

# Bootstrap confidence intervals in nonparametric regression with built-in bias correction

Timothy L. McMurry\*  
*DePaul University*

Dimitris N. Politis†  
*University of California, San Diego*

January 15, 2008

## Abstract

The problem of estimating nonparametric regression with associated confidence intervals is addressed. It is shown that through appropriate choice of infinite order kernel, it is possible to construct bootstrap confidence intervals which do not require either explicit bias correction or suboptimal levels of smoothing at any stage of the estimation. In particular, it is demonstrated that in this setting, consistent estimates are obtained when both the pilot and final smoothings are estimated at the mean square error optimal bandwidth for estimating the regression. The effectiveness of the method is demonstrated through a small simulation study.

## 1 Introduction

This article explores the usefulness of infinite order kernels in constructing bootstrap confidence intervals for nonparametric regressions. In regression, pairs of data  $(X_1, Y_1), \dots, (X_n, Y_n)$  are observed, and the objective is to estimate  $r(x) := E[Y|X = x]$ . Nonparametric techniques allow  $r$  to be estimated without imposing strong assumptions on its underlying form. Unfortunately, this flexibility can make it difficult to ascertain whether observed features are physically meaningful or the result of random variability in the data. Confidence intervals provide a tool to assess the impact of this variability on the estimate.

The bootstrap approach to constructing confidence intervals for kernel regression estimates has been studied by several authors including Härdle and Bowman [8], Härdle and Marron [9], Neumann and Polzehl [13], and Claeskens and van Keilegom [2]. In addition, Yang and Wahba [19] investigated a related problem in spline regression. Although implementations differ, broadly speaking the bootstrap procedure is as follows: first, a pilot regression is used to calculate the residuals, which approximate the true unobserved errors; second, the residuals are resampled, creating ‘new’ errors; third, the resampled errors are added to an estimate of the regression function, creating a new bootstrap data set; finally, the bootstrap data set is smoothed. An estimate for the likely range of the true unknown function is generated by repeating this process several times. Unfortunately, all nonparametric regression estimates, including both the pilot and final smoothings are biased. Unless this bias is adequately accounted for, the resulting confidence intervals will not achieve the expected level of coverage.

---

\*Research partially supported by a DePaul University Faculty Summer Research Grant.

†Research partially supported by NSF grant SES-04-18136 funded jointly by the Economics and Statistics Divisions of NSF.

Previous authors have used various tricks to account for this bias. Härdle and Bowman [8] construct a kernel estimate for the second derivative,  $r''(x)$ , and use this estimate to explicitly adjust for the bias; the estimate of the second derivative is known to be consistent, but it is difficult to choose the bandwidth. Härdle and Marron [9] estimate the residuals using the optimal bandwidth, but the resampled residuals are then added to an oversmoothed estimate of  $r$ ; they then smooth the bootstrapped data using the optimal bandwidth. Neumann and Polzehl [13] use only one bandwidth, but it is smaller than the mean square error optimal. Yang and Wahba [19] avoid bias correction by studying average coverage probability across the function rather than pointwise intervals. Each of the pointwise techniques requires one bandwidth for which a natural estimate does not exist.

In this article we show that in the setting studied by Härdle and Bowman [8], the need for explicit bias correction and the associated suboptimal bandwidths can be overcome through the use of the infinite order kernel regression estimator described in McMurry and Politis [11]. The infinite order estimator automatically adapts to whatever degree of smoothness is present in the underlying function, resulting in estimates with such rapid rates of convergence that explicit correction becomes unnecessary.

The remainder of the paper is organized as follows: Section 2 contains background on the infinite order kernel estimator; Section 3 outlines the procedure, technical details, assumptions, and presents the main theorem; Section 4 contains simulation results; technical proofs have been placed in Section 5.

## 2 Description of the Estimator

As in Härdle and Bowman [8], we address fixed design regression, here using the Gasser-Müller [6] estimator with the class of kernels introduced in McMurry and Politis [11]. To summarize briefly, we assume the data are of the form  $(x_1, Y_1), \dots, (x_n, Y_n)$ , where  $Y_i = r(x_i) + \epsilon_i$ , and where  $\epsilon_1, \dots, \epsilon_n$  are errors with mean 0 and variance  $\sigma^2$ . In addition, for the sake of simplicity, we assume the data has been ordered and scaled so that  $0 < x_1 < x_2 < \dots < x_n < 1$ . The Gasser-Müller estimator is then defined by

$$\hat{r}_h(x) = \frac{1}{h} \sum_{i=1}^n Y_i \int_{s_{i-1}}^{s_i} K\left(\frac{x-u}{h}\right) du, \quad (1)$$

where  $s_0 = 0$ ,  $s_n = 1$ , and  $s_i = (x_i + x_{i+1})/2$  for  $i = 1, \dots, n-1$ . In the above,  $K$  is the kernel function, which determines the weights given to various observations, and  $h$  is the bandwidth parameter, which balances a trade-off between oversmoothing and undersmoothing.

The kernel  $K$  is assumed to be in a subset of the class of general infinite order flat-top kernels defined in Politis [14]. In particular, it is assumed to satisfy the following definition.

**Definition 1** *A general infinite order flat-top kernel  $K$  is defined in terms of its Fourier transform  $\lambda$ , which in turn is defined as follows. Fix a constant  $c > 0$ . Let*

$$\lambda(s) = \begin{cases} 1 & \text{if } |s| \leq c \\ g(|s|) & \text{if } |s| > c, \end{cases}$$

where the function  $g$  is chosen to make  $\lambda(s)$  and  $s\lambda(s)$  integrable, and to make  $\lambda(s)$  infinitely differentiable for all  $s$ . The flat top kernel is now given by

$$K(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lambda(s) e^{-isx} ds, \quad (2)$$

i.e., the inverse Fourier transform of  $\lambda(s)$ .

An example of a kernel satisfying Definition 1 is given in Section 4. Related kernels have been investigated for use in density estimation [3, 4, 5, 16, 18], spectral density estimation [17], and time series autoregression [12].

In addition, we impose some restrictions on the model which generated the data. First, the design points are roughly assumed to be given by quantiles of a reasonably behaved density function; this allows uneven spacing of design points while requiring that the distance between adjacent points decreases at a rate proportional to  $1/n$ . We also make some requirements of the errors and the underlying function  $r$ . These constraints are made precise in the following assumptions.

**Assumption 1** *The design points are given by*

$$x_i = Q\left(\frac{i-1/2}{n}\right), \quad i = 1, \dots, n,$$

where

$$Q(u) = F^{-1}(u),$$

and

$$F(x) = \int_0^x f(t)dt.$$

The function  $f$  is assumed to be a Lipschitz continuous positive density function on  $[0, 1]$ .

**Assumption 2** *The errors  $\epsilon_i$ ,  $i = 1, \dots, n$  are independent and identically distributed with mean 0, and variance  $\sigma^2$ .*

**Assumption 3** *The unknown function  $r$  has at least one bounded continuous derivative on  $(0, 1)$ .*

Under these assumptions, the asymptotic performance of the estimator is established in [11] where the convergence rates were shown to be given by the following theorem restated here in sharpened form.<sup>1</sup>

**Theorem 1** *(McMurry and Politis) If  $r$  has  $k$  continuous derivatives, then under Assumptions 1–3, the asymptotic variance and bias of the infinite order kernel estimator,  $\hat{r}_h(x)$ , as  $n \rightarrow \infty$ ,  $h \rightarrow 0$ , and  $n^2h^3 \rightarrow \infty$  are given by*

$$\text{var}[\hat{r}_h(x)] = \frac{\sigma^2}{nh} \frac{1}{f(x)} \int_{-\infty}^{\infty} K^2(z)dz + O\left(\frac{1}{n}\right) + O\left(\frac{1}{n^2h^3}\right),$$

and

$$\text{bias}[\hat{r}_h(x)] = E[\hat{r}_h(x)] - r(x) = o(h^k) + O(1/n).$$

If  $r$  is infinitely differentiable, the bias becomes  $o(h^m) + O(1/n)$  for all positive real  $m$ .

If  $r$  has  $k$  continuous derivatives, it can easily be seen from the above theorem that choosing  $h$  proportional to  $n^{-1/(2k+1)}$  makes the above estimator converge at the mean square optimal rate.

**Remark:** The bias is likely dominated by the  $o(h^k)$  term, even when  $r$  is very smooth. The  $o(h^k)$  term embodies two systemic inaccuracies. The first is the convolution  $\int r(x-hv)K(v)dv - r(x)$  which is typical in kernel smoothing, and the second results because the tails of the kernel extend beyond the design interval. Even if the first portion were to vanish exactly, the second will persist. Any

---

<sup>1</sup>The bias given in [11] is  $O(h^k)$ ; however, without modification the same proof shows that the bias is actually  $o(h^k)$ .

kernel  $K(x)$  satisfying Definition 1, and in particular the kernel defined by equation (3), has non-zero tails that vanish faster than  $x^{-k}$  for any  $k > 0$ . If, for example, the tails vanish exponentially fast, then the resulting bias contributed by the right tail is  $|h^{-1} \int_1^\infty K((x-z)/h) dz| \leq c_1 \exp[-c_2/h]$ , for positive constants  $c_1$  and  $c_2$ , which when roughly balanced with variance leads to a bandwidth of approximately  $(\log n)^{-1}$ . If the constants are chosen properly, the variance becomes  $O(\log n/n)$  and the bias becomes  $O(1/\sqrt{n})$ ; these rates agree with both the standard parametric rate of convergence and the limiting rate for finite smoothness  $n^{-k/(2k+1)}$  as  $k$  becomes large.

**Remark:** The automatic bias reduction evident in Theorem 1 does not affect the rate at which the variance converges, although it likely makes the leading constant somewhat larger; this can be seen in the simulations in [11], where the integrated variance was slightly larger for the infinite order estimators than for traditional second order estimators.

### 3 Bootstrap Confidence Bands

The proposed bootstrap algorithm is as follows:

1. Perform a pilot estimate on the data using the estimator given in equation (1), optimal bandwidth  $h$ , and kernel satisfying Definition 1; this generates an initial estimate  $\hat{r}_h(x_i)$ , for  $i = 1, 2, \dots, n$ .
2. Calculate the residuals by  $\hat{\epsilon}_i := Y_i - \hat{r}_h(x_i)$ .
3. Discard the residuals from the intervals  $x_i < \eta$  and  $x_i > 1 - \eta$ , where  $0 < \eta < 1/2$ ; this is done to reduce the impact of the bias caused by the edges of the interval. While asymptotic results remain valid for any such fixed value of  $\eta$ , in practice one would want to choose  $\eta$  only so large as to minimize boundary effects. A good rule of thumb is to ensure that all retained residuals are from points estimated using at least the first negative side-lobe on either side of the kernel (see Figure 1). Let  $n_s$  be the number of residuals that are retained, and re-order/re-name the latter to be denoted by  $\hat{\epsilon}_i, i = 1, \dots, n_s$ .
4. Center the remaining residuals giving  $\tilde{\epsilon}_i = \hat{\epsilon}_i - \frac{1}{n_s} \sum_j \hat{\epsilon}_j$  for  $i = 1, \dots, n_s$ .
5. Generate the  $\epsilon_i^*$  by sampling with replacement from  $\tilde{\epsilon}_1, \dots, \tilde{\epsilon}_{n_s}$ , and generate new data  $Y_i^* = \hat{r}_h(x_i) + \epsilon_i^*$ .
6. Smooth the data  $(x_1, Y_1^*), \dots, (x_n, Y_n^*)$  again using the optimal bandwidth  $h$ , resulting in the bootstrap curve  $\hat{r}_h^*(x)$ .

Under conditions which will be made precise in the following assumption and theorem, the bootstrap distribution of  $\hat{r}_h^*(x_i)$  about its mean, which can be estimated via the preceding algorithm, asymptotically approximates the distribution of  $\hat{r}_h(x)$  about  $r(x)$ .

**Assumption 4** *The bandwidth  $h$  decreases at a rate proportional to  $n^{-1/(2k+1)}$  when  $r$  has  $k$  continuous derivatives; this is the mean square error optimal rate.*

**Theorem 2** *Under Assumptions 1–4, the distribution of the bootstrapped estimator about its mean converges to the distribution of  $\hat{r}_h(x)$  about  $r(x)$  in the Mallows metric.*

$$d_2 \left( \sqrt{nh}[\hat{r}_h(x) - r(x)], \sqrt{nh}[r_h^*(x) - E^* \hat{r}_h^*(x)] \right) \rightarrow_P 0.$$

Since  $r_h^*(x)$  can be simulated, and since

$$E^* \hat{r}_h^*(x) = \sum_{i=1}^n \hat{r}_h(x_i) \int_{s_{i-1}}^{s_i} K_h(x-u) du$$

$$= \sum_{i=1}^n \sum_{j=1}^n Y_j \int_{s_{j-1}}^{s_j} K_h(x_i - u) du \int_{s_{i-1}}^{s_i} K_h(x - u) du,$$

the bootstrap distribution can be estimated. Theorem 2 then gives a consistent estimate of the distribution of  $\hat{r}_h(x)$  that can be immediately used to construct confidence intervals for  $\hat{r}(x)$ .

Note that Theorem 2 assumes that all of the pilot estimate used to obtain the residuals, the estimate of the regression to which the bootstrapped residuals are added, and the final smoothing of the bootstrapped data are performed at the mean square optimal bandwidth. In addition, Theorem 2 requires no explicit bias correction. As such, the entire procedure can be performed using only one bandwidth, which can be estimated from the data by a technique such as the rule of thumb described in [11] or cross-validation. Recall that the techniques in [2, 8, 9, 13] require at least one level of smoothing for which there is not a natural estimate.

**Remark:** It is crucial that  $h$  decreases at a rate at least as fast as specified in Assumption 4 in order for the bias correction that is implicitly built-in in Theorem 2 to take effect. To elaborate, if  $h = O(n^{-1/(2k+1)})$  when  $r$  has  $k$  continuous derivatives, then  $\text{bias}(\hat{r}_h(x)) = E[\hat{r}_h(x)] - r(x) = o(h^k)$  uniformly on closed subsets of  $(0, 1)$  from Theorem 1. Consequently,  $\sqrt{nh} \text{bias}(\hat{r}_h(x)) = o(1)$ , which makes the bias of the distribution of  $\sqrt{nh}[\hat{r}_h(x) - r(x)]$  asymptotically negligible.

## 4 Simulations

A small simulation study was undertaken to explore whether or not the proposed confidence intervals have approximately the desired coverage. For each of the simulations, the kernel used was defined by equation (2), along with

$$\lambda(s) = \begin{cases} 1 & \text{for } |s| \leq 0.05 \\ \exp[-0.25 \exp[-0.25/(|s| - 0.05)^2]/(|s| - 1)^2] & \text{for } 0.05 < |s| < 1 \\ 0 & \text{for } |s| \geq 1. \end{cases} \quad (3)$$

A graph of this kernel is shown in Figure 1.

In both simulations, there were 100 evenly spaced design points, and 33 evenly spaced point-wise 95% confidence intervals were constructed in the interval  $[0.1, 0.9]$ . The constant  $\eta$ , which determines which residuals are discarded due to boundary bias, was chosen to be 0.15, which is approximately the distance from the center of the kernel to the end of the first negative side lobe (see Figure 1) at the chosen bandwidths. Bandwidths were chosen by the rule of thumb suggested in [11], and developed further in [15]. Each experiment was performed 1000 times, with 10,000 bootstrap replications each.

**SIMULATION I:** As in Härdle and Bowman [8], data was simulated from the model  $Y_i = \sin(4\pi x_i) + \epsilon_i$  with two different error distributions for the  $\epsilon$  (normal and translated exponential). Simulated coverage is shown in Table 1, and a regression with associated confidence intervals is shown in Figure 2. It should be noted that the infinite order kernel technique seems to perform slightly better than the explicit bias correction used in Härdle and Bowman [8] where they observed coverage of 86-87% for bootstrap confidence intervals ([8], p. 107).

**SIMULATION II:** For comparison, we conducted a second simulation study using a different regression function. As in [7, 11], data was simulated from the model  $r(x) = 2 - 2x + 3 \exp[-100(x - .5)^2] + \epsilon$ . Once again, errors were simulated using three different error variances and both normal and translated exponential errors. Simulated coverage is shown in Table 1. Coverage is similar to the coverage observed in the first simulation.

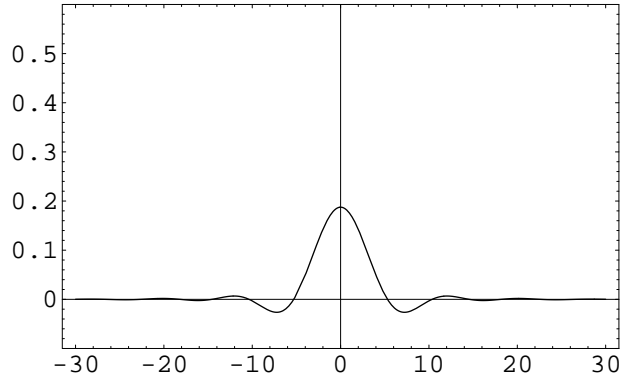


Figure 1: An example of an infinite order kernel.

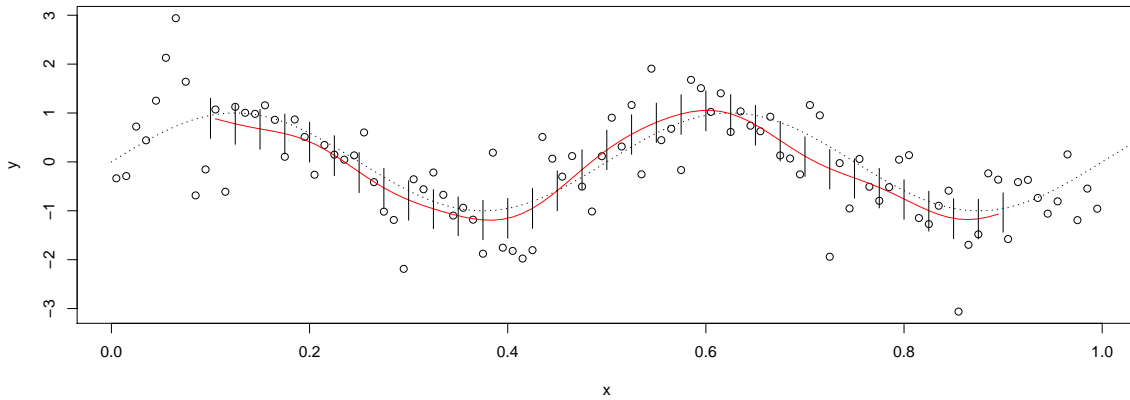


Figure 2: Simulation with  $r(x) = \sin(4\pi x)$  and  $\epsilon_i$  distributed normally with standard deviation 0.7. The dots are the data, the solid line is the estimate, the dashed line is the true function, and the vertical bars represent our bootstrap (pointwise) 95% confidence intervals.

	$\sigma$	Coverage	
		Normal Errors	Translated Exponential
Simulation 1	0.1	0.926	0.915
	0.4	0.928	0.920
	0.7	0.927	0.918
Simulation 2	0.1	0.910	0.901
	0.4	0.922	0.915
	0.7	0.927	0.919

Table 1: Empirical coverage of bootstrap confidence intervals for the two simulations based on Theorem 2 with nominal level 95%.

## 5 Technical Proofs

PROOF OF THEOREM 1: Since the sole difference between Theorem 1 in this paper and Theorems 1 and 2 in [11] lies in the last line of the bias calculation, only enough detail to justify the change is shown here. The remaining details can be found in [11]; in addition, similar results are derived in [6, 10]. In the following, let  $\check{f}(s)$  denote the fourier transform of a function  $f(x)$ .

$$\begin{aligned} |\text{bias}(\hat{r}_h(x))| &= \left| \frac{1}{h} \sum_{i=1}^n r(x_i) \int_{s_{i-1}}^{s_i} K\left(\frac{x-u}{h}\right) du - r(x) \right| \\ &\leq \left| \int_{-\infty}^{\infty} r(u) K_h(x-u) du - r(x) \right| + O(1/n) + o(h^m), \end{aligned}$$

where the last term is  $o(h^m)$  for all positive  $m$ .

$$\begin{aligned} \left| \int_{-\infty}^{\infty} r(u) K_h(x-u) du - r(x) \right| &= \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} \check{r}(s) \check{K}_h(s) e^{-isx} ds - \frac{1}{2\pi} \int_{-\infty}^{\infty} \check{r}(s) e^{-isx} ds \right| \\ &= \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} (\lambda(hs) - 1) \check{r}(s) e^{-isx} ds \right| \\ &\leq \frac{1}{2\pi} \int_{|s| > c/h} \frac{s^k}{s^k} |\check{r}(s)| ds \\ &\leq \frac{1}{2\pi} \left(\frac{h}{c}\right)^k \int_{|s| > c/h} s^k |\check{r}(s)| ds \\ &= o(h^k). \end{aligned}$$

The final term is  $o(h^k)$  rather than  $O(h^k)$  because the integral also converges to 0 as  $h \rightarrow 0$ .  $\square$

Before proving Theorem 2, we first state and prove a preliminary result that was also mentioned without proof by Härdle and Bowman [8].

**Lemma 1** For random variables  $U$  and  $V$ ,

$$d_2(U, V)^2 = d_2(U, V - E[V])^2 + E[U - V]^2 - E[U]^2.$$

PROOF: Let  $a := E[U]$  and  $b := E[V]$ . Then,

$$E[|U - (V - b)|^2] = E[|U - V|^2] + 2bE[(U - V)] + b^2.$$

Therefore,

$$d_2(U, V - b)^2 \leq E[|U - V|^2] + 2bE[(U - V)] + b^2.$$

So, if we choose  $U$  and  $V$  such that  $E[|U - V|^2] = d_2(U, V)^2$ , then we have

$$d_2(U, V - b)^2 \leq d_2(U, V)^2 + 2bE[(U - V)] + b^2.$$

Now choose  $U$  and  $V$  such that

$$E[|U - (V - b)|^2] = d_2(U, V - b)^2.$$

Then,

$$\begin{aligned} d_2(U, V - b)^2 &= E[|U - V|^2] + 2bE[(U - V)] + b^2 \\ &\geq d_2(U, V)^2 + 2bE[(U - V)] + b^2. \end{aligned}$$

Since  $E[U - V]^2 - E[U]^2 = 2bE[(U - V)] + b^2$ , this establishes the desired result.  $\square$

PROOF OF THEOREM 2: By Lemma 8.8 of Bickel and Freedman [1],

$$\begin{aligned} &d_2\left(\sqrt{nh}[\hat{r}_h(x) - r(x)], \sqrt{nh}[r_h^*(x) - E^*\hat{r}_h^*(x)]\right) \\ &= d_2\left(\sqrt{nh}[\hat{r}_h(x) - E\hat{r}_h(x)], \sqrt{nh}[r_h^*(x) - E^*\hat{r}_h^*(x)]\right) \\ &\quad - nh[E\hat{r}_h(x) - r(x)]^2 \end{aligned} \tag{4}$$

The first term is essentially a variance term, and the second is essentially a squared bias term. From our Theorem 1, the bias is of order  $o(h^k) + O(1/n)$ , thus, the second term above is uniformly  $nh(o(h^{2k}) + O(1/n^2) + o(h^k/n))$  on  $[\eta, 1 - \eta]$ . So, if  $h \sim n^{-1/(2k+1)}$ , the bias is  $o(1)$ .

We now focus on the variance term. Let  $F$ ,  $F_n$ ,  $\hat{F}_n$ , and  $\tilde{F}_n$  denote respectively the distribution functions of the errors, and the empirical distributions of the errors, the residuals, and the mean centered residuals. By Lemma 8.9 in [1],

$$d_2\left(\sqrt{nh}[\hat{r}_h(x) - E\hat{r}_h(x)], \sqrt{nh}[r_h^*(x) - E^*\hat{r}_h^*(x)]\right) \leq nh \sum_i w_i^2(x) d_2(F, \tilde{F}_n)^2,$$

where  $w_i = \int_{s_{i-1}}^{s_i} K_h(x - u) du$ . By the triangle inequality,

$$d_2(F, \tilde{F}_n)^2 \leq 2 \left[ d_2(F, F_n)^2 + d_2(F_n, \tilde{F}_n)^2 \right].$$

Then, by lemma 8.4 of Bickel and Freedman [1],  $d_2(F, F_n) \rightarrow_P 0$ .

For the second half of the result we apply Lemma 1 above to get:

$$d_2(F_n, \tilde{F}_n)^2 = d_2(F_n, \hat{F}_n)^2 - \left[ \frac{1}{n_s} \sum_i (\epsilon_i - \hat{\epsilon}_i) \right]^2 + \left[ \frac{1}{n_s} \sum_i \epsilon_i \right]^2.$$

So,

$$E \left[ d_2(F_n, \tilde{F}_n)^2 \right] \leq E \left[ d_2(F_n, \hat{F}_n)^2 \right] + \frac{\sigma^2}{n_s}.$$

By the definition of Mallow's metric, we may consider the joint distribution of  $\epsilon_i$  and  $\hat{\epsilon}_i$  which puts probability mass  $1/n_s$  at each pair  $(\epsilon_i, \hat{\epsilon}_i)$ . Therefore,

$$\begin{aligned} E \left[ d_2(F_n, \hat{F}_n)^2 \right] &\leq E \left[ \frac{1}{n_s} \sum_i (\epsilon_i - \hat{\epsilon}_i)^2 \right] \\ &= \frac{1}{n_s} \sum_i \text{MSE}(\hat{r}_h(x_i)) \rightarrow 0 \end{aligned}$$

and the theorem is proven.  $\square$

## References

- [1] P. Bickel and D. Freedman. Some asymptotic theory for the bootstrap. *Annals of Statistics*, 9(6):1196–1217, 1981.
- [2] G. Claeskens and I. Van Keilegom. Bootstrap confidence bands for regression curves and their derivatives. *Annals of Statistics*, 31(6):1852–1884, 2003.
- [3] K. Davis. Mean square error properties of density estimates. *Annals of Statistics*, 3(4):1025–1030, 1975.
- [4] K. Davis. Mean integrated square error properties of density estimates. *Annals of Statistics*, 5(3):530–535, 1977.
- [5] L. Devroye. A note on the usefulness of superkernels in density estimates. *Annals of Statistics*, 20(4):2037–2056, 1992.
- [6] Th. Gasser and H.-G. Müller. Kernel estimation of regression functions. In Th. Gasser and M. Rosenblatt, editors, *Smoothing Techniques for Curve Estimation*, number 757 in Springer Lecture Notes in Mathematics, pages 23–68. Springer-Verlag, Berlin, 1979.
- [7] Th. Gasser and H.-G. Müller. Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, 11:171–185, 1984.
- [8] W. Härdle and A. W. Bowman. Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands. *Journal of the American Statistical Association*, 83(401):102–110, 1988.
- [9] W. Härdle and J. S. Marron. Bootstrap simultaneous error bars for nonparametric regression. *Annals of Statistics*, 19(2):778–796, 1991.
- [10] J. D. Hart. *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer, New York, 1997.
- [11] T. McMurry and D. N. Politis. Nonparametric regression with infinite order flat-top kernels. *Journal of Nonparametric Statistics*, 16(3-4):549–562, 2004.
- [12] T. McMurry and D. N. Politis. Minimally biased nonparametric regression and autoregression. To appear in *RevStat*, 2008.
- [13] M. H. Neumann and J. Polzehl. Simultaneous bootstrap confidence bands in nonparametric regression. *Journal of Nonparametric Statistics*, 9:307–333, 1998.
- [14] D. N. Politis. On nonparametric function estimation with infinite order flat-top kernels. In Ch. A. Charalambides, Markos V. Koutras, and N. Balakrishnan, editors, *Probability and Statistical Models with Applications*, pages 469–483. Chapman & Hall/CRC, Boca Raton, 2001.
- [15] D. N. Politis. Adaptive bandwidth choice. *Journal of Nonparametric Statistics*, 25(4-5):517–533, 2003.
- [16] D. N. Politis and J. P. Romano. On a family of smoothing kernels of infinite order. In M. Tarter and M. Lock, editors, *Computing Science and Statistics, Proceedings of the 25th Symposium on the Interface*, pages 141–145. The Interface Foundation of North America, April 1993.

- [17] D. N. Politis and J. P. Romano. Bias-corrected nonparametric spectral estimation. *Journal of Time Series Analysis*, 16(1):67–103, 1995.
- [18] D. N. Politis and J. P. Romano. Multivariate density estimation with general flat-top kernels of infinite order. *Journal of Multivariate Analysis*, 68:1–25, 1999.
- [19] Y. Wang and G. Wahba. Bootstrap confidence intervals for smoothing splines and their comparison to bayesian confidence intervals. *Journal of Statistical Computation and Simulation*, 50:263–279, 1995.