

Estimating Error Correlation in Nonparametric Regression

by

Naomi S. Altman

BU-1023-MA

June, 1991

Estimating Error Correlation in Nonparametric Regression

Naomi Simone Altman

Biometrics Unit

Cornell University

June, 1991

This work was supported by a Natural Sciences and Engineering Research Council of Canada postgraduate scholarship, and Hatch Grant 151410 NYF. Much of this work was completed while the author was in the Department of Statistics at Stanford University. The author would like to thank Iain Johnstone for his valuable guidance in supervising the dissertation from which part of this article was taken.

SUMMARY

In nonparametric regression, estimates of the error correlation are needed for bandwidth selection and for determining confidence bands around the regression estimate. They may also be of fundamental interest to the investigator. This paper focuses on the problem of estimating the error correlation function from the residuals of a kernel regression analysis. The data are assumed to have an unknown mean which depends deterministically on an independent variable; the errors are generated by a stationary, but correlated, process. Method of moments estimators (MMEs) of correlation are shown to be consistent under mild regularity conditions.

Keywords: kernel smoothing; correlated errors; method of moments; time series; autocorrelation

1. Introduction

Under regularity conditions on the mean and correlation functions, kernel regression has been shown to be consistent for estimating the mean function (Altman, 1990 and Hart 1991) for the model .

$$y_{n,i} = \mu\left(\frac{i}{n}\right) + \varepsilon_i \quad (1)$$

where n is the sample size, $\mu(x)$ is a smooth deterministic mean function on $[0, 1]$, and ε is a stationary second order error process with mean zero and covariance function

$$\text{cov}(\varepsilon_i, \varepsilon_j) = \sigma^2 \rho(|i - j|). \quad (2)$$

It is therefore natural to attempt to estimate functionals of the error distribution from the residuals. This paper focuses on the use of method of moments estimators (MMEs) to estimate the correlation function of the errors.

Using this model for the correlations, the design points become closer together as the sample size increases, but the error process remains the same. This model is discussed by Hart (1991). The model is of practical application in situations in which the correlation is induced by the measuring device, for example, when the output from a monitor is a filtered sequence. In many applications, the correlation function depends on the distance between design points, and thus on n . However, for these applications, asymptotic results based on model (2) can be viewed as approximations valid for large sample sizes.

The regression residuals have the form:

$$e_{\lambda,n}(i) = y_{n,i} - \hat{\mu}_{\lambda,n}\left(\frac{i}{n}\right). \quad (3)$$

The regression estimator, $\hat{\mu}_{\lambda,n}\left(\frac{i}{n}\right)$, is the kernel estimator of Priestley and Chao (1972),

$$\hat{\mu}_{\lambda,n}(x) = \frac{1}{n\lambda} \sum_{j=0}^n K\left(\frac{x - \frac{j}{n}}{\lambda}\right) y_{n,j}. \quad (4)$$

K is the kernel weight function, and λ is the bandwidth.

2. Expectation of Products of Residuals

The MMEs of the correlation function are rational functions of the residuals. In this section, we compute the asymptotic distribution of products of the residuals, and then use the delta method (Serfling 1980, p.122) and the results of Bartlett (1946) to compute the expectation and variance of the correlation estimates.

Theorem 1 shows that, as $\lambda \rightarrow 0$ and $n\lambda \rightarrow \infty$, the empirical covariances based on the residuals are asymptotically normal, and are consistent estimators of the true covariances. Theorem 2 shows that the MMEs of correlation based on the empirical covariances are also asymptotically normal, and are consistent estimators of the true correlations.

Kernels with the following properties are considered:

- A) K is symmetric about 0.
- B) K has support only on the interval $(-\frac{1}{2}, \frac{1}{2})$.
- C) K is Lipschitz continuous of order $\alpha > 0$.

K is called a kernel of order p if all the first $p - 1$ moments of K are 0, and the p^{th} moment,

$$s_K = \int x^p K(x) dx \quad (5)$$

is not zero. The squared norm of K

$$N_K = \int K^2(x) dx \quad (6)$$

is also needed for the computations that follow.

Altman, 1990, showed that kernel estimators of this type are consistent estimators of the mean function under the following regularity conditions:

- D) The mean function μ has square integrable p^{th} derivative ($p \geq 2$) which is Lipschitz of order γ , $0 < \gamma \leq 1$.

The signal to noise ratio,

$$N(\lambda) = \frac{\int_{\frac{\lambda}{2}}^{1-\frac{\lambda}{2}} (\mu^{(p)}(x))^2 dx}{\sigma^2}. \quad (7)$$

plays an important role in the computations that follow.

E) The correlation function is absolutely summable and

$$\sum_{j=1}^{\infty} \rho(j) = S_{\rho}. \quad (8)$$

F) The correlation function satisfies:

$$\frac{1}{N} \sum_{i=1}^N i |\rho(i)| = o(1)$$

Conditions E and F are mixing conditions which ensures that observations sufficiently far apart are essentially uncorrelated.

For the consistency of the MMEs of variance and correlation based on the residuals from the kernel smooth, further regularity conditions on the errors are needed. A condition which is often used in time series analysis (see, for example, Brockwell and Davis, 1987, chap. 7) and which is sufficient for consistency is:

$$\begin{aligned} \text{G) } \varepsilon_{n,t} &= \sum_{j=-\infty}^{\infty} \psi_j z_{t-j} && \text{with } \sum_{j=-\infty}^{\infty} |\psi_j| < \infty && (9) \\ z_t &\text{ i.i.d.} && \text{with } E(z_t) = 0 \\ &&& E(z_t^4) < \infty. \end{aligned}$$

For results on the variance of the MMEs, the following condition is needed on the fourth moment of the error process:

$$\begin{aligned} \text{H) For all } r, s, t \\ \sum_{n=0}^{\infty} E(\varepsilon_t \varepsilon_{t+s} \varepsilon_{t+r} \varepsilon_{t+r+s+n}) &\text{ converges.} \\ \sum_{n=0}^{\infty} E(\varepsilon_t \varepsilon_{t+s} \varepsilon_{t+r} \varepsilon_{t+r-s+n}) &\text{ converges.} \end{aligned}$$

Theorem 1 : Suppose the data are generated by model (1) with error correlations of form (2). Suppose the kernel function K is of order p and satisfies conditions A – C, the mean function, μ , satisfies condition D, and the error process, ε_t , satisfies conditions E–H. Let $[x]$ denote the largest integer less than x .

Define

$$\hat{\gamma}_{\lambda,n}(s) = \frac{1}{n} \sum_{i=\lfloor \frac{n\lambda}{2} \rfloor}^{n-\lfloor \frac{n\lambda}{2} \rfloor-s} e_{\lambda,n}(i) e_{\lambda,n}(i+s). \quad (10)$$

Then, as $\lambda \rightarrow 0$ and $n\lambda \rightarrow \infty$

$$\sqrt{n} \begin{pmatrix} \hat{\gamma}_{\lambda,n}(0) \\ \vdots \\ \hat{\gamma}_{\lambda,n}(k) \end{pmatrix} \sim \text{asymptotically Normal} \left(\begin{pmatrix} \gamma_{\lambda,n}(0) \\ \vdots \\ \gamma_{\lambda,n}(k) \end{pmatrix}, \frac{V}{n} + O\left(\frac{1}{n^2\lambda}\right) \right)$$

where

$$\begin{aligned} \gamma_{\lambda,n}(s) = & \sigma^2 \left(\rho(s) + \lambda^{2p} \left(\frac{sK}{p!} \right)^2 N(\lambda) + \frac{1+S_p}{n\lambda} (N_K - 2K(0)) \right) + \\ & o(\lambda^{2p}) + o\left(\frac{1}{n\lambda}\right) + o\left(\frac{s}{n}\right). \end{aligned}$$

V is the covariance matrix given by Bartlett's formula (Bartlett, 1946) for the process, ε_t .

The proof of Theorem 1 is in the Appendix.

Theorem 2: Suppose the data and kernel satisfy the conditions of Theorem 1 and define $\hat{\gamma}_{\lambda,n}(s)$ by (10). For fixed s , n and λ , define the method of moments estimator of $\rho(s)$ by

$$\hat{\rho}_{\lambda,n}(s) = \frac{\hat{\gamma}_{\lambda,n}(s)}{\hat{\gamma}_{\lambda,n}(0)}. \quad (11)$$

Then $\hat{\rho}_{\lambda,n}(s)$ is asymptotically normal and, as $\lambda \rightarrow 0$ and $n\lambda \rightarrow \infty$,

$$\begin{aligned} E(\hat{\rho}_{\lambda,n}(s)) = & \frac{\rho(s) + \lambda^{2p} \left(\frac{sK}{p!} \right)^2 N(\lambda) + \frac{(1+2S_p)}{n\lambda} (N_K - 2K(0))}{1 + \lambda^{2p} \left(\frac{sK}{p!} \right)^2 N(\lambda) + \frac{(1+2S_p)}{n\lambda} (N_K - 2K(0))} \\ & + o(\lambda^{2p}) + o\left(\frac{1}{n\lambda}\right) + o\left(\frac{s}{n}\right). \end{aligned}$$

and

$$\text{Var}(\hat{\rho}_{\lambda,n}(s)) = \frac{V_s}{n} + o\left(\frac{1}{n^2\lambda^2}\right)$$

where V_s is given by Bartlett's formula for the process ε_t .

Theorem 2 is an easy consequence of Theorem 1, Theorem 3.3.A of Serfling (1980 p.122) and the results of Bartlett (1946).

Theorem 2 shows the consistency of the MMEs of the correlation function. It also shows that, asymptotically, the mean squared error of the correlation estimates is dominated by the bias, which depends on the signal to noise ratio, $N(\lambda)$, and the sum of the correlations, S_ρ .

While the MMEs are consistent for all lags, it is important to note that the coefficients of the higher order terms increase in magnitude with s . In finite samples, correlations at low lags are estimated much more precisely than correlations at longer lags.

Corollary 2, below, discusses the dependence of the estimation bias on $N(\lambda)$ and S_ρ .

Corollary 2: Under the conditions of Theorem 2, asymptotically,

- a) If $N_K \geq 2K(0)$, the bias of $\hat{\rho}_{\lambda,n}(s)$ is positive.
- b) If $N_K < 2K(0)$, $\hat{\rho}_{\lambda,n}(s)$ has bias which is increasing in λ , and the signal to noise ratio, $N(\lambda)$, and is decreasing in S_ρ .

Proof: Let $r_{\lambda,n}(s) = (\rho(s) + C_{\lambda,n}) / (1 + C_{\lambda,n})$ where

$$C_{\lambda,n} = \lambda^{2p} \left(\frac{s_K}{p!} \right)^2 N(\lambda) + \frac{(1 + 2S_\rho)}{n\lambda} (N_K - 2K(0)). \quad (12)$$

Then

$$E(\hat{\rho}_{\lambda,n}(s)) = r_{\lambda,n}(s) + o(\lambda^{2p}) + o\left(\frac{1}{n\lambda}\right).$$

$r_{\lambda,n}(s)$ is a hyperbola in $C_{\lambda,n}$, with asymptote 1, and singularity at $C_{\lambda,n} = -1$. $r_{\lambda,n}(s)$ is increasing in $C_{\lambda,n}$ for $C_{\lambda,n} > -1$. If $N_K \geq 2K(0)$, then $C_{\lambda,n} > 0$. If $N_K < 2K(0)$, then, for λ sufficiently small, and $n\lambda$ sufficiently large, $C_{\lambda,n} > -1$ and $r_{\lambda,n}(s)$ is increasing in λ and $N(\lambda)$ and decreasing in S_ρ .

Note that for the unimodal kernels which are optimal for estimating the mean function (Gasser and Müller, 1979), N_K is always less than or equal to $2K(0)$.

Corollary 3: Under the conditions of Theorem 2, the asymptotically optimal bandwidth for estimating the correlations is

$$\lambda_\rho = \left[\frac{(N_K - 2K(0))(1 + 2S_\rho)}{2pN(0)} \left(\frac{p!}{s_K} \right)^2 \right]^{\frac{1}{2p+1}} n^{-\frac{1}{2p+1}} \quad (13)$$

The bandwidth λ_ρ should be compared with λ^* , the asymptotically optimal bandwidth for estimating the mean function. Altman, 1990, showed that

$$\begin{aligned}\lambda^* &= \left[\frac{N_K(1+2S_\rho)}{pN(0)} \left(\frac{p!}{s_K} \right)^2 \right]^{\frac{1}{2p+1}} n^{-\frac{1}{2p+1}}. \\ &= \lambda_\rho \left(\frac{2N_K}{N_K - 2K(0)} \right)^{\frac{1}{2p+1}}.\end{aligned}\tag{14}$$

Therefore, different bandwidths are required for estimating the regression and correlation functions.

3. Appendix

3.1. Proof of Theorem 1

Define $\eta_{\lambda,t}$ by

$$\eta_{\lambda,t} = \varepsilon_t - w'_\lambda \varepsilon_t \tag{A.1}$$

where $w_\lambda(i) = \frac{1}{n\lambda} K(\frac{i}{n\lambda})$ for $i = -[\frac{n\lambda}{2}] \dots [\frac{n\lambda}{2}]$ and ε_t is the vector $(\varepsilon_{t-[\frac{n\lambda}{2}]}, \dots, \varepsilon_{t+[\frac{n\lambda}{2}]})$. Let $b_{\lambda,n}(t) = E(\mu(t) - \hat{\mu}_{\lambda,n}(t))$. Note that, for $[\frac{n\lambda}{2}] < i < n - [\frac{n\lambda}{2}]$, the residuals are $e_{\lambda,n}(i) = b_{\lambda,n}(\frac{i}{n}) + \eta_{\lambda,i}$.

$\eta_{\lambda,t}$ is clearly stationary and has the form:

$$\eta_{\lambda,t} = \sum_{j=1}^{\infty} \phi_j z_{t-j}$$

where z_t is defined by (9) and $\sum_{j=1}^{\infty} |\phi_j| < \infty$. Let

$$\hat{\gamma}^*(s) = \frac{1}{n} \sum_{t=[\frac{n\lambda}{2}]}^{n-[\frac{n\lambda}{2}]-s} \eta_{\lambda,t} \eta_{\lambda,t+s}$$

By Brockwell and Davis (1987, Chapter 7):

$$\sqrt{n} \begin{pmatrix} \hat{\gamma}^*(0) \\ \vdots \\ \hat{\gamma}^*(k) \end{pmatrix} \sim AN \left(\begin{pmatrix} \gamma^*(0) \\ \vdots \\ \gamma^*(k) \end{pmatrix}, \frac{V}{n} + O\left(\frac{1}{n^2\lambda}\right) \right)$$

where V is given by Bartlett's formula, and $\gamma^*(s) = E(\eta_{\lambda,t}\eta_{\lambda,t+s})$. The extra term, $O(\frac{1}{n^2\lambda})$, in the variance, may be found by direct computation.

Now

$$\hat{\gamma}_{n,\lambda}(s) = \frac{1}{n} \sum_{t=\lfloor \frac{n\lambda}{2} \rfloor}^{n-\lfloor \frac{n\lambda}{2} \rfloor-s} \left(\eta_{\lambda,t} + b_{\lambda,n}\left(\frac{t}{n}\right) \right) \left(\eta_{\lambda,t+s} + b_{\lambda,n}\left(\frac{t+s}{n}\right) \right) \quad (A.2)$$

We will show that

$$\frac{1}{n} \sum_{t=\lfloor \frac{n\lambda}{2} \rfloor}^{n-\lfloor \frac{n\lambda}{2} \rfloor-s} \eta_{\lambda,t} b_{\lambda,n}\left(\frac{t+s}{n}\right) = o_p(1) \quad (A.3)$$

and

$$\frac{1}{n} \sum_{t=\lfloor \frac{n\lambda}{2} \rfloor}^{n-\lfloor \frac{n\lambda}{2} \rfloor-s} \eta_{\lambda,t+s} b_{\lambda,n}\left(\frac{t}{n}\right) = o_p(1) \quad (A.4)$$

so that,

$$\sqrt{n} \begin{pmatrix} \hat{\gamma}_{\lambda,n}(0) \\ \vdots \\ \hat{\gamma}_{\lambda,n}(k) \end{pmatrix} \sim AN \left(\begin{pmatrix} \gamma^*(0) + \frac{1}{n} \sum_{t=\lfloor \frac{n\lambda}{2} \rfloor}^{n-\lfloor \frac{n\lambda}{2} \rfloor} b_{\lambda,n}^2\left(\frac{t}{n}\right) \\ \vdots \\ \gamma^*(k) + \frac{1}{n} \sum_{t=\lfloor \frac{n\lambda}{2} \rfloor}^{n-\lfloor \frac{n\lambda}{2} \rfloor-k} b_{\lambda,n}\left(\frac{t}{n}\right) b_{\lambda,n}\left(\frac{t+k}{n}\right) \end{pmatrix}, \frac{V}{n} + O\left(\frac{1}{n^2\lambda}\right) \right).$$

Since $\eta_{\lambda,n}$ is stationary, A.3 and A.4 can be replaced by $\frac{1}{n} \sum_{t=\lfloor \frac{n\lambda}{2} \rfloor}^{n-\lfloor \frac{n\lambda}{2} \rfloor-s} \eta_{\lambda,t} b_{\lambda,n}\left(\frac{t}{n}\right)$. By Lemma A.1, $\sum_{s=1}^{\infty} E(\eta_{\lambda,t}\eta_{\lambda,t+s})$ converges absolutely to 0.

By Chebyshev's inequality and Lemmas A.1 and A.2

$$\begin{aligned} P\left(\left| \sum_{t=\lfloor \frac{n\lambda}{2} \rfloor}^{n-\lfloor \frac{n\lambda}{2} \rfloor-s} \eta_{\lambda,t} b_{\lambda,n}\left(\frac{t}{n}\right) \right| > M\sqrt{n} \right) &\leq \frac{1}{M^2 n} b'_{\lambda,n} E(\eta_{\lambda,1}, \dots, \eta_{\lambda,n}) (\eta_{\lambda,1}, \dots, \eta_{\lambda,n})' b_{\lambda,n} \\ &= \frac{\sigma^2}{M^2} \int b_{\lambda,n}^2(x) dx + o(1) \end{aligned}$$

Thus proving A.3 and A.4.

Finally, we show that

$$\gamma_{\lambda,n}(s) = \sigma^2 \left(\rho(s) + \lambda^{2p} \left(\frac{sK}{p!} \right)^2 N(\lambda) + \frac{1+2S\rho}{n\lambda} (N_K - 2K(0)) \right) + o(\lambda^{2p}) + o\left(\frac{1}{n\lambda}\right) + o\left(\frac{s}{n}\right).$$

Define

$$\mathbb{Y}_k^s = E(\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_k)(\boldsymbol{\varepsilon}_{1+s}, \dots, \boldsymbol{\varepsilon}_{k+s})' \quad (\text{A.5})$$

and note that

$$\gamma^*(s) = \sigma^2 \rho(s) - \sigma^2 \sum_{i=-\lfloor \frac{n\lambda}{2} \rfloor}^{\lfloor \frac{n\lambda}{2} \rfloor} w_\lambda(i)(\rho(k+i) + \rho(k-i)) + w'_\lambda \mathbb{Y}_{2\lfloor \frac{n\lambda}{2} \rfloor + 1}^s w_\lambda$$

By Lemma A.3,

$$\sum_{i=-\lfloor \frac{n\lambda}{2} \rfloor}^{\lfloor \frac{n\lambda}{2} \rfloor} w_\lambda(i)(\rho(k+i) + \rho(k-i)) = 2 \frac{K(0)}{n\lambda} (1 + 2S_\rho) + o\left(\frac{1}{n\lambda}\right).$$

By Lemmas A.2, A.4 and A.5,

$$w'_\lambda \mathbb{Y}_{2\lfloor \frac{n\lambda}{2} \rfloor + 1}^s w_\lambda = \sigma^2 \frac{NK}{n\lambda} (1 + 2S_\rho) + o\left(\frac{1}{n\lambda}\right).$$

Finally, by Lemma A.6

$$\sum_{i=1}^n b_{\lambda,n}^2 \left(\frac{i}{n}\right) = \lambda^{2p} \left(\frac{sK}{p!}\right)^2 \int \mu^{(p)}(x)^2 dx + o(\lambda^{2p}).$$

3.2. Lemmas and Technical Propositions

Proposition A.1: If $g(x)$ is Lipschitz continuous of order α with support on $[0, 1]$ then

$$\left| \frac{1}{n} \sum_{i=1}^n g^2\left(\frac{i}{n}\right) - \int g^2(x) dx \right| = o\left(\frac{1}{n}\right). \quad (\text{A.6})$$

and if $0 < c < \frac{i}{n} < 1 - c$ and $\frac{i+j}{n} < 1$ then

$$\frac{1}{n} \left| \sum_{i=1}^{n-j} g\left(\frac{i}{n}\right) \left(g\left(\frac{i+j}{n}\right) - g\left(\frac{i}{n}\right) \right) \right| = o(1). \quad (\text{A.7})$$

Remark 1: A.6 and A.7 are true, with appropriate adjustments, if the interval is rescaled to $[-\frac{1}{2}, \frac{1}{2}]$, and n is replaced by $n\lambda$.

Proposition A.2: If the errors satisfy condition F then as $N \rightarrow \infty$, then, letting $s_j = N - j + 1$,

$$\frac{1}{N} \sum_{j=1}^N j |\rho(j) - \rho(s_j)| = o(1). \quad (A.8)$$

Lemma A.1: Let $\gamma^*(s) = E(\eta_{\lambda,t} \eta_{\lambda,t+s})$, where $\eta_{\lambda,t}$ is defined by A.1. If the process ε_t satisfies condition E then $\sum_{i=1}^{\infty} \gamma^*(s)$ converges absolutely to 0.

Proof of Lemma A.1:

$$\begin{aligned} \gamma^*(s) &= \sigma^2(\rho(s) - \sum_{i=-[\frac{n\lambda}{2}]}^{[\frac{n\lambda}{2}]} w_{\lambda}(i)(\rho(s+i) + \rho(s-i))) \\ &\quad + \sum_{i=-[\frac{n\lambda}{2}]}^{[\frac{n\lambda}{2}]} \sum_{j=-[\frac{n\lambda}{2}]}^{[\frac{n\lambda}{2}]} w_{\lambda}(i)w_{\lambda}(j)\rho(s+j-i) \end{aligned}$$

Since $\sum_{s=1}^{\infty} \rho(s)$ converges absolutely, we can write,

$$\begin{aligned} \sum_{s=1}^{\infty} |\gamma^*(s)| &\leq \sigma^2 \left(1 + 2 \sum_{i=-[\frac{n\lambda}{2}]}^{[\frac{n\lambda}{2}]} |w_{\lambda}(i)| + \left(\sum_{i=-[\frac{n\lambda}{2}]}^{[\frac{n\lambda}{2}]} |w_{\lambda}(i)| \right)^2 \right) \sum_{s=1}^{\infty} |\rho(s)| \\ &< \infty \end{aligned}$$

Accordingly, and since the sum of the weights converges to one,

$$\begin{aligned} \sum_{i=1}^{\infty} \gamma^*(s) &= \sigma^2 \left(1 - 2 \sum_{i=-[\frac{n\lambda}{2}]}^{[\frac{n\lambda}{2}]} w_{\lambda}(i) + \left(\sum_{i=-[\frac{n\lambda}{2}]}^{[\frac{n\lambda}{2}]} w_{\lambda}(i) \right)^2 \right) \sum_{s=1}^{\infty} \rho(s) \\ &= O\left(\frac{1}{n\lambda}\right) \end{aligned}$$

Lemma A.2:(Altman, 1990, Lemma A.4) Let \mathbb{X}_n be the variance matrix of the errors, $\varepsilon_1 \cdots \varepsilon_n$. For a function g defined on $[0, 1]$, define the vector g_n by $g_n(i) = \frac{1}{n}g(\frac{i}{n})$. If $g(x)$ satisfies conditions A.6 and A.7 and the errors satisfy conditions E and A.8 then for $0 < \frac{i}{n} < 1$

$$\left| g'_n \mathbb{X}_n g_n - \sigma^2 \frac{\int g^2(x) dx}{n} (1 + 2S_{\rho}) \right| = o\left(\frac{1}{n}\right). \quad (A.9)$$

Lemma A.3: (Altman, 1990 Lemma 2.1) If the kernel weights (rescaled to $[0, 1]$) satisfy conditions A.6 and A.7, the errors satisfy condition E and A.8, $s \ll n\lambda$ and $\frac{\lambda}{2} < \frac{i}{n} < 1 - \frac{\lambda}{2}$, then

$$\frac{1}{n\lambda} \sum_{j=1}^{\lfloor \frac{n\lambda}{2} \rfloor} K\left(\frac{i+j}{n\lambda}\right) \rho(j+s) = \frac{K(0)}{n\lambda} S_\rho + o\left(\frac{1}{n\lambda}\right). \quad (\text{A.10})$$

Lemma A.4: Let w_λ be the vector of weights defined by (A.1). Then if the kernel function satisfies conditions A.6 and A.7, and the errors satisfy conditions E and A.8, for $\frac{\lambda}{2} < \frac{i}{n} < 1 - \frac{\lambda}{2}$

$$\left| w'_\lambda \mathbb{F}_n w_\lambda - \sigma^2 \frac{NK}{n\lambda} (1 + 2S_\rho) \right| = o\left(\frac{1}{n\lambda}\right).$$

Proof of Lemma A.4: This follows from Lemma A.2, after noting that the number of points receiving non-zero weight is $n\lambda$.

Lemma A.5: Let $\mathbb{F}_{2\lfloor \frac{n\lambda}{2} \rfloor + 1}^1$ be the covariance matrix at lag one, defined by A.5. If the kernel weights satisfy conditions A.6 and A.7 and the errors satisfy condition E and A.8 then for $\frac{\lambda}{2} < \frac{i}{n} < 1 - \frac{\lambda}{2}$

$$\left| w'_\lambda \mathbb{F}_{2\lfloor \frac{n\lambda}{2} \rfloor + 1}^1 w_\lambda - \sigma^2 \frac{NK}{n\lambda} (1 + 2S_\rho) \right| = o\left(\frac{1}{n\lambda}\right).$$

Proof of Lemma A.5: The computation proceeds, in the same way as Lemma A.2, by substituting a simpler matrix, $S_{n\lambda}$ for $\mathbb{F}_{2\lfloor \frac{n\lambda}{2} \rfloor + 1}^1$ and then showing that $|w'_\lambda (\mathbb{F}_{2\lfloor \frac{n\lambda}{2} \rfloor + 1}^1 - S_{n\lambda}) w_\lambda| = o\left(\frac{1}{n\lambda}\right)$. Let $S_{n\lambda}$ be the matrix

$$\sigma^2 \begin{bmatrix} A'_{n\lambda} & \rho(2\lfloor \frac{n\lambda}{2} \rfloor + 1) \\ E_{n\lambda} & A_{n\lambda} \end{bmatrix}$$

where $A_{n\lambda}$ is the vector $(\rho(1), \rho(2), \dots, \rho(\lfloor \frac{n\lambda}{2} \rfloor), \rho(\lfloor \frac{n\lambda}{2} \rfloor - 1) \dots, \rho(1), \rho(2\lfloor \frac{n\lambda}{2} \rfloor))'$, and $E_{n\lambda}$ is the $2\lfloor \frac{n\lambda}{2} \rfloor \times 2\lfloor \frac{n\lambda}{2} \rfloor$ matrix

$$E_{n\lambda}(i, j) = \begin{cases} \rho(|i-j|) & |i-j| \leq k \\ \sigma^2 \rho(2\lfloor \frac{n\lambda}{2} \rfloor + 1 - |i-j|) & |i-j| > k. \end{cases}$$

($E_{n\lambda}$ is similar to R_n of Altman, 1990, Lemma A.4). Define $v_{n\lambda}$, and $z_{n\lambda}$ by

$$\begin{aligned} w'_\lambda &= (w_\lambda(i - \lfloor \frac{n\lambda}{2} \rfloor), v'_{n\lambda}) \\ &= (z'_{n\lambda}, w_\lambda(i + \lfloor \frac{n\lambda}{2} \rfloor)). \end{aligned}$$

By the same method as Lemma A.2,

$$z'_{n\lambda} E_{n\lambda} z_{n\lambda} = \sigma^2 \frac{NK}{n\lambda} (1 + 2S_\rho) + o(\frac{1}{n\lambda})$$

By the Lipschitz condition on K ,

$$|v'_{n\lambda} E_{n\lambda} z_{n\lambda} - z'_{n\lambda} E_{n\lambda} z_{n\lambda}| = o(\frac{1}{n\lambda}).$$

By Lemma A.3,

$$\begin{aligned} v'_{n\lambda} A_{n\lambda} &= A'_{n\lambda} z_{n\lambda} \\ &= 2 \frac{K(0)S_\rho}{n\lambda} + o(\frac{1}{n\lambda}). \end{aligned}$$

Putting this all together, we obtain:

$$\begin{aligned} \left| w'_\lambda S_{n\lambda} w_\lambda - \sigma^2 \frac{NK}{n\lambda} (1 + 2S_\rho) \right| &\leq 2 \left| w_\lambda(i - \lfloor \frac{n\lambda}{2} \rfloor) A'_{n\lambda} z_{n\lambda} \right| \\ &+ \left| w_\lambda(i - \lfloor \frac{n\lambda}{2} \rfloor) w_\lambda(i + \lfloor \frac{n\lambda}{2} \rfloor) \rho(2\lfloor \frac{n\lambda}{2} \rfloor + 1) \right| \\ &+ |v'_{n\lambda} E_{n\lambda} z_{n\lambda} - \sigma^2 \frac{NK}{n\lambda} (1 + 2S_\rho)| \\ &= o(\frac{1}{n\lambda}). \end{aligned}$$

Let $L = \max_x K(x)$. Then, $\mathbb{X}_{2\lfloor \frac{n\lambda}{2} \rfloor + 1}^1$ and $S_{n\lambda}$ are the same on the first $\lfloor \frac{n\lambda}{2} \rfloor$ upper diagonals, and $\lfloor \frac{n\lambda}{2} \rfloor + 2$ lower diagonals, so

$$\begin{aligned} \left| w'_\lambda \mathbb{X}_{2\lfloor \frac{n\lambda}{2} \rfloor + 1}^1 w_\lambda - w'_\lambda S_{n\lambda} w_\lambda \right| \\ \leq 2 \left(\frac{L}{n\lambda} \right)^2 \sum_{j=1}^{\lfloor \frac{n\lambda}{2} \rfloor - 1} (j+1) \left| \rho(j) - \rho(2\lfloor \frac{n\lambda}{2} \rfloor - j) \right| \\ = o(\frac{1}{n\lambda}) \end{aligned}$$

Corollary A.1: Let $\mathbb{X}_{2\lfloor \frac{n\lambda}{2} \rfloor + 1}^s$ be the covariance matrix of defined by A.5. If the kernel weights satisfy conditions A.6 and A.7 and the errors satisfy conditions E and A.8

then for $\frac{\lambda}{2} < \frac{i}{n} < 1 - \frac{\lambda}{2}$ and $s \ll n\lambda$

$$\left| w'_\lambda \mathbb{F}_{2[\frac{n\lambda}{2}]+1}^s w_\lambda - \sigma^2 \frac{N_K}{n\lambda} (1 + 2S_\rho) \right| = o\left(\frac{1}{n\lambda}\right) + o\left(\frac{s}{n}\right).$$

The proof proceeds as for Lemma A.5.

Lemma A.6: Under condition D,

$$|E(e_{\lambda,n}(i))E(e_{\lambda,n}(i+s)) - \lambda^{2p} s_K^2 (\mu^{(p)}(\frac{i}{n}))^2 / (p!)^2| = o(\lambda^{2p}) + o(\frac{1}{n\lambda}).$$

Proof of A.9: Gasser and Müller, (1979) showed $E(e_{\lambda,n}(i)) = (-\lambda)^p s_K(\mu^{(p)}(\frac{i}{n})) / (p!) + o(\lambda^p) + o(\frac{1}{n\lambda})$. The result then follows simply from the Lipschitz condition, on the p^{th} derivative of μ .

References

- Altman, N.S. (1990) "Kernel Smoothing of Data with Correlated Errors," *Journal of the American Statistical Association* 85 749-759.
- Bartlett, M.S. (1946) "On the theoretical specification and sampling properties of auto-correlated time series," *Supplement of the Journal of the Royal Statistical Society* 8 27-41.
- Box, G.E.P. and Jenkins, G.M. (1976) *Time Series Analysis: Forecasting and Control*, Holden-Day, Inc. San Francisco.
- Brockwell, P.J. and Davis, R.A. (1987) *Time Series: Theory and Methods*, Springer-Verlag, New York.
- Gasser, T., Müller, H-G. (1979) "Kernel Estimation of Regression Functions," in *Smoothing Techniques for Curve Estimation* 23-37, Lecture Notes in Mathematics 757. Springer-Verlag, Berlin.
- Hart, J.D. (1991) "Kernel Regression with Time Series Errors," (to appear in the *Journal of the Royal Statistical Society Series B*).
- IMSL Incorporated (1984) *IMSL Library Reference Library*, Houston.
- Nadaraya, E.A. (1964) "On estimating regression," *Theory of Probability and its Applications* 9 141-142.
- Priestly, M.B. and Chao, M.T. (1972) "Non-parametric function fitting," *Journal of the Royal Statistical Society Series B* 34 385-392.
- Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York.
- Silverman, B.W. (1984) "Spline Smoothing: The Equivalent Variable Kernel Method," *Annals of Statistics* 12 898-916.
- Silverman, B.W. (1985) "Some Aspects of the Spline Smoothing Approach to Non-parametric Regression Curve Fitting," *Journal of the Royal Statistical Society Series B* 47 1-52.

Truong, Y.K. (1989) "Nonparametric curve estimation with time series errors," preprint.

Watson, G.S. (1964) "Smooth Regression Analysis," *Sankhyā, Ser. A* 26 359-372.