remark 3. *In the context of estimation of density-weighted conditional covariances as given in Example 3, the focus of the development that follows is the issue of bandwidth selection for estimators of the components θ_0^y and θ_0^z of β. Powell and Stoker (1996, §5) discuss the issue of bandwidth choice for the ratio $\hat{\beta}_n(h) \equiv \left[\hat{\theta}_n^z(h)\right]^{-1}\hat{\theta}_n^y(h)$. The form of the asymptotic MSE-minimizing bandwidth for the ratio $\hat{\beta}_n(h)$ can be shown fairly straightforwardly to be identical to that used for estimating the density-weighted average given by*

$$\hat{\theta}_n^{zu}(h) \equiv [\theta_0^z]^{-1}\,\hat{\theta}_n^u(h). \tag{7}$$

Here $\hat{\theta}_n^u(h)$ refers to a second-order U-statistic of the basic form given in (2) with kernel

$$u(X_i, X_j, h) \equiv g^y(X_i, X_j, h) - g^z(X_i, X_j, h)\beta,$$

where $g^y(\cdot)$ and $g^z(\cdot)$ are the U-statistic kernels given above in (5) and (6), respectively. The asymptotically MSE-optimal bandwidth can be shown to be consistently estimable using preliminary consistent estimates of θ_0^z and β. Further details can be found in the discussion leading up to the statement of Powell and

Stoker (1996, Proposition 5.1). The essential equivalence between the bandwidth selection problem for ratios of density-weighted averages and the corresponding problem for the density-weighted average given by $\hat{\theta}_n^{zu}(h)$ in (7) above makes the extension of the method of bandwidth selection proposed in this paper to the case of ratios of density-weighted averages fairly straightforward.[6]

3 Main Results

Details of the proposal to estimate the mean squared error and select the optimal bandwidth for the class of estimators under consideration appear in Section 3.2 below. These details are preceded by a signficant amount of preliminary discussion in Section 3.1 setting out the relevant notation and regularity conditions that are presumed to underlie the analysis. The estimand will be taken to be scalar-valued for the sake of convenience.[7]

3.1 Setup and assumptions

Define the following, with reference to the expressions given above in (2) and (3):

$$\bar{g}(X_i, h) \equiv E\left[g(X_i, h)|X_i\right] \tag{8}$$

$$\bar{g}_0(X_i) \equiv \lim_{h \to 0} \bar{g}(X_i, h) \tag{9}$$

$$\bar{\theta}_n(h) \equiv \sum_{i=1}^{n} \bar{g}(X_i, h) - (n-1)\theta(h) \tag{10}$$

It is to be noted that the asymptotic behaviour of the basic class of estimator given above as $\hat{\theta}_n(h)$ in (2) depends to a large extent on various properties of the function $\bar{g}_0(X_i)$. In particular, under regularity conditions that ensure the

[6]The careful reader will note that the possibility of employing different bandwidths for estimating the two components of the ratio

$$\hat{\beta}_n(h) \equiv \left[\hat{\theta}_n^z(h)\right]^{-1} \hat{\theta}_n^y(h)$$

(i.e., what might be referred to loosely as the "numerator" and "denominator") has been excluded by the use of a single "h" in the argument of $\hat{\beta}_n(\cdot)$. This is motivated by what appears to be most prevalent in empirical practice.

[7]Cf. the comments in Remark 1 above.

$\sqrt{n}$-asymptotic normality of $\hat{\theta}_n(h)$,[8] the asymptotic behaviour of $\hat{\theta}_n(h) - \theta_0$ is equivalent to that of

$$\frac{2}{n}\sum_{i=1}^{n}\left(\bar{g}_0(X_i) - \theta_0\right),$$

where θ_0 is the estimand given above in (4). In addition, a natural nonparametric estimator of the asymptotic variance of $\hat{\theta}_n(h)$ under these regularity conditions is the empirical variance of $2\bar{g}(X_i, h)$. Given the importance of the functions $\bar{g}(X_i, h)$ and $\bar{g}_0(X_i)$ to the large-sample theory of $\hat{\theta}_n(h)$ under conditions sufficient for $\sqrt{n}$-asymptotic normality, the regularity conditions underlying the development that follows will be couched in terms of the behaviour of $\bar{g}(X_i, h)$ and $\bar{g}_0(X_i)$. To wit:

Assumption 1. *It is always possible to interchange the expectation and* $\lim_{h\to 0}$ *operators. In particular,*

$$E[\bar{g}_0(X_1)] = \theta_0,$$

where θ_0 *is the estimand given above in (4).*

Assumption 2. *There exists a constant* $\alpha > 0$ *and a function* $s(\cdot)$ *with* $E[s(X_1)] \neq 0$ *such that*

$$\bar{g}(X_i, h) - \bar{g}_0(X_i) = s(X_i)h^{\alpha} + \bar{s}(X_i, h),$$

where $E[\|\bar{s}(X_1, h)\|^2] = o(h^{2\alpha})$.

Assumption 3. *There exists a constant* $\gamma > 0$ *and a function* $q(\cdot)$ *with* $E[q(X_1)] \neq 0$ *such that*

$$E\left[\|g(X_1, X_2, h)\|^2 | X_1\right] = q(X_1)h^{-\gamma} + \bar{q}(X_1, h),$$

where $E[\|\bar{q}(X_1, h)\|] = o(h^{-\gamma})$.

Assumption 4. $E\left[\|\bar{g}_0(X_1\|^4)\right] < \infty$*, and for the same constant* $\gamma > 0$ *specified in Assumption 3,*

$$E\left[\|g(X_1, X_2, h)\|^4\right] = O\left(h^{-3\gamma}\right).$$

Assumption 5. $2\alpha > \gamma$*, where* α *and* γ *are the constants specified in Assumptions 2 and 3 above.*

[8] Cf. e.g., Powell et al. (1989, Assumptions 1–4).

Assumption 1 seems necessary to rule out pathological behaviour. Assumptions 2 and 3 also figure in the development of Powell and Stoker (1996), and serve to regulate the bias and variance, respectively, of $\bar{g}(X_i, h)$ as an estimator of $\bar{g}_0(X_i)$. From (3) above, it is also clear that Assumption 2 regulates the bias of $\hat{\theta}_n(h)$ as an estimator of θ_0, since an immediate implication of this condition is that

$$\theta(h) - \theta_0 = E[s(X_1)]h^\alpha + o(h^\alpha). \tag{11}$$

Assumption 3 can also be seen to have an important role in the unconditional variance of the U-statistic kernel $g(\cdot)$, since an immediate consequence is that

$$E\left[\|g(X_1, X_2, h)\|^2\right] = E[q(X_1)]h^{-\gamma} + o(h^{-\gamma}). \tag{12}$$

We note that Assumptions 2 and 3 are satisfied by the three examples in Section 2.1. In particular, in the context of each example presented in Section 2.1, the constant α in Assumption 2 is the order of the kernel function $K(\cdot)$. In general, however, the value of α is reflected in the structure of the U-statistic kernel $g(\cdot)$. The constant γ in Assumption 3 is equal to d in Examples 1 and 3 of Section 2.1, and is equal to $d + 2$ in the context of Example 2.

Assumption 4 is stronger than the more conventional condition of $E[\|\bar{g}_0(X_1)\|^2] < \infty$ typically imposed to ensure the asymptotic normality at rate $\sqrt{n}$ of $\hat{\theta}_n(h)$. The conditions of Assumption 4 will be shown in Section 4 to be important in guaranteeing the efficacy of the bootstrap method proposed in this paper for estimating the mean-squared error of $\hat{\theta}_n(h)$.

Finally, the conditions on the bandwidth sequence $\{h\} = \{h_n\}$ that are necessary for the $\sqrt{n}$-asymptotic normality of $\hat{\theta}_n(h)$ as an estimator of θ_0 are noted. These conditions are given in terms of the constants α and γ specified above in Assumptions 2 and 3, respectively. In particular, the condition

$$h = o\left(n^{-\frac{1}{2\alpha}}\right) \tag{13}$$

is necessary for $\sqrt{n}\,(\theta(h) - \theta_0) = o(1)$, while

$$h^{-1} = o\left(n^{\frac{1}{\gamma}}\right) \tag{14}$$

is required for

$$E\left[\|g(X_1, X_2, h)\|^2\right] = o(n). \tag{15}$$

The condition (15) is in turn necessary for the variance of $\hat{\theta}_n(h)$ to disappear at rate n.[9]

[9] Cf. Powell et al. (1989, Lemma 3.1).

It is clear that the conditions given as (13) and (14) above bound the rate at which h is permitted to converge to zero if the estimator $\hat{\theta}_n(h)$ is to be $\sqrt{n}$-asymptotically normal. As such, Assumption 5 is required in order for both (13) and (14) to hold simultaneously.

Let $\mathcal{H}_n(\alpha, \gamma)$ denote the set of all bandwidths $h = h_n \in (0, \infty)$ satisfying both (13) and (14) above. The statistical problem considered in this paper is to find the mean squared error-minimizing bandwidth within the set $\mathcal{H}_n(\alpha, \gamma)$ for semi-parametric estimation contexts satisfying Assumptions 1–5 above. As first-order asymptotic theory provides no guidance in this situation beyond that provided by the bounds in (13) and (14), the route taken here is to resort to an expansion of the mean squared error function under Assumptions 1–5 that holds for all bandwidths $h \in \mathcal{H}_n(\alpha, \gamma)$. The MSE-optimal bandwidth sequence will be taken to be that that minimizes the leading terms of this expansion.

In order to develop the expansion of the mean squared error function for estimators in the form given as $\hat{\theta}_n(h)$ in (2) above, note from the theory of U-statistics[10] that the finite-sample variance of $\hat{\theta}_n(h)$ is given by

$$\begin{aligned} Var\left[\hat{\theta}_n(h)\right] &= \binom{n}{2}^{-1}\left\{2(n-2)Var\left[\bar{g}(X_1,h)\right] + Var\left[g(X_1,X_2,h)\right]\right\} \\ &= \frac{4}{n}Var\left[\bar{g}(X_1,h)\right] + \frac{2}{n^2}E\left[g^2(X_1,X_2,h)\right] + o(n^{-2}). \end{aligned} \tag{16}$$

Invoking Assumption 2 produces the following characterization of $Var\left[\bar{g}(X_1,h)\right]$:

$$Var\left[\bar{g}(X_1,h)\right] = Var\left[\bar{g}_0(X_1)\right] + 2Cov\left[\bar{g}_0(X_1), s(X_1)\right]h^\alpha + o(h^\alpha). \tag{17}$$

Combining the statements of Assumptions 2 and 3 along with (16) and (17) results in the representation for the mean squared error of $\hat{\theta}_n(h)$ given by

$$\begin{aligned} MSE\left[\hat{\theta}_n(h)\right] &= (\theta(h)-\theta_0)^2 + Var\left[\hat{\theta}_n(h)\right] \\ &= \left\{E\left[s(X_1)\right]\right\}^2 h^{2\alpha} + \frac{4}{n}Var\left[\bar{g}_0(X_1)\right] + \frac{4}{n}Cov\left[\bar{g}_0(X_1), s(X_1)\right]h^\alpha \\ &\quad + \frac{2}{n^2}E\left[q(X_1)\right]h^{-\gamma} + o(h^{2\alpha}) + o(n^{-1}h^\alpha) + o(n^{-2}h^{-\gamma}) \\ &\quad + o(n^{-2}). \end{aligned} \tag{18}$$

It is clear that the first and third terms in (18) are increasing functions of the bandwidth, while the fourth term is decreasing in h. The following elementary argu-

[10]Cf. e.g., Serfling (1980, §5.2).

ment clarifies the order of the largest terms in (18) at an MSE-optimal bandwidth sequence.

Lemma 1. *Suppose $h = h_n$ belongs to a sequence that minimizes (18). In this case $\{E[s(X_1)]\}^2 h^{2\alpha}$ is of larger order than $\frac{4}{n}Cov[\bar{g}_0(X_1), s(X_1)]\, h^{\alpha}$.*

Proof. Suppose not. Then the bandwidth sequence $\{h_n\}$ equating the order of the order $O(n^{-1}h^{\alpha})$ and $O(n^{-2}h^{-\gamma})$ terms in (18) would satisfy $h_n \propto n^{-\frac{1}{\alpha+\gamma}}$, thus implying that $h_n^{2\alpha} \propto n^{-\frac{2\alpha}{\alpha+\gamma}}$ and that $n^{-1}h_n^{\alpha} \propto n^{-\frac{2\alpha+\gamma}{\alpha+\gamma}}$. This, however, implies that $\{E[s(X_1)]\}^2 h_n^{2\alpha}$ is of greater order than $\frac{4}{n}Cov[\bar{g}_0(X_1), s(X_1)]\, h_n^{\alpha}$, a contradiction. □

It is immediate from Lemma 1 that mean squared-error minimization under the assumptions made above proceeds on the basis of equating the orders of $\{E[s(X_1)]\}^2 h^{2\alpha}$ and of $\frac{2}{n^2}E[q(X_1)]\, h^{-\gamma}$. As such, the MSE-optimal bandwidth sequence $\{h_n\}$ satisfies

$$h_n \propto n^{-\frac{2}{2\alpha+\gamma}}. \tag{19}$$

The foregoing discussion is summarized in the following theorem characterizing the mean squared error of the class of estimators under consideration.

Theorem 1. *Under Assumptions 1–5 we have for $n \to \infty$ that*

$$\begin{aligned} MSE\left[\hat{\theta}_n(h_n)\right] &= \frac{4}{n}Var[\bar{g}_0(X_1)] + \{E[s(X_1)]\}^2 h_n^{2\alpha} + \frac{2}{n^2}E[q(X_1)]\, h_n^{-\gamma} \\ &\quad + o\left(n^{-\frac{4\alpha}{2\alpha+\gamma}}\right) \end{aligned}$$

for all $h_n \propto n^{-\frac{2}{2\alpha+\gamma}}$.

Powell and Stoker (1996, Proposition 4.1) exploit the conclusion of Theorem 1 to derive an approximation of the theoretically MSE-optimal bandwidth, which is seen to depend on the functions introduced in Assumptions 2 and 3.[11] As such, estimates of the optimal bandwidth necessarily involve—perhaps in an implicit

[11] In particular, the approximation given in Powell and Stoker (1996, Proposition 4.1) to the MSE-optimal bandwidth has the form

$$h_n = \left\{ \frac{\gamma E[q(X_1)]}{\alpha \left(E[s(X_1)]\right)^2} \right\}^{\frac{1}{2\alpha+\gamma}} n^{-\frac{2}{2\alpha+\gamma}}.$$

fashion—estimation of the bias quantity $s(\cdot)$ introduced in Assumption 2. It follows that one might be concerned about the quality of estimates of the leading constant in the optimal bandwidth that are in turn based on estimation methods that deliver poor estimates of the bias of $\hat{\theta}_n(h)$ with respect to θ_0.

It is clear from (19) above that the MSE-optimal bandwidth h_m for a sample of size $m < n$ will have the form $h_m = km^{-\frac{2}{2\alpha+\gamma}}$, where k is the same leading constant appearing in the formula for the optimal bandwidth appropriate for a sample of size n. For this reason, it is plausible to base estimates of the optimal bandwidth h_n for the full sample of observations on estimates of the leading constant k obtained using subsamples of size $m < n$. This consideration is particularly relevant in cases where k is difficult to estimate using the full sample. While the bootstrap method presented in the following section does not involve explicit estimation of the leading constant k, it is nevertheless effective for essentially the reason just noted.

3.2 Bootstrap estimates of mean squared error

This section presents the details of this paper's proposal for estimating $MSE\left[\hat{\theta}_n(h_n)\right]$ and the associated optimal bandwidth via a resampling procedure. In this connection, further notation and a number of definitions are introduced. Let $\mathcal{X}_n \equiv \{X_1, \ldots, X_n\}$ denote the original random sample of n observations. For $m \leq n$, let $\mathcal{X}_m^* \equiv \{X_1^*, \ldots, X_m^*\}$ denote a random sample of size m from $\mathcal{X}_n$. In what follows, $\mathcal{X}_m^*$ will generally be referred to as an *m-bootstrap sample*. Let $\{h\} \equiv \{h_n\}$ and $\{h_m\}$ denote bandwidth sequences appropriate for samples of sizes n and m respectively, where $h_n, h_m \to 0$ as $n, m \to \infty$. Define the following m-bootstrap analogues of $\hat{\theta}_n(h)$, $\bar{g}(X_i, h)$, $\theta(h)$ and $\bar{\theta}_n(h)$:

$$\hat{\theta}_m^*(h_m) \equiv \binom{m}{2}^{-1} \sum_{i<j} g(X_i^*, X_j^*, h_m); \tag{20}$$

$$\bar{g}^*(X_i^*, h_m) \equiv E\left[g(X_i^*, X_2^*, h_m) \middle| X_i^*, \mathcal{X}_n\right]; \tag{21}$$

$$\theta^*(h_m) \equiv E\left[\bar{g}^*(X_i^*, h_m) \middle| \mathcal{X}_n\right]; \tag{22}$$

$$\bar{\theta}_m^*(h_m) \equiv \theta^*(h_m) + \frac{2}{m} \sum_{i=1}^{m} \left(\bar{g}^*(X_i^*, h_m) - \theta^*(h_m)\right). \tag{23}$$

For clarity, note that

$$\bar{g}^*(X_i^*, h_m) = \frac{1}{n}\sum_{j=1}^{n} g(X_i^*, X_j, h_m); \tag{24}$$

and that

$$\theta^*(h_m) = \frac{1}{n}\sum_{i=1}^{n} \bar{g}^*(X_i, h_m) = \frac{1}{n^2}\sum_{i,j} g(X_i, X_j, h_m). \tag{25}$$

We also have that $\theta^*(h_m) = E\left[g(X_i^*, X_j^*, h_m) \middle| \mathcal{X}_n \right]$, and that

$$E\left[\hat{\theta}_m^*(h_m) \middle| \mathcal{X}_n \right] = \binom{n}{2}^{-1} \sum_{i<j} g(X_i, X_j, h_m) = \hat{\theta}_n(h_m). \tag{26}$$

A natural m-bootstrap estimate of the variance of $\hat{\theta}_n(h)$ is given by $Var\left[\hat{\theta}_m^*(h_m) \middle| \mathcal{X}_n \right]$, whose form is easily derivable from standard discussions of U-statistic theory:[12]

$$\begin{aligned} Var\left[\hat{\theta}_m^*(h_m) \middle| \mathcal{X}_n \right] &= \binom{m}{2}^{-1} \left\{ 2(m-2) Var\left[\bar{g}^*(X_i^*, h_m) \middle| \mathcal{X}_n \right] \right. \\ &\quad \left. + Var\left[g(X_i^*, X_j^*, h_m) \middle| \mathcal{X}_n \right] \right\}, \end{aligned} \tag{27}$$

where

$$\begin{aligned} Var\left[\bar{g}^*(X_i^*, h_m) \middle| \mathcal{X}_n \right] &= \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j,k} g(X_i, X_j, h_m) g(X_i, X_k, h_m) \\ &\quad - \left\{ \frac{1}{n^2} \sum_{i,j} g(X_i, X_j, h_m) \right\}^2; \\ Var\left[g(X_i^*, X_j^*, h_m) \middle| \mathcal{X}_n \right] &= \frac{1}{n^2} \sum_{i,j} g^2(X_i, X_j, h_m) - \left\{ \frac{1}{n^2} \sum_{i,j} g(X_i, X_j, h_m) \right\}^2. \end{aligned}$$

On the other hand, the bias of $\hat{\theta}_n(h)$ as an estimator of θ_0 can be estimated by

$$E\left[\hat{\theta}_m^*(h_m) \middle| \mathcal{X}_n \right] - \hat{\theta}_n(h),$$

[12]Cf. e.g., Serfling (1980, §5.2).

which in light of (26) is given by

$$\hat{\theta}_n(h_m) - \hat{\theta}_n(h). \tag{28}$$

This bias estimate is clearly not useful when $h_m = h = h_n$—in this case the expression in (28) is identically equal to zero—but one would expect practitioners to set $h_m > h_n$ when $m < n$.

Combining (27) and (28) produces an m-bootstrap estimate of the mean squared error of $\hat{\theta}_n(h)$ as an estimate of θ_0. In particular

$$\begin{aligned} MSE\left[\hat{\theta}^*_m(h_m)\middle|\, \mathcal{X}_n\right] &\equiv \left(\hat{\theta}_n(h_m) - \hat{\theta}_n(h)\right)^2 + Var\left[\hat{\theta}^*_m(h_m)\middle|\, \mathcal{X}_n\right] \\ &= E\left[\left(\hat{\theta}^*_m(h_m) - \hat{\theta}_n(h)\right)^2\middle|\, \mathcal{X}_n\right] \end{aligned} \tag{29}$$

denotes the m-bootstrap estimate of $MSE\left[\hat{\theta}_n(h)\right]$. The efficacy of the m-bootstrap in this setting is shown by the conclusion of Theorem 2, which indicates that under certain conditions additional to those assumed above in Theorem 1, the statistic given in (29) is so close in a uniform sense to $MSE\left[\hat{\theta}_m(h_m)\right]$—i.e., to the true mean squared error of the estimator computed using a sample of size $m < n$—that the bandwidth sequence minimizing $MSE\left[\hat{\theta}^*_m(h_m)\middle|\, \mathcal{X}_n\right]$ will also minimize $MSE\left[\hat{\theta}_m(h_m)\right]$. The bandwidth sequence minimizing the m-bootstrap estimate of $MSE\left[\hat{\theta}_m(h_m)\right]$ can then be converted into a bandwidth estimate that is asymptotically equivalent to the MSE-optimal bandwidth for $\hat{\theta}_n(h)$—i.e., the estimator computed using the full sample—by an appropriate rescaling for the amount by which n and m differ. The main result of this paper is stated as follows.

Theorem 2. *Suppose that Assumptions 1–5 hold, and that h_n satisfies both (13) and (14) above. Then as $m, n \to \infty$ with $m \propto n^\delta$ for some constant $\delta \in (0,1)$ with $\delta \leq 1-\delta$,*

$$\begin{aligned} MSE\left[\hat{\theta}^*_m(h_m)\middle|\, \mathcal{X}_n\right] &= \frac{4}{m} Var\left[\bar{g}_0(X_1)\right] + \left\{E\left[s(X_1)\right]\right\}^2 h_m^{2\alpha} + \frac{2}{m^2} E\left[q(X_1)\right] h_m^{-\gamma} \\ &\quad + o_p\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right), \end{aligned}$$

where $h_m = k m^{-\frac{2}{2\alpha+\gamma}}$ for any $k \in (0,\infty)$. This representation holds uniformly for $k \in [\epsilon, \epsilon^{-1}]$, where the constant $\epsilon > 0$ is arbitrary.

Proof. The proof is deferred to Section 4 below. □

Remark 4. *The condition on the resample size* m *given in the statement of Theorem 2 is more stringent than the more commonly encountered requirement that* $m \to \infty$ *with* $m = o(n)$*. The requirement that* $m \propto n^{\delta}$ *for some* $\delta \in (0, 1)$ *with* $\delta \le 1 - \delta$ *ensures that* m *is sufficiently small so that the stochastic remainder term in the expansion of* $MSE\left[\hat{\theta}^*_m(h_m)\middle|\, \mathcal{X}_n\right]$ *is of the same order of magnitude as the remainder term in the expansion of* $MSE\left[\hat{\theta}_m(h_m)\right]$*. Further details appear in Section 4. In Section 3.3 below, it is shown that under the conditions of Theorem 2, a setting of* $\delta = \frac{1}{2}$ *is optimal from the point of view of minimizing the discrepancy between the* m*-bootstrap MSE estimate and the leading terms in the expansion of* $MSE\left[\hat{\theta}_m(h_m)\right]$ *as given in the statement of Theorem 1.*

Remark 5. *It is clear from (29) and the statement of Theorem 2 that the* m*-bootstrap estimate of* $MSE\left[\hat{\theta}_n(h)\right]$ *incorporates a bias estimate involving a pilot bandwidth* $h_n = h(n)$ *satisfying both (13) and (14) above. The proof of Theorem 2 proceeds on the assumption that* h_n *is nonstochastic. In cases when practitioners choose a data-dependent pilot bandwidth* $\hat{h}_n$*,*[13] *a subsidiary argument regarding the asymptotic equivalence of* $\hat{\theta}_n(\hat{h}_n)$ *and* $\hat{\theta}_n(h_n)$ *for a nonstochastic* h_n *satisfying* $\frac{\hat{h}_n}{h_n} \xrightarrow{p} 1$ *would seem to be in order.*[14]

It is immediate from Theorems 1 and 2 that for bandwidth sequences satisfying $h_m = km^{-\frac{2}{2\alpha+\gamma}}$, the m-bootstrap estimate $MSE\left[\hat{\theta}^*_m(h_m)\middle|\, \mathcal{X}_n\right]$ is asymptotically equivalent to $MSE\left[\hat{\theta}_m(h_m)\right]$ up to terms of order $m^{-\frac{4\alpha}{2\alpha+\gamma}}$, where for $\epsilon > 0$ that may be made arbitrarily small, the asymptotic equivalence holds uniformly for values of $k \in [\epsilon, \epsilon^{-1}]$. As such, minimization of the m-bootstrap MSE estimate proposed here produces a bandwidth sequence $\{\hat{h}_m\}$ that is asymptotically equivalent to the sequence of bandwidths that minimize the leading terms of $MSE\left[\hat{\theta}_m(h_m)\right]$. In particular, if h_m is the nonstochastic bandwidth minimizing $MSE\left[\hat{\theta}_m(h_m)\right]$, then the bandwidth $\hat{h}_m$ minimizing $MSE\left[\hat{\theta}^*_m(h_m)\middle|\, \mathcal{X}_n\right]$ satis-

[13]For example, $\hat{h}_n$ might be computed as a plug-in estimate of the MSE-optimal bandwidth according to the proposal of Powell and Stoker (1996, §4.4).

[14]Unfortunately, it is not clear if such a subsidiary argument can be made without a substantial strengthening of the conditions given in the statement of Theorem 2. Cf. also the comments on this subject by Powell and Stoker (1996, p. 311–312).

fies

$$\frac{\hat{h}_m}{h_m} \xrightarrow{p} 1 \tag{30}$$

as $m, n \to \infty$ in accordance with the requirements of Theorem 2. The bandwidth $\hat{h}_m$ is easily seen to involve an implicit estimate of the leading constant in the theoretically optimal bandwidth h_m. As such, rescaling the sequence $\{\hat{h}_m\}$ by the factor $\left(\frac{m}{n}\right)^{\frac{2}{2\alpha+\gamma}}$ will produce a sequence of bandwidths $\{\hat{h}_n\}$ with

$$\hat{h}_n \equiv \hat{h}_m \left(\frac{m}{n}\right)^{\frac{2}{2\alpha+\gamma}} \tag{31}$$

that is in turn asymptotically equivalent to the sequence of nonstochastic minimizers of $MSE\left[\hat{\theta}_n(h)\right]$. The quantity given in (31) can be thought of as an m-bootstrap estimate of the MSE-optimal bandwidth for the estimator computed using the full sample of observations. From (30) and (31) it is easily seen that the discrepancy between the m-bootstrap bandwidth estimate $\hat{h}_n$ and the true MSE-optimal bandwidth $h_{n,opt}$ vanishes at rate

$$\hat{h}_n - h_{n,opt} = o_p\left(n^{-\frac{2}{2\alpha+\gamma}}\right),$$

which indicates performance in large samples at least as accurate as the plug-in bandwidth estimator proposed by Powell and Stoker (1996, §4.4).

3.3 Selection of the resample size

In this section guidance on selecting the resample size m is given. In particular, inspection of the proof of Theorem 2 reveals that the discrepancy between the m-bootstrap MSE estimate $MSE\left[\hat{\theta}^*_m(h_m)\middle| \mathcal{X}_n\right]$ and the leading terms of the expansion of $MSE\left[\hat{\theta}_m(h_m)\right]$ for $h_m \propto m^{-\frac{2}{2\alpha+\gamma}}$ with $m \propto n^\delta$ is of order

$$o_p\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right) + o_p\left(n^{-\delta} m^{-\frac{2\alpha}{2\alpha+\gamma}}\right) + o_p\left(n^{-2\delta} m^{\frac{\gamma}{2\alpha+\gamma}}\right). \tag{32}$$

The expression in (32) is clearly minimized for δ satisfying the requirements of Theorem 2 by setting $\delta = \frac{1}{2}$.[15]

[15]In particular, cf. (47) and (64) below.

4 Proof of Theorem 2

This section presents the proof of the central result of this paper.

Let $\mathcal{H}_m(\alpha, \gamma)$ denote the analogue of the set $\mathcal{H}_n(\alpha, \gamma)$ as defined above in Section 3.1. In particular, $\mathcal{H}_m(\alpha, \gamma)$ denotes the set of all bandwidth sequences appropriate for samples of size m that satisfy conditions (13) and (14) with m appearing in place of n. The following preliminary argument is made.

Lemma 2. *Suppose Assumption 3 holds, and suppose that $h_m \in \mathcal{H}_m(\alpha, \gamma)$. Then*

$$\hat{\theta}^*_m(h_m) - \bar{\theta}^*_m(h_m) = o_p\left(m^{-1} h_m^{-\frac{\gamma}{2}}\right).$$

Proof. The proof appears in Appendix A.1. □

From Lemma 2, it follows that we can write

$$\begin{aligned} \hat{\theta}^*_m(h_m) - \hat{\theta}_n(h_n) &= \bar{\theta}^*_m(h_m) - \hat{\theta}_n(h_n) + o_p\left(m^{-1} h_m^{-\frac{\gamma}{2}}\right) \\ &= \bar{\theta}^*_m(h_m) - \theta^*(h_m) + \theta^*(h_m) - \hat{\theta}_n(h_n) + o_p\left(m^{-1} h_m^{-\frac{\gamma}{2}}\right) \\ &= \frac{2}{m} \sum_{i=1}^{m} \left(\bar{g}^*(X_i^*, h_m) - \theta^*(h_m)\right) + \theta^*(h_m) - \hat{\theta}_n(h_n) + o_p\left(m^{-1} h_m^{-\frac{\gamma}{2}}\right). \end{aligned}$$

Consider $\theta^*(h_m) - \hat{\theta}_n(h_n)$. Recalling that $\theta(h_n) \equiv E\left[\hat{\theta}_n(h_n)\right]$, set

$$\begin{aligned} \theta^*(h_m) - \hat{\theta}_n(h_n) &= \theta^*(h_m) - \theta(h_n) + \theta(h_n) - \hat{\theta}_n(h_n) \\ &= \{\theta^*(h_m) - \theta(h_m)\} + \{\theta(h_m) - \theta_0\} + \{\theta_0 - \theta(h_n)\} \\ &\quad + \left\{\theta(h_n) - \hat{\theta}_n(h_n)\right\}. \end{aligned} \tag{33}$$

The analysis of the quantity in (33) begins by considering $\theta^*(h_m) - \theta(h_m)$ for $h_m \in \mathcal{H}_m(\alpha, \gamma)$. Since

$$\theta^*(h_m) = \frac{1}{n^2} \sum_{i,j} g(X_i, X_j, h_m),$$

it follows that

$$\begin{aligned} E\left[\theta^*(h_m)\right] &= E\left[g(X_1, X_2, h_m)\right] \\ &= \theta(h_m). \end{aligned}$$

In addition,

$$E\left[(\theta^*(h_m)-\theta(h_m))^2\right]=\frac{1}{n^4}\sum_{i,j,k,l}E\left[(g(X_i,X_j,h_m)-\theta(h_m))\,(g(X_k,X_l,h_m)-\theta(h_m))\right]. \tag{34}$$

By independence, all terms in (34) with $i\neq k$ and $j\neq l$ have zero expectation. Suppose that $i=k$ and that $j\neq l$. In this case,

$$\begin{aligned}
& E\left[(g(X_i,X_j,h_m)-\theta(h_m))\,(g(X_i,X_l,h_m)-\theta(h_m))\right]\\
=\; & E\left[E\left[(g(X_i,X_j,h_m)-\theta(h_m))\,(g(X_i,X_l,h_m)-\theta(h_m))\right| X_i]\right]\\
=\; & E\left[(\bar{g}(X_i,h_m)-\theta(h_m))\,(\bar{g}(X_i,h_m)-\theta(h_m))\right]\\
=\; & Var\left[\bar{g}(X_i,h_m)\right]\\
=\; & Var\left[\bar{g}_0(X_1)\right]+2Cov\left[\bar{g}_0(X_1),s(X_1)\right]h_m^{\alpha}+o(h_m^{\alpha})
\end{aligned}$$

by Assumption 2. Now suppose that $i=k$ and $j=l$. Then

$$\begin{aligned}
& E\left[(g(X_i,X_j,h_m)-\theta(h_m))^2\right]\\
=\; & E\left[g^2(X_i,X_j,h_m)-2\theta(h_m)g(X_i,X_j,h_m)+\theta^2(h_m)\right]\\
=\; & E\left[q(X_1)\right]h_m^{-\gamma}+o(h_m^{-\gamma})-\theta^2(h_m)
\end{aligned}$$

by Assumption 3. Therefore for $h_m\in\mathcal{H}_m(\alpha,\gamma)$ and recalling the assumption that $m=O\left(n^{1-\delta}\right)$ for some $\delta\in(0,1)$ as $m,n\to\infty$,

$$\begin{aligned}
& E\left[(\theta^*(h_m)-\theta(h_m))^2\right]\\
=\; & \frac{1}{n^4}\left\{n^2\left(E\left[q(X_1)\right]h_m^{-\gamma}+o(h_m^{-\gamma})-\theta^2(h_m)\right)+2n^2(n-1)\left(Var\left[\bar{g}_0(X_1)\right]\right.\right.\\
& \left.\left.+2Cov\left[\bar{g}_0(X_1),s(X_1)\right]h_m^{\alpha}+o(h_m^{\alpha})\right)\right\}\\
=\; & n^{-2}\left\{E\left[q(X_1)\right]h_m^{-\gamma}-\theta^2(h_m)\right\}+o(n^{-2}h_m^{-\gamma})\\
& +\frac{2(n-1)}{n^2}Var\left[\bar{g}_0(X_1)\right]+\frac{2(n-1)}{n^2}Cov\left[\bar{g}_0(X_1),s(X_1)\right]h_m^{\alpha}+o(n^{-1}h_m^{\alpha})\\
=\; & O\left(n^{-2}h_m^{-\gamma}\right)+O\left(n^{-1}\right)+O\left(n^{-1}h_m^{\alpha}\right) \qquad (35)\\
=\; & o\left(\frac{m}{n^2}\right)+O\left(\frac{1}{n}\right)+o\left(\frac{1}{n\sqrt{m}}\right)\\
=\; & O\left(\frac{1}{n}\right).
\end{aligned}$$

Therefore an application of Chebyshev's inequality produces the result

$$\theta^*(h_m)-\theta(h_m)=o_p\left(\frac{1}{\sqrt{n}}\right)$$

for all $h_m \in \mathcal{H}_m(\alpha,\gamma)$. Then for $h_m \in \mathcal{H}_m(\alpha,\gamma)$ and $h_n \in \mathcal{H}_n(\alpha,\gamma)$,

$$\begin{aligned}\theta^*(h_m) - \hat{\theta}_n(h_n) &= o_p\left(\frac{1}{\sqrt{n}}\right) + \{\theta(h_m) - \theta_0\} + E\left[s(X_1)\right] h_n^\alpha + o\left(h_n^\alpha\right) + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &= \theta(h_m) - \theta_0 + E\left[s(X_1)\right] h_n^\alpha + o\left(h_n^\alpha\right) + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &= \theta(h_m) - \theta_0 + o_p\left(\frac{1}{\sqrt{n}}\right)\end{aligned}$$

as $m, n \to \infty$ with $m = O\left(n^{1-\delta}\right)$ for some $\delta \in (0,1)$.

Now let

$$\begin{aligned}\sigma_m^{*2} &\equiv \sigma_m^{*2}(h_m) \\ &\equiv 4Var\left[\bar{g}^*(X_i^*, h_m)|\mathcal{X}_n\right]. \end{aligned} \tag{36}$$

Note that

$$\begin{aligned}\sigma_m^{*2} &= 4E\left[\left(\bar{g}^*(X_i^*, h_m) - \theta^*(h_m)\right)^2 \middle| \mathcal{X}_n\right] \\ &= 4\left\{E\left[\bar{g}^{*2}(X_i^*, h_m) \middle| \mathcal{X}_n\right] - \theta^{*2}(h_m)\right\}.\end{aligned}$$

Define $Y_{im}^* \equiv Y_{im}^*(h_m) \equiv \frac{2}{\sigma_m^*}\left(\bar{g}^*(X_i^*, h_m) - \theta^*(h_m)\right)$. It will be shown that $\frac{1}{\sqrt{m}}\sum_{i=1}^m Y_{im}^*$ is asymptotically normal under the assumptions that have been made regarding m and n. Two preliminary arguments are made prior to the statement of the asymptotic normality result.

Lemma 3. *Suppose Assumptions 1–5 hold, and that $h_m \in \mathcal{H}_m(\alpha,\gamma)$. Then*

$$|g(X_i, X_j, h_m) - \theta^*(h_m)| = o_p\left(n^{-\frac{\delta}{2}}\right)$$

as $m, n \to \infty$ with $m \propto n^\delta$ for some constant $\delta \in (0,1)$ with $\delta \leq 1-\delta$.

Proof. The proof appears in Appendix A.2. □

Lemma 4. *Under the conditions of Lemma 3,*

$$\begin{aligned}E\left[\sigma_m^{*2}\right] &= 4Var\left[\bar{g}_0(X_1)\right] + 8Cov\left[\bar{g}_0(X_1), s(X_1)\right] h_m^\alpha + o\left(h_m^\alpha\right) \\ &\quad + \frac{2}{m}E\left[g^2(X_1, X_2, h_m)\right] + o(1).\end{aligned}$$

Proof. The proof appears in Appendix A.3. □

The asymptotic normality of $\frac{1}{\sqrt{m}}\sum_{i=1}^m Y_{im}^*$ is given in the following argument.

Proposition 1. *Under the conditions of Lemmas 3 and 4,*

$$\frac{1}{\sqrt{m}}\sum_{i=1}^m Y_{im}^* \xrightarrow{d} W,$$

where W is a standard normal random variable independent of $\mathcal{X}_n$.

Proof. Consider $E\left[|Y_{im}^*|^3\middle|\mathcal{X}_n\right]$. We have

$$\begin{aligned}
& E\left[|Y_{im}^*|^3\middle|\mathcal{X}_n\right] \\
= \quad & \frac{8}{n}\sum_{i=1}^n \left|\frac{\bar{g}^*(X_i,h_m)-\theta^*(h_m)}{\sigma_m^*}\right|^3 \\
\leq \quad & \frac{8}{n^4}\sum_{i=1}^n\sum_{j,k,l}\left|\frac{g(X_i,X_j,h_m)-\theta^*(h_m)}{\sigma_m^*}\right|\left|\frac{g(X_i,X_k,h_m)-\theta^*(h_m)}{\sigma_m^*}\right| \\
& \cdot\left|\frac{g(X_i,X_l,h_m)-\theta^*(h_m)}{\sigma_m^*}\right|.
\end{aligned}$$

Since $\sigma_m^* = \left(E\left[\sigma_m^{*2}\right]\right)^{\frac{1}{2}} + \frac{1}{2}\left(\sigma_m^{*2}-E\left[\sigma_m^{*2}\right]\right)\left(E\left[\sigma_m^{*2}\right]\right)^{-\frac{1}{2}}+\cdots$, we have

$$\begin{aligned}
& \left|\frac{g(X_i,X_j,h_m)-\theta^*(h_m)}{\sigma_m^*}\right| \\
\leq \quad & \frac{|g(X_i,X_j,h_m)-\theta^*(h_m)|}{\sqrt{E\left[\sigma_m^{*2}\right]}} \qquad (37) \\
\leq \quad & \frac{|g(X_i,X_j,h_m)-\theta^*(h_m)|}{2\sqrt{Var\left[\bar{g}_0(X_1)\right]}} \qquad (38) \\
= \quad & o_p\left(n^{-\frac{\delta}{2}}\right), \qquad (39)
\end{aligned}$$

where the inequality (37)–(38) follows from Lemma 4 and the inequality (38)–(39) follows from Lemma 3. Therefore

$$E\left[|Y_{im}^*|^3\middle|\mathcal{X}_n\right] = n^{-4}\cdot n^4\cdot o_p\left(n^{-\frac{3\delta}{2}}\right) = o_p\left(n^{-\frac{3\delta}{2}}\right),$$

and as such,

$$mE\left[\left|\frac{1}{\sqrt{m}}Y_{im}^*\right|^3\middle|\mathcal{X}_n\right]$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{m}} E\left[|Y_{im}^*|^3 \middle| \mathcal{X}_n \right] \\
&= m^{-\frac{1}{2}} \cdot o_p\left(n^{-\frac{3\delta}{2}}\right) \\
&= O_p\left(n^{-\frac{\delta}{2}}\right) o_p\left(n^{-\frac{3\delta}{2}}\right) \\
&= o_p\left(n^{-2\delta}\right) \\
&= o_p(1)
\end{aligned}
$$

as $n \to \infty$. This verifies the condition for Liapunov's central limit theorem.[16] Therefore

$$
\left. \frac{1}{\sqrt{m}} \sum_{i=1}^{m} Y_{im}^* \right| \mathcal{X}_n \xrightarrow{d} W,
$$

where $W \sim N(0,1)$ and W is independent of $\mathcal{X}_n$. □

Corollary 1. *Under the conditions of Proposition 1, there exists a positive constant C_3 such that*

$$
\sup_y \left| P\left[\frac{1}{\sqrt{m}} \sum_{i=1}^{m} Y_{im}^* \leq y \middle| \mathcal{X}_n \right] - \Phi(y) \right| \leq C_3 \cdot \frac{1}{\sqrt{m}} E\left[|Y_{im}^*|^3 \middle| \mathcal{X}_n \right],
$$

where $\Phi(\cdot)$ denotes the distribution function of W as given above.

Proof. This is an application of the Berry-Esséen theorem for triangular arrays.[17] □

Now consider the following, under Assumptions 1–5, $h_n \in \mathcal{H}_n(\alpha, \gamma)$, $h_m \in \mathcal{H}_m(\alpha, \gamma)$, $m, n \to \infty$ with $m \propto n^\delta$ for some constant $\delta \in (0,1)$ with $\delta \leq 1 - \delta$:

$$
\frac{\sqrt{m}}{\sigma_m^*}\left(\hat{\theta}_m^*(h_m) - \hat{\theta}_n(h_n)\right) = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} Y_{im}^* + \frac{\sqrt{m}}{\sigma_m^*}\left(b_m^* + c_m^*\right),
$$

where

$$
b_m^* \equiv \frac{\sqrt{m}}{\sigma_m^*}\left(\theta(h_m) - \theta_0\right) \tag{40}
$$

[16] Cf. e.g., Chung (2001, Theorem 7.1.2).
[17] Cf. e.g., Chung (2001, Theorem 7.4.1).

and

$$
\begin{aligned}
c_m^* \quad \equiv \quad & \frac{\sqrt{m}}{\sigma_m^*} \Big\{ (\theta^*(h_m) - \theta(h_m)) + (\theta_0 - \theta(h_n)) + \Big(\theta(h_n) - \hat{\theta}_n(h_n) \Big) \\
& + \Big(\hat{\theta}_m^*(h_m) - \bar{\theta}_m^*(h_m) \Big) \Big\} . \qquad (41)
\end{aligned}
$$

For $\Phi(\cdot)$ denoting the distribution function of a $N(0,1)$ random variate, set

$$
\Delta^* \equiv \sup_y \left| P \left[\frac{1}{\sqrt{m}} \sum_{i=1}^m Y_{im}^* \leq y \middle| \mathcal{X}_n \right] - \Phi(y) \right| .
$$

Now suppose that

$$
C_3 \cdot \frac{1}{\sqrt{m}} E \left[|Y_{im}^*|^3 \middle| \mathcal{X}_n \right] \leq e^{-\frac{1}{2}}, \qquad (42)
$$

where $C_3 > 0$ is the constant appearing in Corollary 1. Condition (42) clearly holds for n sufficiently large. From Petrov (1975, Theorem 9, p. 121), there exists a constant $C_4 > 0$ such that for every y,

$$
\left| P \left[\frac{1}{\sqrt{m}} \sum_{i=1}^m Y_{im}^* \leq y \middle| \mathcal{X}_n \right] - \Phi(y) \right| \leq \frac{C_4 \Delta^* \log\left(\frac{1}{\Delta^*}\right) + \lambda_2^*}{1 + y^4},
$$

where

$$
\begin{aligned}
\lambda_2^* \quad \equiv \quad & \left| E \left[\left(\frac{1}{\sqrt{m}} \sum_{i=1}^m Y_{im}^* \right)^4 \middle| \mathcal{X}_n \right] - \int_{-\infty}^{\infty} y^4 d\Phi(y) \right| \\
= \quad & \left| E \left[\left(\frac{1}{\sqrt{m}} \sum_{i=1}^m Y_{im}^* \right)^4 \middle| \mathcal{X}_n \right] - 3 \right| \\
= \quad & \left| \frac{1}{m^2} \sum_{i,j,k,l} E \left[Y_{im}^* Y_{jm}^* Y_{km}^* Y_{lm}^* \middle| \mathcal{X}_n \right] - 3 \right| \\
= \quad & \left| \frac{1}{m} E \left[Y_{im}^{*4} \middle| \mathcal{X}_n \right] + \frac{2m(m-1)}{m^2} \left(E \left[Y_{im}^{*2} \middle| \mathcal{X}_n \right] \right)^2 - 3 \right| \\
= \quad & \left| \frac{1}{m} E \left[Y_{im}^{*4} \middle| \mathcal{X}_n \right] - \frac{2}{m} - 1 \right| .
\end{aligned}
$$

From Lemma 3, $|g(X_i, X_j, h_m) - \theta^*(h_m)| = o_p\left(n^{-\frac{\delta}{2}}\right)$, and

$$\begin{aligned}
& E\left[Y_{im}^{*4}|\mathcal{X}_n\right] \\
\leq\; & \frac{16}{n^5}\sum_{i=1}^{n}\sum_{p,q,r,s}\left|\frac{g(X_i,X_p,h_m)-\theta^*(h_m)}{\sigma_m^*}\right|\left|\frac{g(X_i,X_q,h_m)-\theta^*(h_m)}{\sigma_m^*}\right| \\
& \cdot\left|\frac{g(X_i,X_r,h_m)-\theta^*(h_m)}{\sigma_m^*}\right|\left|\frac{g(X_i,X_s,h_m)-\theta^*(h_m)}{\sigma_m^*}\right| \\
=\; & \frac{16}{n^5}\cdot n\cdot n^4\cdot O_p\left(\left|\frac{g(X_i,X_p,h_m)-\theta^*(h_m)}{\sigma_m^*}\right|^4\right) \qquad (43) \\
=\; & o_p\left(n^{-2\delta}\right), \qquad (44)
\end{aligned}$$

where the equality (43)–(44) follows from (39) above. Therefore

$$\begin{aligned}
\lambda_2^* \;&\leq\; \frac{1}{m}\cdot o_p\left(n^{-2\delta}\right)+\frac{2}{m}+1 \\
&\leq\; O_p\left(n^{-\delta}\right)+1.
\end{aligned}$$

So for $W \sim N(0,1)$ independently distributed of $\mathcal{X}_n$ and constants $C_5, C_6 > 0$,

$$\begin{aligned}
& \left|E\left[\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m}Y_{im}^*+b_m^*+c_m^*\right)^2\middle|\mathcal{X}_n\right]-E\left[(W+b_m^*+c_m^*)^2\middle|\,\mathcal{X}_n\right]\right| \\
=\;& 2\left|\int_0^\infty y\left\{P\left[\left|\frac{1}{\sqrt{m}}\sum_{i=1}^{m}Y_{im}^*+b_m^*+c_m^*\right|>y\middle|\mathcal{X}_n\right]\right.\right. \\
& \left.\left. -P\left[|W+b_m^*+c_m^*|>y|\,\mathcal{X}_n\right]\right\}dy\right| \\
=\;& 2\left|\int_0^\infty y\left\{P\left[|W+b_m^*+c_m^*|\leq y|\,\mathcal{X}_n\right]\right.\right. \\
& \left.\left. -P\left[\left|\frac{1}{\sqrt{m}}\sum_{i=1}^{m}Y_{im}^*+b_m^*+c_m^*\right|\leq y\middle|\mathcal{X}_n\right]\right\}dy\right| \\
\leq\;& C_5\int_{-\infty}^{\infty}|y|\left\{1+(y+b_m^*+c_m^*)^4\right\}^{-1}dy\cdot\lambda_2^* \\
\leq\;& C_6\left(1+|b_m^*+c_m^*|\right)\left(O_p\left(n^{-\delta}\right)+1\right).
\end{aligned}$$

Therefore

$$MSE\left[\hat{\theta}_m^*(h_m)\middle|\,\mathcal{X}_n\right]$$

$$
\begin{aligned}
&= E\left[\left(\hat{\theta}_m^*(h_m) - \hat{\theta}_n(h_n)\right)^2 \middle| \mathcal{X}_n\right] \\
&= \frac{\sigma_m^{*2}}{m} E\left[\left(\frac{\sqrt{m}}{\sigma_m^*}\left(\hat{\theta}_m^*(h_m) - \hat{\theta}_n(h_n)\right)\right)^2 \middle| \mathcal{X}_n\right] \\
&= \frac{\sigma_m^{*2}}{m} E\left[\left(\frac{1}{\sqrt{m}}\sum_{i=1}^{m} Y_{im}^* + b_m^* + c_m^*\right)^2 \middle| \mathcal{X}_n\right] \\
&= \frac{\sigma_m^{*2}}{m} E\left[(W + b_m^* + c_m^*)^2 \middle| \mathcal{X}_n\right] + R_m^*(h_m),
\end{aligned}
$$

where

$$
\begin{aligned}
|R_m^*(h_m)| &\leq \frac{\sigma_m^{*2}}{m} \cdot C_6 \left(1 + |b_m^* + c_m^*|\right)\left(O_p\left(n^{-\delta}\right) + 1\right) \\
&= O_p\left(\frac{\sigma_m^{*2}}{m}\left(1 + |b_m^* + c_m^*|\right)\right).
\end{aligned}
$$

From Lemma 3 we have that

$$
\begin{aligned}
\sigma_m^{*2} &\leq \frac{4}{n^3}\sum_{i=1}^{n}\sum_{j,k} |g(X_i, X_j, h_m) - \theta^*(h_m)|\,|g(X_i, X_k, h_m) - \theta^*(h_m)| \\
&= O\left(n^{-3} \cdot n \cdot n^2\right) o_p\left(n^{-\delta}\right) \\
&= o_p\left(n^{-\delta}\right).
\end{aligned}
$$

Therefore

$$
\frac{\sigma_m^{*2}}{m} = m^{-1} \cdot o_p\left(n^{-\delta}\right) = o_p\left(n^{-2\delta}\right), \tag{45}
$$

and

$$
\begin{aligned}
\frac{\sigma_m^{*2}}{m}|b_m^* + c_m^*| &\leq \frac{\sigma_m^*}{\sqrt{m}}|\theta(h_m) - \theta_0| + \frac{\sigma_m^*}{\sqrt{m}}\left|o_p\left(\frac{1}{\sqrt{n}}\right) + o_p\left(m^{-1}h^{-\frac{\gamma}{2}}\right)\right| \\
&= o_p\left(n^{-\delta}h_m^{\alpha}\right) + o_p\left(n^{-\delta-\frac{1}{2}}\right) + o_p\left(n^{-\delta}m^{-1}h_m^{-\frac{\gamma}{2}}\right) \\
&= o_p\left(n^{-\delta}h_m^{\alpha}\right) + o_p\left(n^{-\delta-\frac{1}{2}}\right) + o_p\left(n^{-2\delta}h_m^{-\frac{\gamma}{2}}\right) \\
&= o_p\left(n^{-\delta}h_m^{\alpha}\right) + o_p\left(n^{-2\delta}h_m^{-\frac{\gamma}{2}}\right),
\end{aligned} \tag{46}
$$

for $h_m \in \mathcal{H}_m(\alpha,\gamma)$. Combining (45) and (46) produces

$$\begin{aligned} |R_m^*(h_m)| &= o_p\left(n^{-2\delta}\right) + o_p\left(n^{-\delta}h_m^{\alpha}\right) + o_p\left(n^{-2\delta}h_m^{-\frac{\gamma}{2}}\right) \\ &= o_p\left(n^{-\delta}h_m^{\alpha}\right) + o_p\left(n^{-2\delta}h_m^{-\frac{\gamma}{2}}\right). \end{aligned} \tag{47}$$

Now consider the leading term in the expansion of $MSE\left[\hat{\theta}_m^*(h_m)\middle|\, \mathcal{X}_n\right]$. In particular, consider $\frac{\sigma_m^{*2}}{m}E\left[(W+b_m^*+c_m^*)^2\middle|\,\mathcal{X}_n\right]$, where $W \sim N(0,1)$ is independently distributed of $\mathcal{X}_n$ and b_m^* and c_m^* are as defined above in (40) and (41), respectively.

We have

$$\begin{aligned} & E\left[\frac{\sigma_m^{*2}}{m}E\left[(W+b_m^*+c_m^*)^2\middle|\,\mathcal{X}_n\right]\right] \\ =\ & E\left[\frac{\sigma_m^{*2}}{m}E\left[W^2+2W\left(b_m^*+c_m^*\right)+\left(b_m^*+c_m^*\right)^2\middle|\,\mathcal{X}_n\right]\right] \\ =\ & E\left[\frac{\sigma_m^{*2}}{m}\left(1+E\left[(b_m^*+c_m^*)^2\middle|\,\mathcal{X}_n\right]\right)\right] \\ =\ & \frac{E\left[\sigma_m^{*2}\right]}{m} + E\left[(\theta(h_m)-\theta_0)^2 + 2\left(\theta(h_m)-\theta_0\right)E\left[\frac{\sigma_m^*}{\sqrt{m}}c_m^*\middle|\,\mathcal{X}_n\right] + E\left[\frac{\sigma_m^{*2}}{m}c_m^{*2}\middle|\,\mathcal{X}_n\right]\right] \\ =\ & \frac{E\left[\sigma_m^{*2}\right]}{m} + (\theta(h_m)-\theta_0)^2 + 2\left(\theta(h_m)-\theta_0\right)E\left[\frac{\sigma_m^*}{\sqrt{m}}c_m^*\right] + E\left[\frac{\sigma_m^{*2}}{m}c_m^{*2}\right]. \end{aligned}$$

Note that for $h_m \in \mathcal{H}_m(\alpha,\gamma)$ and $h_n \in \mathcal{H}_n(\alpha,\gamma)$,

$$\begin{aligned} E\left[\frac{\sigma_m^*}{\sqrt{m}}c_m^*\right] &= \left\{\theta(h_m)-\theta(h_m)\right\} + \left\{-E\left[s(X_1)\right]h_n^{\alpha} + o\left(h_n^{\alpha}\right)\right\} + \left\{\theta(h_n)-\theta(h_n)\right\} \\ &\quad + \left\{\theta(h_m)-\theta(h_m)\right\} \\ &= O\left(h_n^{\alpha}\right) \\ &= o\left(h_m^{\alpha}\right) \end{aligned}$$

from the assumption that $m \propto n^{\delta}$ for some $\delta \in (0,1)$ with $\delta \leq 1-\delta$. In addition,

$$\begin{aligned} & E\left[\frac{\sigma_m^{*2}}{m}c_m^{*2}\right] \\ =\ & O\left(E\left[(\theta^*(h_m)-\theta(h_m))^2\right] + E\left[(\theta^*(h_m)-\theta(h_m))\left(\theta_0-\theta(h_n)\right)\right]\right. \end{aligned}$$

$$
\begin{aligned}
&+E\left[\left(\theta^*(h_m)-\theta(h_m)\right)\left(\theta(h_n)-\hat{\theta}_n(h_n)\right)\right] \\
&+E\left[\left(\theta^*(h_m)-\theta(h_m)\right)\left(\hat{\theta}^*_m(h_m)-\bar{\theta}^*_m(h_m)\right)\right]+\left(\theta_0-\theta(h_n)\right)^2 \\
&+\left(\theta_0-\theta(h_n)\right)E\left[\theta(h_n)-\hat{\theta}_n(h_n)\right]+\left(\theta_0-\theta(h_n)\right)E\left[\hat{\theta}^*_m(h_m)-\bar{\theta}^*_m(h_m)\right] \\
&+E\left[\left(\theta(h_n)-\hat{\theta}_n(h_n)\right)^2\right]+E\left[\left(\theta(h_n)-\hat{\theta}_n(h_n)\right)\left(\hat{\theta}^*_m(h_m)-\bar{\theta}^*_m(h_m)\right)\right] \\
&+E\left[\left(\hat{\theta}^*_m(h_m)-\bar{\theta}^*_m(h_m)\right)^2\right]\Bigg) \\
=\;& o\left(m^{-2}h_m^{-\gamma}\right)+o\left(h_m^{2\alpha}\right)+o\left(m^{-2}h_m^{-\gamma}\right)+o\left(m^{-2}h_m^{-\gamma}\right) \\
=\;& o\left(h_m^{2\alpha}\right)+o\left(m^{-2}h_m^{-\gamma}\right),
\end{aligned}
$$

where, after noting that $m \propto n^{\delta}$ for some constant $\delta \in (0,1)$ with $\delta \leq 1-\delta$, appeal was made to the following results established above:

$$
E\left[\left(\theta^*(h_m)-\theta(h_m)\right)^2\right]=O\left(n^{-1}\right)=o\left(m^{-1}\right)=h_m^{-\gamma}o\left(m^{-2}\right)
$$

from (35);

$$
\left(\theta_0-\theta(h_n)\right)^2=O\left(h_n^{2\alpha}\right)=o\left(h_m^{2\alpha}\right)
$$

by Assumption 2;

$$
E\left[\left(\theta(h_n)-\hat{\theta}_n(h_n)\right)^2\right]=Var\left[\hat{\theta}_n(h_n)\right]=O\left(n^{-2}h_n^{-\gamma}\right)=o\left(m^{-2}h_m^{-\gamma}\right)
$$

from (16)–(17); and

$$
E\left[\left(\hat{\theta}^*_m(h_m)-\bar{\theta}^*_m(h_m)\right)^2\right]=o\left(m^{-2}h_m^{-\gamma}\right)
$$

by inspection of the proof of Lemma 2.

From Lemma 4 we have the representation

$$
E\left[\sigma_m^{*2}\right]=4Var\left[\bar{g}_0(X_1)\right]+8Cov\left[\bar{g}_0(X_1),s(X_1)\right]h_m^{\alpha}+o\left(h_m^{\alpha}\right)+\frac{2}{m}E\left[g^2(X_1,X_2,h_m)\right]+o(1).
$$

Note that for $h_m \in \mathcal{H}_m(\alpha,\gamma)$, $o\left(m^{-1}\right)=h_m^{-\gamma}o\left(m^{-2}\right)$, and so the leading term of the expansion of $MSE\left[\hat{\theta}^*_m(h_m)\middle|\mathcal{X}_n\right]$ has mean

$$
\frac{4}{m}Var\left[\bar{g}_0(X_1)\right]+\frac{2}{m^2}E\left[g^2(X_1,X_2,h_m)\right]+O\left(m^{-1}h_m^{\alpha}\right)+o\left(m^{-2}h_m^{-\gamma}\right)
$$

$$
\begin{aligned}
& + (\theta(h_m) - \theta_0)^2 + o\left(h_m^{2\alpha}\right) + o\left(h_m^{2\alpha}\right) + o\left(m^{-2} h_m^{-\gamma}\right) \\
= \;& \frac{4}{m} Var\left[\bar{g}_0(X_1)\right] + (\theta(h_m) - \theta_0)^2 + \frac{2}{m^2} E\left[g^2(X_1, X_2, h_m)\right] + O\left(m^{-1} h_m^{\alpha}\right) + o\left(h_m^{2\alpha}\right) \\
& + o\left(m^{-2} h_m^{-\gamma}\right). \qquad (48)
\end{aligned}
$$

It remains to complete the argument—via an appeal to Chebyshev's inequality—by first considering the order of

$$
E\left[\frac{\sigma^{*4}}{m^2}\left\{E\left[\left(W + b_m^* + c_m^*\right)^2 \middle| \mathcal{X}_n\right]\right\}^2\right],
$$

where W, b_m^* and c_m^* are as defined previously.

In particular,

$$
\begin{aligned}
& E\left[\frac{\sigma_m^{*4}}{m^2}\left\{E\left[\left(W + b_m^* + c_m^*\right)^2 \middle| \mathcal{X}_n\right]\right\}^2\right] \\
= \;& E\left[\frac{\sigma_m^{*4}}{m^2}\left\{W^2 + E\left[\left(b_m^* + c_m^*\right)^2 \middle| \mathcal{X}_n\right] + 2W E\left[\left(b_m^* + c_m^*\right)^2 \middle| \mathcal{X}_n\right]\right\}^2\right] \\
= \;& E\left[\frac{\sigma_m^{*4}}{m^2}\left\{\left(W^2 + E\left[\left(b_m^* + c_m^*\right)^2 \middle| \mathcal{X}_n\right]\right)^2\right.\right. \\
& \left.\left. + 4W E\left[b_m^* + c_m^* \middle| \mathcal{X}_n\right]\left(W^2 + E\left[\left(b_m^* + c_m^*\right)^2 \middle| \mathcal{X}_n\right]\right) + 4W^2\left(E\left[b_m^* + c_m^* | \mathcal{X}_n\right]\right)^2\right\}\right] \\
= \;& E\left[\frac{\sigma_m^{*4}}{m^2}\left(W^2 + E\left[\left(b_m^* + c_m^*\right)^2 | \mathcal{X}_n\right]\right)^2\right] + 4E\left[\frac{\sigma^{*4}}{m^2}\left(E\left[b_m^* + c_m^* | \mathcal{X}_n\right]\right)^2\right] \\
= \;& E\left[\frac{\sigma_m^{*4}}{m^2}\left(W^4 + 2W^2 E\left[\left(b_m^* + c_m^*\right)^2 \middle| \mathcal{X}_n\right] + \left(E\left[\left(b_m^* + c_m^*\right)^2 \middle| \mathcal{X}_n\right]\right)^2\right)\right] \\
& + 4E\left[\frac{\sigma_m^{*4}}{m^2}\left(E\left[b_m^* + c_m^* | \mathcal{X}_n\right]\right)^2\right] \\
= \;& 3E\left[\frac{\sigma_m^{*4}}{m^2}\right] + 2E\left[\frac{\sigma_m^{*4}}{m^2} E\left[\left(b_m^* + c_m^*\right)^2 | \mathcal{X}_n\right]\right] + 5E\left[\frac{\sigma_m^{*4}}{m^2}\left(E\left[\left(b_m^* + c_m^*\right)^2 | \mathcal{X}_n\right]\right)^2\right] \\
= \;& 3E\left[\frac{\sigma_m^{*4}}{m^2}\right] + 2E\left[\frac{\sigma_m^{*2}}{m} E\left[(\theta(h_m) - \theta_0)^2 + 2(\theta(h_m) - \theta_0)\frac{\sigma_m^*}{\sqrt{m}} c_m^* + \frac{\sigma_m^{*2}}{m} c_m^{*2} \middle| \mathcal{X}_n\right]\right] \\
& + 5E\left[\left(E\left[(\theta(h_m) - \theta_0)^2 + 2(\theta(h_m) - \theta_0)\frac{\sigma_m^*}{\sqrt{m}} c_m^* + \frac{\sigma_m^{*2}}{m} c_m^{*2} \middle| \mathcal{X}_n\right]\right)^2\right] \\
= \;& 3E\left[\frac{\sigma_m^{*4}}{m^2}\right] + 2\left\{(\theta(h_m) - \theta_0)^2 E\left[\frac{\sigma_m^{*2}}{m}\right] + 2(\theta(h_m) - \theta_0) E\left[\frac{\sigma_m^{*2}}{m} \cdot \frac{\sigma_m^*}{\sqrt{m}} c_m^*\right]\right.
\end{aligned}
$$

$$
\begin{aligned}
& +E\left[\frac{\sigma_m^{*2}}{m}\cdot\frac{\sigma_m^{*2}}{m}c_m^{*2}\right]\Bigg\} + 5E\left[(\theta(h_m)-\theta_0)^4 + 4\left(\theta(h_m)-\theta_0\right)^2\frac{\sigma_m^{*2}}{m}c_m^{*2} + \frac{\sigma_m^{*4}}{m^2}c_m^{*4}\right. \\
& \left. +4\left(\theta(h_m)-\theta_0\right)^3\frac{\sigma_m^{*}}{\sqrt{m}}c_m^{*} + 2\left(\theta(h_m)-\theta_0\right)^2\frac{\sigma_m^{*2}}{m}c_m^{*2} + 4\left(\theta(h_m)-\theta_0\right)\frac{\sigma_m^{*}}{\sqrt{m}}c_m^{*}\cdot\frac{\sigma_m^{*2}}{m}c_m^{*2}\right] \\
\leq\ & 3E\left[\frac{\sigma_m^{*4}}{m^2}\right] + 2\left\{\left(\theta(h_m)-\theta_0\right)^2 E\left[\frac{\sigma_m^{*2}}{m}\right] + 2\left(\theta(h_m)-\theta_0\right)\left(E\left[\frac{\sigma_m^{*4}}{m^2}\right]\right)^{\frac{1}{2}}\left(E\left[\frac{\sigma_m^{*2}}{m}c_m^{*2}\right]\right)^{\frac{1}{2}}\right. \\
& \left. +\left(E\left[\frac{\sigma_m^{*4}}{m^2}\right]\right)^{\frac{1}{2}}\left(E\left[\frac{\sigma_m^{*4}}{m^2}c_m^{*4}\right]\right)^{\frac{1}{2}}\right\} + 5\left\{\left(\theta(h_m)-\theta_0\right)^4 + 4\left(\theta(h_m)-\theta_0\right)^2 E\left[\frac{\sigma_m^{*2}}{m}c_m^{*2}\right]\right. \\
& +E\left[\frac{\sigma_m^{*4}}{m^2}c_m^{*4}\right] + 4\left(\theta(h_m)-\theta_0\right)^3 E\left[\frac{\sigma_m^{*}}{\sqrt{m}}c_m^{*}\right] + 2\left(\theta(h_m)-\theta_0\right)^2 E\left[\frac{\sigma_m^{*2}}{m}c_m^{*2}\right] \\
& \left. +4\left(\theta(h_m)-\theta_0\right)\left(E\left[\frac{\sigma_m^{*2}}{m}c_m^{*2}\right]\right)^{\frac{1}{2}}\left(E\left[\frac{\sigma_m^{*4}}{m^2}c_m^{*4}\right]\right)^{\frac{1}{2}}\right\}, \qquad (49)
\end{aligned}
$$

where the inequality follows from several applications of Hölder's inequality.

The argument is made that each term in (49) is of order $O\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right)$ for $h_m \propto m^{-\frac{2}{2\alpha+\gamma}}$.

For $h_m \propto m^{-\frac{2}{2\alpha+\gamma}}$, Assumption 2 implies that $\theta(h_m)-\theta_0 = O\left(m^{-\frac{2\alpha}{2\alpha+\gamma}}\right)$. Arguments already made yield

$$
\begin{aligned}
E\left[\frac{\sigma_m^{*2}}{m}\right] &= O\left(m^{-2}h_m^{-\gamma}\right) = O\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right); & (50)\\
E\left[\frac{\sigma_m^{*}}{\sqrt{m}}c_m^{*}\right] &= o\left(h_m^{\alpha}\right) = o\left(m^{-\frac{2\alpha}{2\alpha+\gamma}}\right); & (51)\\
E\left[\frac{\sigma_m^{*2}}{m}c_m^{*2}\right] &= o\left(h_m^{2\alpha}\right) = o\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right). & (52)
\end{aligned}
$$

for $h_m \propto m^{-\frac{2}{2\alpha+\gamma}}$.

Now consider $E\left[\frac{\sigma_m^{*4}}{m^2}c_m^{*4}\right]$. In this connection the fourth moment conditions of Assumption 4 are crucial. Note that from the theory of U-statistics[18] and for $h_n \in \mathcal{H}_n(\alpha,\gamma)$,

$$
E\left[\left(\hat{\theta}_n(h_n)-\theta(h_n)\right)^4\right]
$$

[18]Cf. e.g., Serfling (1980, §5.2).

$$
\begin{aligned}
&= \binom{n}{2}^{-4} \Big\{ O\left(n^2(n-1)^2 \left(E\left[(g(X_1,X_2,h_n)-\theta(h_n))^2\right]\right)^2\right) \\
&\quad + O\left(n(n-1)E\left[(g(X_1,X_2,h_n)-\theta(h_n))^4\right]\right) \\
&\quad + O\left(n^2\left(E\left[(\bar{g}(X_1,h_n)-\theta(h_n))^2\right]\right)^2\right) + O\left(nE\left[(\bar{g}(X_1,h_n)-\theta(h_n))^4\right]\right) \Big\} \\
&= \binom{n}{2}^{-4} \left\{ O\left(n^2(n-1)^2 h_n^{-2\gamma}\right) + O\left(n(n-1)h_n^{-3\gamma}\right) + O\left(n^2\right) + O(n) \right\} \\
&= O\left(n^{-8}\right)\left\{ O\left(n^4 h_n^{-2\gamma}\right) + O\left(n^2 h_n^{-3\gamma}\right) + O\left(n^2\right) + O(n) \right\} \\
&= O\left(n^{-4}h_n^{-2\gamma}\right) + O\left(n^{-6}h_n^{-3\gamma}\right) \\
&= o\left(n^{-2}\right) + o\left(n^{-3}\right) \\
&= o\left(m^{-2}\right) \\
&= O\left(h_m^{4\alpha}\right) \\
&= O\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right) \qquad (53)
\end{aligned}
$$

for $h_m \propto m^{-\frac{2}{2\alpha+\gamma}}$.

Next,

$$
\begin{aligned}
& E\left[(\theta^*(h_m)-\theta(h_m))^4\right] \\
&= \frac{1}{n^8}\sum_{i_1,\ldots,i_8} E\left[(g(X_{i_1},X_{i_2},h_m-\theta(h_m))\cdots(g(X_{i_7},X_{i_8},h_m)-\theta(h_m))\right],
\end{aligned}
$$

where all terms in the summation with $i_1 \neq i_2 \neq \cdots \neq i_8$ have zero expectation. Then

$$
\begin{aligned}
& E\left[(\theta^*(h_m)-\theta(h_m))^4\right] \\
&= n^{-8}\Big\{ O\left(nE\left[(\bar{g}(X_1,h_m)-\theta(h_m))^4\right]\right) + O\left(n^2\left(E\left[(\bar{g}(X_1,h_m)-\theta(h_m))^2\right]\right)^2\right) \\
&\quad + O\left(n^2 E\left[(g(X_1,X_2,h_m)-\theta(h_m))^4\right]\right) \\
&\quad + O\left(n^2(n-1)^2\left(E\left[(g(X_1,X_2,h_m)-\theta(h_m))^2\right]\right)^2\right)\Big\} \\
&= n^{-8}\left\{ O(n) + O\left(n^2\right) + O\left(n^2 h_m^{-3\gamma}\right) + O\left(n^4 h_m^{-2\gamma}\right)\right\} \\
&= O\left(n^{-6}h_m^{-3\gamma}\right) + O\left(n^{-4}h_m^{-2\gamma}\right) \\
&= o\left(m^{-6}h_m^{-3\gamma}\right) + o\left(m^{-4}h_m^{-2\gamma}\right) \\
&= o\left(m^{-\frac{12\alpha}{2\alpha+\gamma}}\right) + o\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right) \\
&= o\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right) \qquad (54)
\end{aligned}
$$

for $h_m \propto m^{-\frac{2}{2\alpha+\gamma}}$.

In addition, for $\kappa(X_i^*, X_j^*, h_m)$ as defined in (73) in Appendix A.1,

$$\begin{aligned} & E\left[\left(\hat{\theta}_m^*(h_m) - \bar{\theta}_m^*(h_m)\right)^4 \middle| \mathcal{X}_n\right] \\ = & \binom{m}{2}^{-4} \sum_{i_1<i_2} \cdots \sum_{i_7<i_8} E\left[\kappa(X_{i_1}^*, X_{i_2}^*, h_m) \cdots \kappa(X_{i_7}^*, X_{i_8}^*, h_m) \middle| \mathcal{X}_n\right] \\ = & O\left(m^{-8}\right)\left\{O\left(m^2(m-1)^2\left(E\left[\kappa^2(X_i^*, X_j^*, h_m) \middle| \mathcal{X}_n\right]\right)^2\right)\right. \\ & \left.+O\left(m(m-1)E\left[\kappa^4(X_i^*, X_j^*, h_m) \middle| \mathcal{X}_n\right]\right)\right\} \\ = & O\left(m^{-4}E\left[g^2(X_i^*, X_j^*, h_m) \middle| \mathcal{X}_n\right]\right) + O\left(m^{-6}E\left[g^4(X_i^*, X_j^*, h_m) \middle| \mathcal{X}_n\right]\right). \end{aligned}$$

It follows that

$$\begin{aligned} & E\left[\left(\hat{\theta}_m^*(h_m) - \bar{\theta}_m^*(h_m)\right)^4\right] \\ = & O\left(m^{-4}E\left[g^2(X_1, X_2, h_m)\right]\right) + O\left(m^{-6}E\left[g^4(X_1, X_2, h_m)\right]\right) \\ = & O\left(m^{-4}h_m^{-2\gamma}\right) + O\left(m^{-6}h_m^{-3\gamma}\right) \\ = & O\left(m^{-\frac{12\alpha}{2\alpha+\gamma}}\right) + O\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right) \\ = & O\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right) \end{aligned} \tag{55}$$

for $h_m \propto m^{-\frac{2}{2\alpha+\gamma}}$.

Combining the condition of Assumption 2, (53)–(55), and Minkowski's inequality yields

$$\begin{aligned} & \left(E\left[\frac{\sigma_m^{*4}}{m^2}c_m^{*4}\right]\right)^{\frac{1}{4}} \\ \leq & \left(E\left[(\theta^*(h_m) - \theta(h_m))^4\right]\right)^{\frac{1}{4}} + \left((\theta_0 - \theta(h_n))^4\right)^{\frac{1}{4}} + \left(E\left[\left(\theta(h_n) - \hat{\theta}_n(h_n)\right)^4\right]\right)^{\frac{1}{4}} \\ & + \left(E\left[\left(\hat{\theta}_m^*(h_m) - \bar{\theta}_m^*(h_m)\right)^4\right]\right)^{\frac{1}{4}} \\ = & O\left(m^{-\frac{2\alpha}{2\alpha+\gamma}}\right) \end{aligned}$$

for $h_m \propto m^{-\frac{2}{2\alpha+\gamma}}$. Therefore

$$E\left[\frac{\sigma_m^{*4}}{m^2}c_m^{*4}\right] = O\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right). \tag{56}$$

It remains to consider the order of $E\left[\frac{\sigma_m^{*4}}{m^2}\right]$. Naturally, $E\left[\sigma_m^{*4}\right] = \left(E\left[\sigma_m^{*2}\right]\right)^2 + Var\left[\sigma_m^{*2}\right]$. A representation for $E\left[\sigma_m^{*2}\right]$ is available from Lemma 4, so the immediate focus is on $Var\left[\sigma_m^{*2}\right]$. In particular, the argument is made that $Var\left[\sigma_m^{*2}\right] = O(1)$. In this connection, define

$$\bar{g}_0^*(X_i^*) \equiv \lim_{h_m \to 0} \bar{g}^*(X_i^*, h_m) = \lim_{h_m \to 0} \frac{1}{n} \sum_{j=1}^{n} g(X_i^*, X_j, h_m)$$

and

$$\theta_0^* \equiv \frac{1}{n} \sum_{i=1}^{n} \bar{g}_0^*(X_i).$$

The following preliminary argument is made.

Lemma 5. *Under the conditions of Lemmas 3 and 4,*

$$\bar{g}^*(X_i^*, h_m) - \bar{g}_0(X_i^*) = E\left[s(X_1)\right] h_m^{\alpha} + o\left(h_m^{\alpha}\right) + o_p\left(n^{-\frac{\delta}{2}}\right).$$

Proof. The proof appears in Appendix A.4. □

It is immediate from Lemma 5 that $Var\left[\bar{g}^*(X_i^*, h_m)|\mathcal{X}_n\right] = Var\left[\bar{g}_0^*(X_i^*)|\mathcal{X}_n\right]$. Note that

$$\begin{aligned} Var\left[\bar{g}_0^*(X_i^*)|\mathcal{X}_n\right] &= \frac{1}{n} \sum_{i=1}^{n} \left(\bar{g}_0^*(X_i) - \theta_0^*\right)^2 \\ &= \frac{1}{n} \sum_{i=1}^{n} \bar{g}_0^{*2}(X_i) - \theta_0^{*2} \\ &= \frac{1}{n} \sum_{i=1}^{n} \bar{g}_0^{*2}(X_i) - \frac{1}{n^2} \sum_{i,j} \bar{g}_0^*(X_i)\bar{g}_0^*(X_j). \end{aligned}$$

Therefore

$$\begin{aligned} & Var\left[Var\left[\bar{g}_0^*(X_i^*)|\mathcal{X}_n\right]\right] \\ = & Var\left[\frac{1}{n} \sum_{i=1}^{n} \bar{g}_0^{*2}(X_i)\right] + Var\left[\frac{1}{n^2} \sum_{i,j} \bar{g}_0^*(X_i)\bar{g}_0^*(X_j)\right] \\ & -2Cov\left[\frac{1}{n} \sum_{i=1}^{n} \bar{g}_0^{*2}(X_i), \frac{1}{n^2} \sum_{i,j} \bar{g}_0^*(X_i)\bar{g}_0^*(X_j)\right] \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n} Var\left[\bar{g}_0^{*2}(X_1)\right] + \frac{1}{n^4} Var\left[\sum_{i=1}^{n} \bar{g}_0^{*2}(X_i) + 2\sum_{i\neq j} \bar{g}_0^{*}(X_i)\bar{g}_0^{*}(X_j)\right] \\
&\quad -\frac{2}{n^3}\sum_{i,j,k} Cov\left[\bar{g}_0^{*2}(X_i), \bar{g}_0^{*}(X_j)\bar{g}_0^{*}(X_k)\right]. \qquad (57)
\end{aligned}
$$

As such,

$$
\begin{aligned}
& Var\left[\bar{g}_0^{*2}(X_1)\right] \\
= \;& E\left[\bar{g}_0^{*4}(X_1)\right] - \left(E\left[\bar{g}_0^{*2}(X_1)\right]\right)^2 \\
\leq \;& E\left[\bar{g}_0^{*4}(X_1)\right] \\
= \;& \lim_{h_m\to 0} \frac{1}{n^4}\sum_{j,k,p,q} E\left[g(X_1,X_j,h_m)g(X_1,X_k,h_m)g(X_1,X_p,h_m)g(X_1,X_q,h_m)\right] \\
= \;& \lim_{h_m\to 0} \frac{1}{n^4}\sum_{j,k,p,q} E\left[\bar{g}^4(X_1,h_m)\right] \\
= \;& E\left[\bar{g}_0^{4}(X_1)\right] \\
= \;& O(1), \qquad (58)
\end{aligned}
$$

where the last equality follows from Assumption 4. In addition,

$$
\begin{aligned}
& Var\left[\bar{g}_0^{*}(X_1)\bar{g}_0^{*}(X_2)\right] \\
= \;& E\left[\bar{g}_0^{*2}(X_1)\bar{g}_0^{*2}(X_2)\right] - \left(E\left[\bar{g}_0^{*}(X_1)\right]E\left[\bar{g}_0^{*}(X_2)\right]\right)^2 \\
\leq \;& \left(E\left[\bar{g}_0^{*2}(X_1)\right]\right)^2 \\
= \;& \lim_{h_m\to 0} E\left[\frac{1}{n^2}\sum_{j,k} E\left[g(X_1,X_j,h_m)g(X_1,X_k,h_m)|X_1\right]\right] \\
= \;& \lim_{h_m\to 0} E\left[\bar{g}^2(X_1,h_m)\right] \\
= \;& E\left[\bar{g}_0^{2}(X_1)\right] \\
= \;& O(1), \qquad (59)
\end{aligned}
$$

where the last equality also follows from Assumption 4.

Now

$$
Cov\left[\sum_{i=1}^{n} \bar{g}_0^{*2}(X_i), 2\sum_{i\neq j} \bar{g}_0^{*}(X_i)\bar{g}_0^{*}(X_j)\right]
$$

$$
\begin{aligned}
&= 2\sum_{i=1}^{n}\sum_{j\neq k} Cov\left[\bar{g}_0^{*2}(X_i), \bar{g}_0^*(X_j)\bar{g}_0^*(X_k)\right] \\
&= 4n\sum_{k=2}^{n} Cov\left[\bar{g}_0^{*2}(X_1), \bar{g}_0^*(X_1)\bar{g}_0^*(X_k)\right] \\
&= 4n(n-1)\left\{E\left[\bar{g}_0^{*3}(X_1)\right] - E\left[\bar{g}_0^{*2}(X_1)\right]E\left[\bar{g}_0^*(X_1)\right]\right\}E\left[\bar{g}_0^*(X_2)\right] \\
&= 4n(n-1)\lim_{h_m\to 0}\left\{E\left[\frac{1}{n^3}\sum_{j,k,l} g(X_1,X_j,h_m)g(X_1,X_k,h_m)g(X_1,X_l,h_m)\right]\right. \\
&\quad \left. -E\left[\frac{1}{n^2}\sum_{j,k} g(X_1,X_j,h_m)g(X_1,X_k,h_m)\right]E\left[\frac{1}{n}\sum_{j} g(X_1,X_j,h_m)\right]\right\} \\
&\quad \cdot E\left[\frac{1}{n}\sum_{j} g(X_1,X_j,h_m)\right] \\
&= 4n(n-1)\lim_{h_m\to 0}\left\{E\left[\bar{g}^3(X_1,h_m)\right] - E\left[\bar{g}^2(X_1,h_m)\right]E\left[\bar{g}(X_1,h_m)\right]\right\} \\
&\quad \cdot E\left[\bar{g}(X_1,h_m)\right] \\
&= 4n(n-1)\left\{E\left[\bar{g}_0^3(X_1)\right] - E\left[\bar{g}_0^2(X_1)\right]\theta_0\right\}\theta_0 \\
&= O(n^2); \qquad (60)
\end{aligned}
$$

and

$$
\begin{aligned}
&\sum_{i,j,k} Cov\left[\bar{g}_0^{*2}(X_i), \bar{g}_0^*(X_j)\bar{g}_0^*(X_k)\right] \\
&= \sum_{i,j} Cov\left[\bar{g}_0^{*2}(X_i), \bar{g}_0^{*2}(X_j)\right] + \sum_{i=1}^{n}\sum_{j\neq k} Cov\left[\bar{g}_0^{*2}(X_i), \bar{g}_0^*(X_j)\bar{g}_0^*(X_k)\right] \\
&= nVar\left[\bar{g}_0^{*2}(X_1)\right] + 2n\sum_{k=2}^{n} Cov\left[\bar{g}_0^{*2}(X_1), \bar{g}_0^*(X_1)\bar{g}_0^*(X_k)\right] \\
&= nVar\left[\bar{g}_0^{*2}(X_1)\right] + 2n(n-1)\left\{E\left[\bar{g}_0^{*3}(X_1)\right] - E\left[\bar{g}_0^{*2}(X_1)\right]E\left[\bar{g}_0^*(X_1)\right]\right\} \\
&= nVar\left[\bar{g}_0^{*2}(X_1)\right] + 2n(n-1)\left\{E\left[\bar{g}_0^{*3}(X_1)\right] - E\left[\bar{g}_0^{*2}(X_1)\right]\theta_0\right\} \\
&\leq nE\left[\bar{g}_0^{*4}(X_1)\right] + 2n(n-1)\left\{E\left[\bar{g}_0^{*3}(X_1)\right] - E\left[\bar{g}_0^{*2}(X_1)\right]\theta_0\right\} \\
&= O(n^2). \qquad (61)
\end{aligned}
$$

Combining (57), (58), (59), (60) and (61), we have that $Var\left[Var\left[\bar{g}_0^*(X_i^*)|\mathcal{X}_n\right]\right] = O(1)$. Therefore

$$
Var\left[\sigma_m^{*2}\right] = 16Var\left[Var\left[\bar{g}^*(X_i^*, h_m)|\mathcal{X}_n\right]\right]
$$

$$\begin{aligned} &= 16Var\left[Var\left[\bar{g}_0^*(X_i^*)|\mathcal{X}_n\right]\right] \\ &= O(1). \end{aligned}$$

As such, for $h_m \propto m^{-\frac{2}{2\alpha+\gamma}}$,

$$\begin{aligned} E\left[\frac{\sigma^{*4}}{m^2}\right] &= m^{-2}\left(E\left[\sigma_m^{*2}\right]\right)^2 + m^{-2}\cdot O(1) \\ &= \left(O\left(m^{-1}\cdot m^{-1}h_m^{-\gamma}\right)\right)^2 + O\left(m^{-2}\right) \\ &= \left(O\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right)\right)^2 \\ &= O\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right). \end{aligned} \tag{62}$$

Returning to (49) above, we have from Assumption 2, (50)–(56) and (62) that for $h_m \propto m^{-\frac{2}{2\alpha+\gamma}}$,

$$\begin{aligned} & E\left[\frac{\sigma_m^{*4}}{m^2}\left\{E\left[\left.(W+b_m^*+c_m^*)^2\right|\mathcal{X}_n\right]\right\}^2\right] \\ \leq\ & O\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right) + O\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right)O\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right) \\ & +O\left(m^{-\frac{2\alpha}{2\alpha+\gamma}}\right)\left\{O\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right)\right\}^{\frac{1}{2}}\left\{o\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right)\right\}^{\frac{1}{2}} \\ & +\left\{O\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right)\right\}^{\frac{1}{2}}\left\{o\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right)\right\}^{\frac{1}{2}} + O\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right) + O\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right)o\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right) \\ & +o\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right) + O\left(m^{-\frac{6\alpha}{2\alpha+\gamma}}\right)o\left(m^{-\frac{2\alpha}{2\alpha+\gamma}}\right) + O\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right)o\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right) \\ & +O\left(m^{-\frac{2\alpha}{2\alpha+\gamma}}\right)\left\{o\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right)\right\}^{\frac{1}{2}}\left\{o\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right)\right\}^{\frac{1}{2}} \\ =\ & O\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right) + O\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right) + o\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right) + o\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right) + O\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right) \\ & +o\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right) + o\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right) + o\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right) + o\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right) + o\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right) \\ =\ & O\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right). \end{aligned} \tag{63}$$

It follows from (63) that for $h_m \propto m^{-\frac{2}{2\alpha+\gamma}}$,

$$m^{\frac{4\alpha}{2\alpha+\gamma}}\left(\frac{\sigma_m^{*2}}{m}E\left[\left.(W+b_m^*+c_m^*)^2\right|\mathcal{X}_n\right] - E\left[\frac{\sigma_m^{*2}}{m}\left(W+b_m^*+c_m^*\right)^2\right]\right) = o_p(1),$$

and so it is possible to write, with an appeal to Lemma 1,

$$
\begin{aligned}
& \frac{\sigma_m^{*2}}{m} E\left[\left(W + b_m^* + c_m^*\right)^2 \middle| \mathcal{X}_n \right] \\
= & \frac{4}{m} Var\left[\bar{g}_0(X_1)\right] + \left(\theta(h_m) - \theta_0\right)^2 + \frac{2}{m^2} E\left[g^2(X_1, X_2, h_m)\right] + O\left(m^{-1} h_m^{\alpha}\right) \\
& + o\left(h_m^{2\alpha}\right) + o\left(h_m^{2\alpha}\right) + o\left(m^{-2} h_m^{-\gamma}\right) + o_p\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right) \\
= & \frac{4}{m} Var\left[\bar{g}_0(X_1)\right] + \left(\theta(h_m) - \theta_0\right)^2 + \frac{2}{m^2} E\left[g^2(X_1, X_2, h_m)\right] \\
& + o_p\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right) \qquad (64)
\end{aligned}
$$

for all $h_m \propto m^{-\frac{2}{2\alpha+\gamma}}$.

Now consider $|R_m^*(h_m)|$ as given in (47) above. When $h_m \propto m^{-\frac{2}{2\alpha+\gamma}}$,

$$
h_m^{\alpha} \propto m^{-1} h_m^{-\frac{\gamma}{2}},
$$

so $o_p\left(n^{-\delta} h_m^{\alpha}\right) = o_p\left(n^{-\delta} m^{-1} h_m^{-\frac{\gamma}{2}}\right) = o_p\left(n^{-2\delta} h_m^{-\frac{\gamma}{2}}\right)$, which implies that

$$
\begin{aligned}
|R_m^*(h_m)| &= o_p\left(n^{-\delta} h_m^{\alpha}\right) \\
&= o_p\left(m^{-1} m^{-\frac{2\alpha}{2\alpha+\gamma}}\right) \\
&= o_p\left(m^{-\frac{4\alpha-\gamma}{2\alpha+\gamma}}\right) \\
&= o_p\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right). \qquad (65)
\end{aligned}
$$

Thus (48), (64) and (65) imply the representation

$$
\begin{aligned}
MSE\left[\hat{\theta}_m^*(h_m) \middle| \mathcal{X}_n\right] &= \frac{4}{m} Var\left[\bar{g}_0(X_1)\right] + \left\{E\left[s(X_1)\right]\right\}^2 h_m^{2\alpha} + \frac{2}{m^2} E\left[q(X_1)\right] h_m^{-\gamma} \\
& \quad + o_p\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right) \qquad (66)
\end{aligned}
$$

for all $h_m \propto m^{-\frac{2}{2\alpha+\gamma}}$.

Now suppose $h_m = k m^{-\frac{2}{2\alpha+\gamma}}$, where $k \in (0, \infty)$. Pick $\epsilon > 0$. Assuming that $k \in [\epsilon, \epsilon^{-1}]$, define

$$
\mu_m(k) \equiv \frac{4}{m} Var\left[\bar{g}_0(X_1)\right] + \left\{E\left[s(X_1)\right]\right\}^2 h_m^{2\alpha} + \frac{2}{m^2} E\left[q(X_1)\right] h_m^{-\gamma}
$$

and

$$\rho_m(k) \equiv \frac{\sigma_m^{*2}}{m} E\left[(W + b_m^* + c_m^*)^2 \middle| \mathcal{X}_n \right] - \mu_m(k).$$

In order to extend the representation of (66) uniformly for values of the scaling constant $k \in [\epsilon, \epsilon^{-1}]$, the argument is made that for any $\eta > 0$,

$$P\left[\sup_{k \in [\epsilon, \epsilon^{-1}]} |\rho_m(k)| > \eta m^{-\frac{4\alpha}{2\alpha+\gamma}} \right] \to 0 \tag{67}$$

as $m \to \infty$.

Let $M_{\rho_m}(t)$ denote the moment-generating function of $\rho_m(k)$ for $k \in [\epsilon, \epsilon^{-1}]$. Assume that $t \in \left(0, m^{\frac{4\alpha}{2\alpha+\gamma}}\right)$ so that $|t\rho_m(k)|$ is bounded in probability for sufficiently large m. Arguing via Taylor's theorem, there exists a constant $c > 0$ such that

$$\log M_{\rho_m}(t) \leq tE\left[\rho_m(k)\right] + ct^2 Var\left[\rho_m(k)\right].$$

An application of Markov's inequality accordingly produces

$$P\left[|\rho_m(k)| > \eta m^{-\frac{4\alpha}{2\alpha+\gamma}} \right] \leq \exp\left\{ -\eta m^{-\frac{4\alpha}{2\alpha+\gamma}} t + tE\left[\rho_m(k)\right] + ct^2 Var\left[\rho_m(k)\right] \right\}, \tag{68}$$

while results established earlier give the bounds

$$E\left[\rho_m(k)\right] = o\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right) \tag{69}$$

and

$$Var\left[\rho_m(k)\right] = O\left(m^{-\frac{8\alpha}{2\alpha+\gamma}}\right). \tag{70}$$

Now pick

$$t = m^{-\frac{2\alpha}{2\alpha+\gamma}} \log m. \tag{71}$$

Combining (69)–(71) in (68), we find that there exists a constant $m_0 \geq 1$ not depending on k such that for $m \geq m_0$,

$$P\left[|\rho_m(k)| > \eta m^{-\frac{4\alpha}{2\alpha+\gamma}} \right] \leq \exp\left(-\eta \log m\right) = m^{-\eta} \tag{72}$$

for all $\eta > 0$.

Now consider that for $k, k' \in [\epsilon, \epsilon^{-1}]$,

$$\begin{aligned} |\rho_m(k) - \rho_m(k')| &\leq 2\,|\mu_m(k') - \mu_m(k)| + o_p\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right) \\ &\leq 2m^{-\frac{4\alpha}{2\alpha+\gamma}}\left(\{E\,[s(X_1)]\}^2\left|k'^{2\alpha} - k^{2\alpha}\right| + E\,[q(X_1)]\left|k'^{-\gamma} - k^{-\gamma}\right|\right) \\ &\quad + o_p\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right) \\ &= O_p\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right). \end{aligned}$$

As such, $|\rho_m(k) - \rho_m(k')| = O_p\left(m^{-\frac{4\alpha}{2\alpha+\gamma}}\right)$ uniformly as $m \to \infty$ for $\{k, k'\} \subset [\epsilon, \epsilon^{-1}]$.

Define a partition $K_m \equiv \{k_{im}\}$ of the interval $[\epsilon, \epsilon^{-1}]$ with

$$k_{im} \equiv \epsilon + i m^{-\zeta},$$

where for some $\zeta > 0$, $i = 0, 1, 2, \ldots, m^{\zeta}(\epsilon^{-1} - \epsilon)$. The cardinality of K_m is accordingly $O\left(m^{\zeta}\right)$. Exploiting (72), it follows from the chaining argument in the previous paragraph that

$$P\left[\sup_{k \in K} |\rho_m(k)| > \eta m^{-\frac{4\alpha}{2\alpha+\gamma}}\right] \leq m^{\zeta - \eta}$$

for m sufficiently large. Set $\zeta \in (0, \eta)$ and (67) follows, which is the desired result.

5 Numerical Evidence

This section presents results of a modest simulation experiment comparing the performance of the m-bootstrap bandwidth estimate given above in (31) with a number of implementations of the "plug-in" bandwidth estimator proposed by Powell and Stoker (1996, §4.4). In the experiment considered here, the objective is to estimate the average density

$$\theta_0 \equiv E\,[f(X_1)]$$

of a sample of scalar-valued observations given by $\mathcal{X}_n \equiv \{X_1, \ldots, X_n\}$. One hundred Monte Carlo replications of $\mathcal{X}_n$ were drawn from a standard normal distribution. The sample sizes were fixed at $n = 50$ throughout, and the estimator $\hat{\theta}_n(h)$ was constructed using a standard normal smoothing kernel.

In the case of the simulations reported here, the true value of θ_0 is straightforwardly calculated as

$$\theta_0 = \frac{1}{2\sqrt{\pi}} \approx .2821,$$

while the constants α and γ of Assumptions 2 and 3, respectively, are given by $\alpha = 2$ and $\gamma = 1$. The true MSE-minimizing bandwidth for the estimator and data-generating process considered in this experiment can be approximated to order $o\left(n^{-\frac{2}{5}}\right)$ by the expression given in the statement of Powell and Stoker (1996, Proposition 4.1), which in this context can be shown to be

$$h_{n,opt} = 8^{\frac{1}{5}} n^{-\frac{2}{5}}.$$

For $n = 50$, the true MSE-optimal bandwidth is accordingly given by $h_{50,opt} \approx .3170$.[19] Estimates of the true MSE-optimal bandwidth are naturally called for when $h_{n,opt}$ cannot be computed due to insufficient information regarding the underlying data-generating process.

Table 1 summarizes the simulated performance of three different plug-in estimators of $h_{n,opt}$ as well as three different implementations of the m-bootstrap estimator of the same estimand. The plug-in estimates of $h_{n,opt}$ were computed in accordance with the proposal given in Powell and Stoker (1996, §4.4), which involves replacement of the unknown quantities in the leading constant of $h_{n,opt}$ with natural empirical counterparts. In particular, the plug-in bandwidth estimator has the form

$$\tilde{h}_{n,h_0,\tau} \equiv \left(\frac{\gamma \hat{Q}_n(h_0)}{\alpha \hat{S}_n^2(\tau, h_0)} \right)^{\frac{1}{2\alpha+\gamma}} \cdot n^{-\frac{2}{2\alpha+\gamma}},$$

where

$$\hat{Q}_n(h_0) \equiv \binom{n}{2}^{-1} \sum_{i<j} h_0^{\gamma} g^2(X_i, X_j, h_0)$$

and

$$\hat{S}_n(\tau, h_0) \equiv \frac{\hat{\theta}_n(\tau h_0) - \hat{\theta}_n(h_0)}{(\tau h_0)^{\alpha} - h_0^{\alpha}}.$$

Here $h_0 > 0$ denotes a pilot bandwidth and $\tau \neq 1$ a positive secondary smoothing parameter, while α, γ, $g(\cdot,\cdot,\cdot)$ and $\hat{\theta}_n(\cdot)$ are all as given above in Section 3.1.

[19] The general expression for $h_{n,opt}$ is reproduced in n. 11 above. Cf. also the tabulated values of $h_{n,opt}$ in Powell and Stoker (1996, Table 1) for the data-generating process considered by the simulations reported here.

The selection of h_0 and τ is obviously an integral part of implementing the plug-in bandwidth estimator of $h_{n,opt}$. In the simulations reported here, h_0 is fixed at $h_{n,opt} = .3170$, while $\tau \in \left\{1.1, \frac{2}{3} \times 1.1, \frac{3}{2} \times 1.1\right\}$. The initial setting $\tau = 1.1$ was settled upon after some experimentation by the author.

The m-bootstrap estimator as given by the expression in (31) was implemented in the simulations reported here by first computing the m-bootstrap MSE estimator given above in (29) repeatedly for bandwidths $h_m = km^{-\frac{2}{5}}$, where the scaling constant k ranges over a grid of 100 equally spaced values covering the interval $[0.01, 3.00]$. These initial computations were followed by finding the setting of $k = k^*$ in the grid covering $[0.01, 3.00]$ that produces the smallest realized value of $MSE\left[\hat{\theta}^*_m\left(km^{-\frac{2}{5}}\right)\middle| \mathcal{X}_n\right]$ in the first step. The bandwidth estimate $\hat{h}_m = k^*m^{-\frac{2}{5}}$ is then rescaled in accordance with the expression given in (31) to produce a bandwidth estimate $\hat{h}_{n,m}$ appropriate for the full sample. In each of the initial computations of $MSE\left[\hat{\theta}^*_m\left(km^{-\frac{2}{5}}\right)\middle| \mathcal{X}_n\right]$, the pilot bandwidth $h_n \propto n^{-\frac{2}{5}}$ required for bias estimation was taken to be the plug-in bandwidth estimator of Powell and Stoker (1996, §4.4) described above with $h_0 = h_{n,opt}$ and $\tau = 1.1$. The resample size m was set to diverge in accordance with the optimal rate suggested above in Section 3.3. In particular, the setting $m = \lfloor \kappa\sqrt{n} \rfloor$ was used, where $\kappa = 1, \frac{2}{3}, \frac{3}{2}$ in order to investigate the sensitivity of the results in small samples to variation in this particular scaling constant.[20]

The results displayed in Table 1 suggest a degree of sensitivity of the bandwidth estimates considered to changes in τ and m, although the m-bootstrap procedure seems to be relatively more insensitive to changes in the resample size than the plug-in estimates are to variation in τ. Both sets of implementations of the bandwidth estimators considered here lead to biased estimates of the optimal bandwidth.[21] Use of the m-bootstrap appears to lead to bandwidth estimates with significantly greater precision than those produced by the plug-in procedure.

Table 2 reports the simulated sampling behaviour of the average density estimates constructed using each of the implementations of the plug-in and m-bootstrap bandwidth estimators described above. The results summarized in Table 2 seem to suggest that use of the m-bootstrap leads to average density estimators with sampling performance superior to that induced by use of the plug-in bandwidth estimates.

[20] For any x, $\lfloor x \rfloor$ denotes the largest integer less than or equal to x.

[21] The degree of bias in the bandwidth estimates evident from a glance at Table 1 is perhaps unsurprising given the small sample size used in the simulations.

6 Conclusion

This paper has presented a bootstrap method for estimating the mean squared error and the associated asymptotically optimal bandwidth for density-weighted averages. For samples of size n, the bootstrap procedure described above involves the resampling of $m < n$ observations without replacement from the original sample. This method is shown to work in the sense that the m-bootstrap mean squared error estimate $MSE\left[\hat{\theta}^*_m(h_m)\middle|\,\mathcal{X}_n\right]$ is so close in a uniform sense to the actual mean squared error $MSE\left[\hat{\theta}_m(h_m)\right]$ of the estimator computed for a sample of size m that the bandwidth that minimizes one is asymptotically equivalent to the bandwidth that minimizes the other. An estimate of the asymptotically optimal bandwidth for the estimator computed using the full sample of size n can then be generated by an appropriate rescaling of the bandwidth found to minimize $MSE\left[\hat{\theta}^*_m(h_m)\middle|\,\mathcal{X}_n\right]$.[22] Specific guidance on selecting the resample size m in applications was also given. In particular, for resamples of size $m \propto n^\delta$ with $\delta \in (0, 1)$ a constant satisfying $\delta \leq 1 - \delta$, it was shown that a setting of $m \propto \sqrt{n}$ is sufficient to minimize the stochastic order of magnitude of the discrepancy between the leading terms of the expansion of $MSE\left[\hat{\theta}_m(h_m)\right]$ and its m-bootstrap estimate when these quantities are evaluated at an optimal bandwidth sequence.

Natural alternatives to the bootstrap method presented above include plug-in methods of the sort described by Powell and Stoker (1996, §4.4) and methods based on the full-sample bootstrap coupled with an explicit method of bias correction. Simulation evidence presented above in Section 5 seems to suggest that the m-bootstrap method of bandwidth estimation presented in this paper leads to estimators of density-weighted averages with sampling performance in small samples superior to that induced by certain implementations of the plug-in estimator of the optimal bandwidth suggested by Powell and Stoker (1996, §4.4). Further work on the relationship between the sampling behaviour of bandwidth estimators and the sampling behaviour of the semiparametric estimators in which they are embedded would appear to be fruitful from both a theoretical and an applied viewpoint.

[22]Cf. the discussion surrounding (31) above.

A Proofs of lemmas not proved in the main text

A.1 Proof of Lemma 2

Set

$$\kappa(X_i^*, X_j^*, h_m) \equiv g(X_i^*, X_j^*, h_m) - \bar{g}^*(X_i^*, h_m) - \bar{g}^*(X_j^*, h_m) + \theta^*(h_m). \tag{73}$$

As such,

$$\hat{\theta}_m^*(h_m) - \bar{\theta}_m^*(h_m) = \binom{m}{2}^{-1} \sum_{i<j} \kappa(X_i^*, X_j^*, h_m),$$

and

$$E\left[\left(\hat{\theta}_m^*(h_m) - \bar{\theta}_m^*(h_m)\right)^2 \middle| \mathcal{X}_n\right] = \binom{m}{2}^{-1} \sum_{i<j} \sum_{k<l} E\left[\kappa(X_i^*, X_j^*, h_m)\kappa(X_k^*, X_l^*, h_m) \middle| \mathcal{X}_n\right].$$

By the independence—conditional on $\mathcal{X}_n$—of the elements of $\mathcal{X}_m^*$,

$$E\left[\left(\hat{\theta}_m^*(h_m) - \bar{\theta}_m^*(h_m)\right)^2 \middle| \mathcal{X}_n\right] = \binom{m}{2}^{-2} \sum_{i<j} E\left[\kappa^2(X_i^*, X_j^*, h_m)|\mathcal{X}_n\right].$$

Now $O\left(E\left[\kappa^2(X_i^*, X_j^*, h_m)\middle| \mathcal{X}_n\right]\right) = O\left(E\left[g^2(X_i^*, X_j^*, h_m)\middle| \mathcal{X}_n\right]\right)$, and

$$E\left[g^2(X_i^*, X_j^*, h_m)\middle| \mathcal{X}_n\right] = \frac{1}{n^2}\sum_{i,j} g^2(X_i, X_j, h_m),$$

which has expectation by Assumption 3 equal to

$$\begin{aligned} E\left[E\left[g^2(X_i^*, X_j^*, h_m)\middle| \mathcal{X}_n\right]\right] &= E\left[g^2(X_1, X_2, h_m)\right] \\ &= E[q(X_1)]h_m^{-\gamma} + o\left(h_m^{-\gamma}\right) \end{aligned}$$

for $h_m \in \mathcal{H}_m(\alpha, \gamma)$.

Therefore for $h_m \in \mathcal{H}_m(\alpha, \gamma)$,

$$\begin{aligned} E\left[E\left[\left(\hat{\theta}_m^*(h_m) - \bar{\theta}_m^*(h_m)\right)^2 \middle| \mathcal{X}_n\right]\right] &= \binom{m}{2}^{-2} \cdot O(m^2) \cdot O\left(E\left[E\left[g^2(X_i^*, X_j^*, h_m)\middle| \mathcal{X}_n\right]\right]\right) \\ &= O\left(m^{-2}\right) \cdot O\left(h_m^{-\gamma}\right) \\ &= O\left(m^{-2}h_m^{-\gamma}\right) \\ &= o(1). \end{aligned}$$

As such,

$$\begin{aligned} m^2 h_m^{\gamma} E\left[E\left[\left(\hat{\theta}_m^*(h_m) - \bar{\theta}_m^*(h_m)\right)^2 \middle| \mathcal{X}_n\right]\right] &= m^2 h_m^{\gamma} E\left[\left(\hat{\theta}_m^*(h_m) - \bar{\theta}_m^*(h_m)\right)^2\right] \\ &= o(1), \end{aligned}$$

from which it follows, via Markov's inequality, that $m h_m^{\frac{\gamma}{2}} \left(\hat{\theta}_m^*(h_m) - \bar{\theta}_m^*(h_m)\right) = o_p(1)$ as $m \to \infty$.

A.2 Proof of Lemma 3

Note that

$$|g(X_i, X_j, h_m) - \theta^*(h_m)| = \left| \frac{1}{n^2} \sum_{k,l} \{g(X_i, X_j, h_m) - g(X_k, X_l, h_m)\} \right|,$$

and that

$$\begin{aligned} & E\left[\left(\sum_{k,l} \{g(X_k, X_l, h_m) - g(X_i, X_j, h_m)\}\right)^2\right] \\ = \;& n^2 E\left[(g(X_k, X_l, h_m) - g(X_i, X_j, h_m))^2\right] \\ & +2n^2(n-1)E\left[(g(X_k, X_l, h_m) - g(X_i, X_j, h_m))\,(g(X_k, X_q, h_m) - g(X_i, X_j, h_m))\right]. \end{aligned}$$

But

$$\begin{aligned} & E\left[(g(X_k, X_l, h_m) - g(X_i, X_j, h_m))^2\right] \\ = \;& 2E\left[g^2(X_k, X_l, h_m)\right] - 2E\left[E\left[g(X_k, X_l, h_m)g(X_i, X_j, h_m)|\, X_k, X_l\right]\right] \\ = \;& 2\left(E\left[g^2(X_k, X_l, h_m)\right] - \theta^2(h_m)\right) \\ = \;& 2E\left[q(X_1)\right] h_m^{-\gamma} + o(h_m^{-\gamma}) - 2\theta_0^2 + O\left(h_m^{\alpha}\right), \end{aligned}$$

and

$$\begin{aligned} & E\left[(g(X_k, X_l, h_m) - g(X_i, X_j, h_m))\,(g(X_k, X_q, h_m) - g(X_i, X_j, h_m))\right] \\ = \;& E\left[E\left[(g(X_k, X_l, h_m) - g(X_i, X_j, h_m))\,(g(X_k, X_q, h_m) - g(X_i, X_j, h_m))\,|X_k, X_i, X_j\right]\right] \\ = \;& E\left[(\bar{g}(X_k, h_m) - g(X_i, X_j, h_m))\,(\bar{g}(X_k, h_m) - g(X_i, X_j, h_m))\right] \\ = \;& E\left[E\left[(\bar{g}(X_k, h_m) - g(X_i, X_j, h_m))\,(\bar{g}(X_k, h_m) - g(X_i, X_j, h_m))\,|X_i, X_j\right]\right] \\ = \;& E\left[E\left[\bar{g}^2(X_k, h_m)\right] - \theta(h_m)g(X_i, X_j, h_m) - \theta(h_m)g(X_i, X_j, h_m) + g^2(X_i, X_j, h_m)\right] \\ = \;& E\left[\bar{g}^2(X_k, h_m)\right] - 2\theta^2(h_m) + E\left[g^2(X_i, X_j, h_m)\right] \\ = \;& Var\left[\bar{g}(X_k, h_m)\right] - \theta^2(h_m) + E\left[q(X_1)\right] h_m^{-\gamma} + o\left(h_m^{-\gamma}\right) \\ = \;& Var\left[\bar{g}_0(X_1)\right] + 2Cov\left[\bar{g}_0(X_1), s(X_1)\right] h_m^{\alpha} + o\left(h_m^{\alpha}\right) - \theta_0^2 + O\left(h_m^{\alpha}\right) + E\left[q(X_1)\right] h_m^{-\gamma} + o\left(h_m^{-\gamma}\right). \end{aligned}$$

From this it follows that

$$\begin{aligned} & \frac{1}{n^4} E\left[\left(\sum_{k,l} |g(X_k, X_l, h_m) - g(X_i, X_j, h_m)|\right)^2\right] \\ = \;& \frac{2}{n^2} E\left[q(X_1)\right] h_m^{-\gamma} + o\left(n^{-2} h_m^{-\gamma}\right) - \frac{2\theta_0^2}{n^2} + O\left(n^{-2} h_m^{\alpha}\right) + \frac{2}{n} Var\left[\bar{g}_0(X_1)\right] \\ & + \frac{4}{n} Cov\left[\bar{g}_0(X_1), s(X_1)\right] h_m^{\alpha} + o\left(n^{-1} h_m^{\alpha}\right) - \frac{\theta_0^2}{n} + O\left(n^{-1} h_m^{\alpha}\right) + \frac{1}{n} E\left[q(X_1)\right] h_m^{-\gamma} \end{aligned}$$

$$
\begin{aligned}
& +o\left(n^{-1}h_m^{-\gamma}\right)+O\left(n^{-2}\right) \\
= \quad & O\left(n^{-1}\right)+O\left(n^{-1}h_m^{\alpha}\right)+O\left(n^{-1}h_m^{-\gamma}\right) \\
= \quad & O\left(n^{-1}\right)+o\left(\frac{1}{n\sqrt{m}}\right)+o\left(\frac{m}{n}\right) \\
= \quad & o\left(\frac{m}{n}\right) \\
= \quad & o\left(n^{1-\delta-1}\right) \\
= \quad & o\left(n^{-\delta}\right).
\end{aligned}
$$

By Chebyshev's inequality, therefore,

$$
\frac{1}{n^2}\sum_{k,l}|g(X_k,X_l,h_m)-g(X_i,X_j,h_m)|=o_p\left(n^{-\frac{\delta}{2}}\right),
$$

which implies that $|g(X_i,X_j,h_m)-\theta^*(h_m)|=o_p\left(n^{-\frac{\delta}{2}}\right)$.

A.3 Proof of Lemma 4

We have

$$
\begin{aligned}
& E\left[\sigma_m^{*2}\right] \\
= \quad & \frac{4}{n}\sum_{i=1}^{n}E\left[g(X_i,X_2,h_m)g(X_i,X_3,h_m)\right]-\frac{4}{n^4}\left\{n^2E\left[g^2(X_1,X_2,h_m)\right]\right. \\
& \left.+2n^2(n-1)E\left[g(X_1,X_2,h_m)g(X_1,X_3,h_m)\right]+n^2(n-1)^2\left(E\left[g(X_1,X_2,h_m)\right]\right)^2\right\} \\
= \quad & \frac{4}{n}\sum_{i=1}^{n}E\left[E\left[g(X_i,X_2,h_m)g(X_i,X_3,h_m)|\,X_i\right]\right]-4\left\{\frac{1}{n^2}E\left[g^2(X_1,X_2,h_m)\right]\right. \\
& \left.+\frac{2(n-1)}{n^2}E\left[E\left[g(X_1,X_2,h_m)g(X_1,X_3,h_m)|\,X_1\right]\right]+\frac{(n-1)^2}{n^2}\theta^2(h_m)\right\} \\
= \quad & \frac{4}{n}\sum_{i=1}^{n}E\left[\bar{g}^2(X_i,h_m)\right]-4\left\{\frac{1}{n^2}E\left[g^2(X_1,X_2,h_m)\right]+\frac{2(n-1)}{n^2}E\left[\bar{g}^2(X_1,h_m)\right]+\frac{(n-1)^2}{n^2}\theta^2(h_m)\right\} \\
= \quad & 4\left\{E\left[\bar{g}^2(X_1,h_m)\right]-\theta^2(h_m)+\left(\frac{1}{n^2}-\frac{2}{n}\right)\left(E\left[\bar{g}^2(X_1,h_m)\right]-\theta^2(h_m)\right)+\frac{1}{n^2}E\left[\bar{g}^2(X_1,h_m)\right]\right. \\
& \left.-\frac{1}{n^2}E\left[g^2(X_1,X_2,h_m)\right]\right\} \\
= \quad & \frac{4(n^2+2-2n)}{n^2}Var\left[\bar{g}(X_1,h_m)\right]-\frac{4}{n^2}Var\left[g(X_1,X_2,h_m)\right] \\
= \quad & 4Var\left[\bar{g}(X_1,h_m)\right]+O\left(n^{-1}\right)-\frac{4}{n^2}Var\left[g(X_1,X_2,h_m)\right] \\
= \quad & mVar\left[\hat{\theta}_m(h_m)\right]+o\left(m^{-1}h_m^{-\gamma}\right)-\left(\frac{2}{m}+\frac{4}{n^2}\right)E\left[g^2(X_1,X_2,h_m)\right]+O(n^{-1})+O\left(n^{-2}\right)
\end{aligned}
$$

$$= \quad mVar\left[\hat{\theta}_m(h_m)\right] + o(1) + O\left(\frac{2}{m}E\left[g^2(X_1, X_2, h_m)\right]\right).$$

for $h_m \in \mathcal{H}_m(\alpha,\gamma)$. Note that $O\left(\frac{2}{m}E\left[g^2(X_1,X_2,h_m)\right]\right) = o(1)$ for $h_m \in \mathcal{H}_m(\alpha,\gamma)$. The desired result follows from (16) and (17) above.

A.4 Proof of Lemma 5

Note that

$$\begin{aligned}
E\left[\bar{g}^*(X_i^*, h_m)\right] &= E\left[E\left[\bar{g}^*(X_i^*, h_m)|\mathcal{X}_n\right]\right] \\
&= E\left[\frac{1}{n}\sum_{i=1}^n \bar{g}^*(X_i, h_m)\right] \\
&= E\left[\frac{1}{n^2}\sum_{i,j} g(X_i, X_j, h_m)\right] \\
&= \theta(h_m),
\end{aligned}$$

and that

$$\begin{aligned}
E\left[\bar{g}_0^*(X_i^*)\right] &= E\left[E\left[\bar{g}_0^*(X_i^*)|\mathcal{X}_n\right]\right] \\
&= E\left[\frac{1}{n}\sum_{i=1}^n \bar{g}_0^*(X_i)\right] \\
&= E\left[\bar{g}_0^*(X_1)\right] \\
&= \lim_{h_m\to 0}\frac{1}{n}\sum_{j=1}^n E\left[g(X_1, X_j, h_m)\right] \\
&= \theta_0.
\end{aligned}$$

Therefore $E\left[\bar{g}^*(X_i^*, h_m) - \bar{g}_0^*(X_i^*)\right] = \theta(h_m) - \theta_0$.

Now consider $E\left[\left(\bar{g}^*(X_i^*, h_m) - \bar{g}_0^*(X_i^*)\right)^2\right]$. We have

$$\begin{aligned}
&E\left[\left(\bar{g}^*(X_i^*, h_m) - \bar{g}_0^*(X_i^*)\right)^2\right] \\
= \;& E\left[E\left[\left(\bar{g}^*(X_i^*, h_m) - \bar{g}_0^*(X_i^*)\right)^2\middle|\, \mathcal{X}_n\right]\right] \\
= \;& E\left[\frac{1}{n}\sum_{i=1}^n \left(\bar{g}^*(X_i, h_m) - \bar{g}_0^*(X_i)\right)^2\right] \\
= \;& E\left[\left(\bar{g}^*(X_1, h_m) - \bar{g}_0^*(X_1)\right)^2\right] \\
= \;& E\left[\left(\frac{1}{n}\sum_{j=1}^n \left(g(X_1, X_j, h_m) - \bar{g}_0^*(X_1)\right)\right)^2\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n^2}\sum_{j,k} E\left[\left(g(X_1,X_j,h_m)-\bar{g}_0^*(X_1)\right)\left(g(X_1,X_k,h_m)-\bar{g}_0^*(X_1)\right)\right] \\
&= \frac{1}{n^2}\Big\{nE\left[\left(g(X_1,X_2,h_m)-\bar{g}_0^*(X_1)\right)^2\right] \\
&\quad +n(n-1)E\left[\left(g(X_1,X_2,h_m)-\bar{g}_0^*(X_1)\right)\left(g(X_1,X_3,h_m)-\bar{g}_0^*(X_1)\right)\right]\Big\}.
\end{aligned}
$$

Now

$$
\begin{aligned}
& E\left[\left(g(X_1,X_2,h_m)-\bar{g}_0^*(X_1)\right)^2\right] \\
=\ & E\left[g^2(X_1,X_2,h_m)-2g(X_1,X_2,h_m)\bar{g}_0^*(X_1)+\bar{g}_0^{*2}(X_1)\right] \\
=\ & E\left[g^2(X_1,X_2,h_m)\right]-2\cdot\lim_{h_m\to 0}E\left[g(X_1,X_2,h_m)\cdot\frac{1}{n}\sum_{j=1}^{n}g(X_1,X_j,h_m)\right] \\
& +\lim_{h_m\to 0}E\left[\frac{1}{n^2}\sum_{j,k}g(X_1,X_j,h_m)g(X_1,X_k,h_m)\right] \\
=\ & E\left[g^2(X_1,X_2,h_m)\right]-2\cdot\lim_{h_m\to 0}E\left[g(X_1,X_2,h_m)\cdot\frac{1}{n}\sum_{j=1}^{n}E\left[g(X_1,X_j,h_m)|X_j\right]\right] \\
& +\lim_{h_m\to 0}E\left[\frac{1}{n^2}\sum_{j,k}E\left[g(X_1,X_j,h_m)g(X_1,X_k,h_m)|X_1\right]\right] \\
=\ & E\left[g^2(X_1,X_2,h_m)\right]-2\cdot\lim_{h_m\to 0}E\left[g(X_1,X_2,h_m)\bar{g}(X_1,h_m)\right]+\lim_{h_m\to 0}E\left[\bar{g}^2(X_1,h_m)\right] \\
=\ & E\left[g^2(X_1,X_2,h_m)\right]-2\cdot\lim_{h_m\to 0}E\left[\bar{g}(X_1,h_m)E\left[g(X_1,X_2,h_m)|X_1\right]\right] \\
& +E\left[\lim_{h_m\to 0}\bar{g}^2(X_1,h_m)\right] \\
=\ & E\left[g^2(X_1,X_2,h_m)\right]-2\cdot\lim_{h_m\to 0}E\left[\bar{g}^2(X_1,h_m)\right]+E\left[\bar{g}_0^2(X_1)\right] \\
=\ & E\left[g^2(X_1,X_2,h_m)\right]-E\left[\bar{g}_0^2(X_1)\right] \\
=\ & E\left[q(X_1)\right]h_m^{-\gamma}+o\left(h_m^{-\gamma}\right)-Var\left[\bar{g}(X_1,h_m)\right]+2Cov\left[\bar{g}_0(X_1),s(X_1)\right]h_m^{\alpha}+o\left(h_m^{\alpha}\right) \\
=\ & O\left(h_m^{-\gamma}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
& E\left[\left(g(X_1,X_2,h_m)-\bar{g}_0^*(X_1)\right)\left(g(X_1,X_3,h_m)-\bar{g}_0^*(X_1)\right)\right] \\
=\ & E\left[\left(E\left[\left(g(X_1,X_2,h_m)-\bar{g}_0^*(X_1)\right)|X_1\right]\right)^2\right] \\
=\ & E\left[\left(\bar{g}(X_1,h_m)-\bar{g}_0^*(X_1)\right)^2\right] \\
=\ & E\left[\bar{g}^2(X_1,h_m)-2\bar{g}_0^*(X_1)\bar{g}(X_1,h_m)+\bar{g}_0^{*2}(X_1)\right] \\
=\ & E\left[\bar{g}^2(X_1,h_m)\right]-2\cdot\lim_{h_m\to 0}E\left[\bar{g}(X_1,h_m)\cdot\frac{1}{n}\sum_{j=1}^{n}g(X_1,X_j,h_m)\right]
\end{aligned}
$$

$$
\begin{aligned}
& + \lim_{h_m \to 0} E\left[\frac{1}{n^2}\sum_{j,k} g(X_1, X_j, h_m) g(X_1, X_k, h_m)\right] \\
= \;& E\left[\bar{g}^2(X_1, h_m)\right] - 2 \cdot \lim_{h_m \to 0} E\left[\bar{g}^2(X_1, h_m)\right] + \lim_{h_m \to 0} E\left[\bar{g}^2(X_1, h_m)\right] \\
= \;& E\left[\bar{g}^2(X_1, h_m)\right] - E\left[\bar{g}_0^2(X_1)\right] \\
= \;& Var\left[\bar{g}(X_1, h_m)\right] + \theta^2(h_m) - Var\left[\bar{g}_0(X_1)\right] - \theta_0^2 \\
= \;& 2Cov\left[\bar{g}_0(X_1), s(X_1)\right] h_m^\alpha + o(h_m^\alpha) + \left(\theta(h_m) + \theta_0\right)\left(\theta(h_m) - \theta_0\right) \\
= \;& O\left(h_m^\alpha\right).
\end{aligned}
$$

Therefore

$$
\begin{aligned}
& \frac{1}{n^2}\left\{nE\left[\left(g(X_1, X_2, h_m) - \bar{g}_0^*(X_1)\right)^2\right]\right. \\
& \left. + n(n-1)E\left[\left(g(X_1, X_2, h_m) - \bar{g}_0^*(X_1)\right)\left(g(X_1, X_3, h_m) - \bar{g}_0^*(X_1)\right)\right]\right\} \\
= \;& O\left(n^{-1} h_m^{-\gamma}\right) + \frac{n-1}{n} \cdot O\left(h_m^\alpha\right) \\
= \;& o\left(\frac{m}{n}\right) + o\left(m^{-\frac{1}{2}}\right) + o\left(n^{-1} m^{-\frac{1}{2}}\right) \\
= \;& o\left(\frac{m}{n}\right) \\
= \;& o\left(n^{1-\delta-1}\right) \\
= \;& o\left(n^{-\delta}\right).
\end{aligned}
$$

It follows that $n^{\frac{\delta}{2}}\left(\bar{g}^*(X_i^*, h_m) - \bar{g}_0^*(X_i^*) - (\theta(h_m) - \theta_0)\right) = o_p(1)$.

Table 1: Bandwidth estimates for average density estimation, $n = 50$

	mean	s.d.
$h_{n,opt}$	.3170	—
$\tilde{h}_{n,\tau_1}$	.4802	.5301
$\tilde{h}_{n,\tau_2}$	.3857	.2090
$\tilde{h}_{n,\tau_3}$	.3670	.1010
$\hat{h}_{n,m_1}$	.4491	.0531
$\hat{h}_{n,m_2}$	.4265	.0456
$\hat{h}_{n,m_3}$	.4670	.0580

Notes:

1. Simulated random samples each of size $n = 50$ were drawn from a standard normal distribution. The estimates of the average densities were constructed using standard normal smoothing kernels. Reported means and standard deviations (s.d.) are based on 100 Monte Carlo replications.

2. $h_{n,opt}$ refers to the approximation of the true MSE-optimal bandwidth for $n = 50$ given in the statement of Powell and Stoker (1996, Proposition 4.1).

3. $\tilde{h}_{n,\tau_j}$, for $j = 1, 2, 3$ refer to three implementations of the plug-in bandwidth estimator suggested by Powell and Stoker (1996, eq. (4.35)) with pilot bandwidth h_0 equal to $h_{n,opt}$ and three settings of the tuning parameter τ—in particular, $\tau_1 = 1.1$, $\tau_2 = \frac{2}{3} \times \tau_1$ and $\tau_3 = \frac{3}{2} \times \tau_1$. The setting $\tau_1 = 1.1$ was settled upon after some experimentation by the author.

4. $\hat{h}_{n,m_j}$, for $j = 1, 2, 3$ refer to three implementations of the m-bootstrap bandwidth estimator given above in (31) with corresponding resample sizes $m_1 = \lfloor \sqrt{n} \rfloor$, $m_2 = \left\lfloor \frac{2}{3} \times m_1 \right\rfloor$ and $m_3 = \left\lfloor \frac{3}{2} \times m_1 \right\rfloor$, where $\lfloor x \rfloor$ indicates the largest integer less than or equal to x. In each of these implementations, $MSE\left[\hat{\theta}^*_m(h_m) \middle| \mathcal{X}_n\right]$ was minimized for $h_m = km^{-\frac{2}{5}}$ by choice of k over an evenly spaced grid of 100 points covering the interval $[0.01, 3.00]$. In each of these implementations, the pilot bandwidth h_n used to construct $MSE\left[\hat{\theta}^*_m(h_m) \middle| \mathcal{X}_n\right]$ was taken to be $\tilde{h}_{n,\tau_1}$ as described above.

Table 2: Estimates of average densities, $n = 50$

	bias	s.d.	MSE
$\hat{\theta}_n\left(h_{n,opt}\right)$	-.0041	.0320	.0010
$\hat{\theta}_n\left(\tilde{h}_{n,\tau_1}\right)$	-.0056	.0289	.0009
$\hat{\theta}_n\left(\tilde{h}_{n,\tau_2}\right)$	-.0115	.0093	.0002
$\hat{\theta}_n\left(\tilde{h}_{n,\tau_3}\right)$	-.0468	.0040	.0022
$\hat{\theta}_n\left(\hat{h}_{n,m_1}\right)$	-.0038	.0043	.0000
$\hat{\theta}_n\left(\hat{h}_{n,m_2}\right)$	-.0020	.0037	.0000
$\hat{\theta}_n\left(\hat{h}_{n,m_3}\right)$	-.0052	.0047	.0000

Notes:

1. Simulated random samples each of size $n = 50$ were drawn from a standard normal distribution. As such, the object of estimation is $\theta_0 \equiv E\left[f(X_1)\right] = \frac{1}{2\sqrt{\pi}} \approx .2821$.
2. Simulated biases, standard deviations (s.d.) and mean squared errors (MSE) were based on 100 Monte Carlo replications. The estimates of the average densities were constructed using standard normal smoothing kernels.
3. $h_{n,opt}$ and $\tilde{h}_{n,\tau_j}$, $\hat{h}_{n,m_j}$ for $j = 1, 2, 3$ are as described in the notes to Table 1.

References

Chung, K. L. (2001) *A Course in Probability Theory,* third ed. (San Diego: Academic Press)

Goldstein, L., and K. Messer (1992) 'Optimal plug-in estimators for nonparametric functional estimation.' *Annals of Statistics* 20, 1306–1328

Hall, P., and J. S. Marron (1987) 'Estimation of integrated squared density derivatives.' *Statistics and Probability Letters* 6, 109–115

Härdle, W., and A. B. Tsybakov (1993) 'How sensitive are average derivatives?' *Journal of Econometrics* 58, 31–48

Härdle, W., and T. M. Stoker (1989) 'Investigating smooth multiple regression by the method of average derivatives.' *Journal of the American Statistical Association* 84, 986–995

Horowitz, J. L. (1998) *Semiparametric Methods in Econometrics* (New York: Springer-Verlag)

Jaeckel, L. A. (1972) 'Estimating regression coefficients by minimizing the dispersion of the residuals.' *Annals of Mathematical Statistics* 43, 1449–1458

Jones, M. C., and S. J. Sheather (1991) 'Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives.' *Statistics and Probability Letters* 11, 511–514

Jurečková, J. (1971) 'Nonparametric estimate of regression coefficients.' *Annals of Mathematical Statistics* 42, 1328–1338

Nishiyama, Y., and P. M. Robinson (2005) 'The bootstrap and Edgeworth correction for semiparametric averaged derivatives.' *Econometrica* 73, 903–948

Petrov, V. V. (1975) *Sums of Independent Random Variables* (New York: Springer-Verlag)

Powell, J. L. (1994) 'Estimation of semiparametric models.' In *Handbook of Econometrics,* ed. R. F. Engle and D. L. McFadden, vol. 4 (Amsterdam: Elsevier) pp. 2443–2521

Powell, J. L., and T. M. Stoker (1996) 'Optimal bandwidth choice for density-weighted averages.' *Journal of Econometrics* 75, 291–316

Powell, J. L., J. H. Stock, and T. M. Stoker (1989) 'Semiparametric estimation of index coefficients.' *Econometrica* 57, 1403–1430

Robinson, P. M. (1988) 'Root-n-consistent semiparametric regression.' *Econometrica* 56, 931–954

Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics* (New York: Wiley)