Efficient nonparametric estimation and inference for the volatility function †

FRANCESCO GIORDANO AND MARIA LUCIA PARRELLA

Supplemental online materials

Appendix B. Additional theoretical results

In this Section we show a self-contained method to estimate the unknown functionals for the asymptotic bandwidth parameter in Local Linear Polynomial estimator. We extend the method in [1] to the case of dependent data. Moreover, we use a different technical approach as in [2] and [3] in order to deal with the Neural Networks estimator for unknown functions defined on non compact sets. Assume the following model

\[ X_t = \sigma(X_{t-1}) \epsilon_t, \quad t \in \mathbb{N}, \]  

(S.1)

and consider the following assumptions

Assumptions (a)

(a1) The errors \( \epsilon_t \) have a continuous and positive density function with \( E(\epsilon_t^2) = 1, \ E(\epsilon_t) = E(\epsilon_t^3) = 0, \ E(\epsilon_t^4) < \infty \).

(a2) The function \( \sigma(\cdot) \) is positive and has a continuous second derivative.

(a3) There exist some constants \( M \) and \( \alpha \) such that,
   i) \( 0 < M < \infty, \ \sigma(y) \leq M(1 + |y|) \) and \( M \left[ E|\epsilon_t|^4 \right]^{1/4} < 1 \) for all \( y \in \mathbb{R} \);
   ii) \( 0 \leq \alpha \leq M, \ \sigma(y) - \alpha|y| = o(1) \) for \( |y| \to \infty \).

(a4) The process \( \{X_t\} \) is strictly stationary.

(a5) The density function \( f_X(\cdot) \) of the (stationary) measure of the process \( \mu_X \) exists; it is bounded, continuous and positive on every compact set in \( \mathbb{R} \).

Assumptions (b)

(b1) \( \sum_{k=1}^{d_n} |c_k| \leq \Delta_n \).

(b2) \( d_n \to \infty, \ \Delta_n \to \infty \) as \( n \to \infty \).

(b3) The activation function, \( \psi(\cdot) \), is strictly increasing, sigmoidal and has a continuous second derivative.

Remark 1 Under the assumptions (a1)–(a5), it can be shown that the process is geometrically ergodic and exponentially \( \alpha \)-mixing (see [4]).

Let us consider, for some \( \lambda > 0 \),

\[ Q_1(n) := \frac{\Delta_n^2 d_n \log(\Delta_n^2 d_n)}{\sqrt{n}}, \quad Q_2(n) := \frac{\Delta_n^4}{n^{1-\lambda}}, \quad \mathcal{F}_n := \left\{ q : \sum_{k=1}^{d_n} |c_k| \leq \Delta_n \right\}, \]  

(S.2)

where the neural network function is \( q(x; \eta) = c_0 + \sum_{k=1}^{d_n} c_k \psi(b_{k0} + b_{k1} x) \) with \( \eta = (c_0, c_1, \ldots, c_{d_n}, b_1, \ldots, b_{d_n})^T \) and \( b_k = (b_{k0}, b_{k1})^T \) for each \( k = 1, \ldots, d_n \). The function
ψ(·) is sigmoidal in the sense that it is an increasing function such that \( \lim_{u \to \infty} \psi(u) = 1 \) and \( \lim_{u \to -\infty} \psi(u) = 0 \).

\( \mathcal{F}_n \) is the class of feedforward neural networks with bounded weights, that is \( \sum_{k=1}^{d_n} |c_k| \leq \Delta_n \). Now \( \mathcal{F} = \bigcup_{n=1}^{\infty} \mathcal{F}_n \) is the class of general feedforward neural networks. \( \mathcal{F} \) is dense with respect to the class of squared integrable functions using a predefined measure (see [5]). Moreover, if \( \Delta_n \equiv \Delta \), one can show, under some further technical assumptions, that \( \mathcal{F} \) is dense with respect to the class of squared integrable functions using a predefined measure with a bound which only depends on \( d_n^{-1/2} \) (see [3] and references therein). So, \( d_n \) means as the smoothing parameter and it needs to diverge. Instead, \( \Delta_n \) can be either constant or divergent. However, what we need is that either \( Q_1(n) \) or \( Q_2(n) \) should go to zero when \( n \to \infty \).

Under model (S.1), the Neural Network estimator \( q(x, \hat{\eta}) \) can be written as

\[
q(x, \hat{\eta}) = \arg \min_{f \in \mathcal{F}_n} \frac{1}{n-1} \sum_{k=1}^{n-1} (X_{k+1}^2 - f(X_k))^2
\]

(S.3)

First of all, we report some preliminary results for the Neural Networks estimator. The following Lemma 1 extends the results for the consistency in [2] with respect to the Neural Network estimator, \( q(x, \hat{\eta}) \), using assumptions (a) and (b). Moreover, the same consistency, as in [2], is shown in the Lemma 2 for the Neural Network estimator of the second derivative for the unknown function \( \sigma^2(x) \).

**LEMMA 1** Under assumptions (a1) – (a5) and (b), the estimator \( q(x; \hat{\eta}) \) of \( \sigma^2(x) \), defined in (S.3), is consistent in the sense that:

- if \( Q_1(n) \to 0 \) as \( n \to \infty \), then
  \[
  E \int (q(x; \hat{\eta}) - \sigma^2(x))^2 \, d\mu_X(x) \to 0 \quad n \to \infty;
  \]
- if, additionally, \( Q_2(n) \to 0 \) for some \( \lambda > 0 \), then
  \[
  \int (q(x; \hat{\eta}) - \sigma^2(x))^2 \, d\mu_X(x) \overset{a.s.}{\to} 0 \quad n \to \infty.
  \]

**Proof:** It is sufficient to apply Theorem 3.2 in [2] with respect to the estimator \( q(x, \hat{\eta}) \). Based on the previous Remark 1, the process in (S.1) is exponentially \( \alpha \)-mixing and the activation function for Neural Network estimator is sigmoidal, continuous and strictly increasing by (b3). So the conditions for the Theorem 3.2 in [2] are satisfied. \( \square \)

**LEMMA 2** Under the same assumptions as in Lemma 1, the estimator of the second derivative of \( \sigma^2(x) \) is consistent in the sense that:

- if \( Q_1(n) \to 0 \) as \( n \to \infty \), then
  \[
  E \int \left( q^{(2)}(x; \hat{\eta}) - \sigma_{(2)}^2(x) \right)^2 \, d\mu_X(x) \to 0 \quad n \to \infty;
  \]
- if, additionally, \( Q_2(n) \to 0 \) for some \( \lambda > 0 \), then
  \[
  \int \left( q^{(2)}(x; \hat{\eta}) - \sigma_{(2)}^2(x) \right)^2 \, d\mu_X(x) \overset{a.s.}{\to} 0 \quad n \to \infty,
  \]
where \( \sigma^2_{(2)}(x) \) is the second derivative of \( \sigma^2(x) \).

**Proof:** Define with \( G \) the class of all functions \( \sigma^2(x) \) satisfying the assumptions (a2) and (a3). Now we can write

\[
\int \left( q^{(2)}(x; \tilde{\eta}) - \sigma^2_{(2)}(x) \right)^2 d\mu_X(x) \leq \|D^2\|^2 \int \left( q(x; \tilde{\eta}) - \sigma^2(x) \right)^2 d\mu_X(x)
\]

where \( \|D^2\|^2 = \sup_{f \in G} \int (f''(x))^2 d\mu_X(x) \).

By assumptions, the linear operator \( D^2 \) is bounded. So \( \|D^2\|^2 < \infty \). Finally, using Lemma 1 we obtain the result.

The next two lemmas are used in Propositions 1 and 2.

**Lemma 3** Under the same assumptions as in Lemma 1, 
\[
\frac{1}{n - 1} \sum_{t=1}^{n-1} (q(X_t; \tilde{\eta}) - \sigma^2(X_t))^2
\]

is consistent in the sense that:

- if \( Q_1(n) \rightarrow 0 \) as \( n \rightarrow \infty \), then
  \[
  \frac{1}{n - 1} \sum_{t=1}^{n-1} (q(X_t; \tilde{\eta}) - \sigma^2(X_t))^2 \xrightarrow{p} 0 \quad n \rightarrow \infty;
  \]

- if, additionally, \( Q_2(n) \rightarrow 0 \) for some \( \lambda > 0 \), then
  \[
  \frac{1}{n - 1} \sum_{t=1}^{n-1} (q(X_t; \tilde{\eta}) - \sigma^2(X_t))^2 \xrightarrow{a.s.} 0 \quad n \rightarrow \infty.
  \]

**Proof:** By Theorem 3.2 in [2], which uses the same line of the proof as in Theorem 16.1 in [3], we have that

\[
W_1 := \sup_{f \in F_n} \left\{ \frac{1}{n - 1} \sum_{k=1}^{n-1} \left| f(X_k) - X_{k+1}^2 \right|^2 - E \left\{ \left| f(X_t) - X_{t+1}^2 \right|^2 \right\} \right\} \xrightarrow{p(a.s.)} 0
\]

for \( n \rightarrow \infty \). The above convergence is in probability or almost sure according to the two conditions, \( Q_1(n) \rightarrow 0 \) and \( Q_2(n) \rightarrow 0 \), respectively. Below, we use only the convergence in probability because the almost sure convergence follows exactly the same technique.

The neural network estimator \( q(X_t; \tilde{\eta}) \in F_n \) for some \( n > n_0 \). Using model (S.1) we can write

\[
W_2 := \frac{1}{n - 1} \sum_{k=1}^{n-1} (q(X_k; \tilde{\eta}) - \sigma^2(X_k))^2 - E \left\{ (q(X_t; \tilde{\eta}) - \sigma^2(X_t))^2 \right\} +
\]

\[+
\frac{1}{n - 1} \sum_{k=1}^{n-1} (\sigma^2(X_k)\varepsilon_{k+1}^2 - \sigma^2(X_k))^2 - E \left\{ \sigma^2(X_t)\varepsilon_{t+1}^2 - \sigma^2(X_t)^2 \right\} +
\]

\[- \frac{2}{n - 1} \sum_{k=1}^{n-1} \left| q(X_k; \tilde{\eta}) - \sigma^2(X_k) \right| \left| \sigma^2(X_k)\varepsilon_{k+1}^2 - \sigma^2(X_k) \right| +
\]

\[+ 2E \left\{ \left| q(X_k; \tilde{\eta}) - \sigma^2(X_k) \right| \left| \sigma^2(X_k)\varepsilon_{k+1}^2 - \sigma^2(X_k) \right| \right\}.
\]

(S.4)
Therefore $W_2 \xrightarrow{p} 0$. Consider the terms in the second line of (S.4). By assumptions (a) it follows that $E \left\{ (\sigma^2(X_i) \varepsilon_{i+1}^2 - \sigma^2(X_i))^2 \right\} = E [\sigma^4(X_i)] E \left\{ (\varepsilon_i^2 - 1)^2 \right\} := c_1$ with $0 < c < \infty$. By ergodicity of the process $\{X_t\}$ and using the assumptions (a) we have that

$$
\frac{1}{n-1} \sum_{k=1}^{n-1} (\sigma^2(X_k) \varepsilon_{k+1}^2 - \sigma^2(X_k))^2 \xrightarrow{a.s.} c_1 \quad n \to \infty.
$$

Since $E [\sigma^4(X_i)] < \infty$ then $E \left\{ (q(X_i; \hat{\eta}) - \sigma^2(X_i))^2 \right\} < \infty$ and by Lemma 1

$$
E \left\{ (q(X_i; \hat{\eta}) - \sigma^2(X_i))^2 \right\} \xrightarrow{p} 0
$$

when $n \to \infty$. Using Schwartz’s inequality we have that

$$
E \left\{ |q(X_k; \hat{\eta}) - \sigma^2(X_k)| \left| \sigma^2(X_k) \varepsilon_{k+1}^2 - \sigma^2(X_k) \right| \right\} \xrightarrow{p} 0.
$$

So,

$$
W_3 := \frac{1}{n-1} \sum_{k=1}^{n-1} (q(X_k; \hat{\eta}) - \sigma^2(X_k))^2 - \frac{2}{n-1} \sum_{k=1}^{n-1} (q(X_k; \hat{\eta}) - \sigma^2(X_k)) \left| \sigma^2(X_k) \varepsilon_{k+1}^2 - \sigma^2(X_k) \right|
$$

and $W_3 \xrightarrow{p} 0$. Since $E \left\{ (q(X_i; \hat{\eta}) - \sigma^2(X_i))^2 \right\} < \infty$ then it follows that

$$
\frac{1}{n-1} \sum_{k=1}^{n-1} (q(X_k; \hat{\eta}) - \sigma^2(X_k))^2 \xrightarrow{p} c_2
$$

when $n \to \infty$, with $0 \leq c_2 < \infty$. So, it implies that

$$
\frac{1}{n-1} \sum_{k=1}^{n-1} (q(X_k; \hat{\eta}) - \sigma^2(X_k)) \left| \sigma^2(X_k) \varepsilon_{k+1}^2 - \sigma^2(X_k) \right| \xrightarrow{p} c_2/2
$$

when $n \to \infty$. But, by Schwartz’s inequality we can write

$$
\frac{1}{n-1} \sum_{k=1}^{n-1} (q(X_k; \hat{\eta}) - \sigma^2(X_k)) \left| \sigma^2(X_k) \varepsilon_{k+1}^2 - \sigma^2(X_k) \right| \leq
$$

$$
\leq \left[ \frac{1}{n-1} \sum_{k=1}^{n-1} (q(X_k; \hat{\eta}) - \sigma^2(X_k))^2 \right]^{1/2} \left[ \frac{1}{n-1} \sum_{k=1}^{n-1} (\sigma^2(X_k) \varepsilon_{k+1}^2 - \sigma^2(X_k))^2 \right]^{1/2}.
$$

If we apply convergence, we have $c_2/2 \leq \sqrt{c_1 c_2}$. Since $c_1$ can be considered an arbitrary constant because it depends on the fourth moment of $\varepsilon_i$, while $c_2$ does not, the inequality is true if and only if $c_2 = 0$. This completes the proof. \qed

**Lemma 4** Under the same assumptions as in Lemma 1, \( \frac{1}{n-1} \sum_{i=1}^{n-1} (q^{(2)}(X_i; \hat{\eta}) - \sigma_{(2)}^2(X_i))^2 \) is consistent in the sense that:
• if $Q_1(n) \to 0$ as $n \to \infty$, then
\[
\frac{1}{n-1} \sum_{t=1}^{n-1} \left( q(X_t; \hat{\eta}) - \sigma^2_{(2)}(X_t) \right)^2 \overset{p}{\to} 0 \quad n \to \infty;
\]

• if, in addition, $Q_2(n) \to 0$ for some $\lambda > 0$, then
\[
\frac{1}{n-1} \sum_{t=1}^{n-1} \left( q(X_t; \hat{\eta}) - \sigma^2_{(2)}(X_t) \right)^2 \overset{a.s.}{\to} 0 \quad n \to \infty.
\]

**Proof:** As in the proof of Lemma 2, let $G$ be the class of all functions $\sigma^2(x)$ which satisfy the assumptions (a2) and (a3). Now, we have that
\[
\frac{1}{n-1} \sum_{k=1}^{n-1} \left( q(X_k; \hat{\eta}) - \sigma^2_{(2)}(X_k) \right)^2 \leq \left\| d_n^2 \right\|^2 \frac{1}{n-1} \sum_{k=1}^{n-1} [f(X_k)]^2
\]
where $\left\| d_n^2 \right\|^2 = \sup_{f \in G} \frac{1}{n-1} \sum_{k=1}^{n-1} [f''(x_k)]^2$ and $\left\| \hat{d}_n^2 \right\|^2 = \sup_{f \in G} \frac{1}{n-1} \sum_{k=1}^{n-1} [f''(X_k)]^2$,
with the stochastic process $\{X_t\}$ defined in (S.1) and $\| \cdot \|$ the norm of $L_2$ space with respect to the empirical measure. Based on assumption (a3), every function in $G$ has a bounded second derivative and so
\[
\lim_{n \to \infty} \frac{1}{n-1} \sum_{k=1}^{n-1} [f''(x_k)]^2 < \infty
\]
for every sequence $\{x_k\} \in \mathbb{R}$, $k = 1, 2, \ldots$.

Based on assumptions (a) and ergodicity of the stochastic process $\{X_t\}$ we have that
\[
\left\| \hat{d}_n^2 \right\|^2 \overset{a.s.}{\to} c < \infty.
\]
Finally, using Lemma 3 it follows that
\[
\left\| \hat{d}_n^2 \right\|^2 \frac{1}{n-1} \sum_{k=1}^{n-1} \left( q(X_k; \hat{\eta}) - \sigma^2(X_k) \right)^2 \overset{a.s.}{\to} 0 \quad n \to \infty.
\]

The above convergence is in probability if $Q_1(n) \to 0$, when $n \to \infty$. If, in addition, $Q_2(n) \to 0$, when $n \to \infty$, then there will be almost sure convergence. This completes the proof. \(\square\)

### B.1. Consistency for the Functional of the bias

Let $I_x = [x - a/2, x + a/2]$, with $a > 0$ for all $x \in \mathbb{R}$. According to assumption (a5) it follows that $\mu_{X}(I_x) > 0$. Moreover, the number of observed values in $I_x$ from (S.1) tends to infinity when $n \to \infty$ with probability 1.

Using model (S.1), we can write the functional of the bias, $B_{\omega_x}$, as
\[
B_{\omega_x} = C_1^2 \int_{I_x} \left( \sigma^2_{(2)}(x) \right)^2 f_X(x) \omega_x.
\]
Similarly, we can write its estimator as \( \hat{B}_{\omega}^{x} \), that is

\[
\hat{B}_{\omega}^{x} = \frac{C^2}{1/n - 1} \sum_{t=1}^{n-1} \frac{[q^{(2)}(X_t; \hat{\eta})]^2 \mathbb{I}(X_t \in I_x)}{\mathbb{I}(X_t \in I_x)}.
\]

(S.6)

**Proposition 1**  Under the same assumptions as in Lemma 1, \( \hat{B}_{\omega}^{x} \), defined in (S.6), is consistent in the sense that:

- If \( Q_1(n) \to 0 \) as \( n \to \infty \), then

\[
\hat{B}_{\omega}^{x} \xrightarrow{p} B_{\omega}^{x} \quad n \to \infty
\]

- if, additionally, \( Q_2(n) \to 0 \) for some \( \lambda > 0 \), then

\[
\hat{B}_{\omega}^{x} \xrightarrow{a.s.} B_{\omega}^{x} \quad n \to \infty
\]

where \( B_{\omega}^{x} \) is defined in (S.5).

**Proof:** For the sake of simplicity we consider only the convergence in probability. The almost sure convergence follows the same technique. The estimator in (S.6) can be written as

\[
\hat{B}_{\omega}^{x} = \frac{C^2}{1/n - 1} \sum_{t=1}^{n-1} \frac{[q^{(2)}(X_t; \hat{\eta})]^2 \mathbb{I}(X_t \in I_x)}{\mathbb{I}(X_t \in I_x)}.
\]

The quantity \( C^2 \) is known. By ergodicity of the stochastic process \( \{X_t\} \) it follows that

\[
\frac{1}{n - 1} \sum_{t=1}^{n-1} \mathbb{I}(X_t \in I_x) \xrightarrow{a.s.} \mu_X(I_x).
\]

Using assumptions (a) and again ergodicity of the stochastic process \( \{X_t\} \) we have that

\[
\frac{1}{n - 1} \sum_{t=1}^{n-1} \left( \sigma^{(2)}(X_t) \right)^2 \mathbb{I}(X_t \in I_x) \xrightarrow{a.s.} \int_{I_x} \left( \sigma^{(2)}(x) \right)^2 f_X(x) dx.
\]

By Lemma 4 \( \frac{1}{n - 1} \sum_{t=1}^{n-1} [q^{(2)}(X_t; \hat{\eta})]^2 \) and \( \frac{1}{n - 1} \sum_{t=1}^{n-1} \left( \sigma^{(2)}(X_t) \right)^2 \) converge in probability to the same limit. But the result is the same if we consider

\[
\frac{1}{n - 1} \sum_{t=1}^{n-1} \left[ q^{(2)}(X_t; \hat{\eta}) \right]^2 \mathbb{I}(X_t \in I_x).
\]

Therefore, it follows that

\[
\frac{1}{n - 1} \sum_{t=1}^{n-1} \left[ q^{(2)}(X_t; \hat{\eta}) \right]^2 \mathbb{I}(X_t \in I_x) \xrightarrow{p} \int_{I_x} \left( \sigma^{(2)}(x) \right)^2 f_X(x) dx.
\]
Since \( d\omega_{I_x}(u) = du/\mu_X(I_x) \) and \( \mu_X(I_x) > 0 \) we have that \( \hat{\omega}_{I_x} \xrightarrow{p} \omega_{I_x} \). The proof is complete. \( \square \)

Let \( m_{4\varepsilon} = E(\varepsilon_1^4) \) and \( \hat{m}_{4\varepsilon} = \frac{\sum_{t=1}^{n-1} X_t^4}{\sum_{t=1}^{n-1} [q(X_t; \hat{\eta})]^2} \).

**Corollary 1** Using the same conditions as in Proposition 1, the estimator \( \hat{m}_{4\varepsilon} \), is consistent in the sense that:

- if \( Q_1(n) \to 0 \) as \( n \to \infty \), then \( \hat{m}_{4\varepsilon} \xrightarrow{p} m_{4\varepsilon} \) \( n \to \infty \)
- if, in addition, \( Q_2(n) \to 0 \) for some \( \lambda > 0 \), then \( \hat{m}_{4\varepsilon} \xrightarrow{a.s.} m_{4\varepsilon} \) \( n \to \infty \).

**Proof:** As in the previous proofs, we analyze the convergence in probability since the almost sure convergence is straightforward. The estimator \( \hat{m}_{4\varepsilon} \) can be written as

\[
\hat{m}_{4\varepsilon} = \frac{\frac{1}{n-1} \sum_{t=1}^{n-1} X_t^4}{\frac{1}{n-1} \sum_{t=1}^{n-1} [q(X_t; \hat{\eta})]^2}.
\]

Based on assumptions (a) and ergodicity of the stochastic process \( \{X_t\} \), it follows that

\[
\frac{1}{n-1} \sum_{t=1}^{n-1} X_t^4 \xrightarrow{a.s.} E[\sigma^4(X_t)] m_{4\varepsilon} \quad n \to \infty
\]

and

\[
\frac{1}{n-1} \sum_{t=1}^{n-1} \sigma^4(X_t) \xrightarrow{a.s.} E[\sigma^4(X_t)] \quad n \to \infty
\]

since \( E[\sigma^4(X_t)] < \infty \) and \( m_{4\varepsilon} < \infty \). Using Lemma 3, it implies that \( \frac{1}{n-1} \sum_{t=1}^{n-1} \sigma^4(X_t) \) and \( \frac{1}{n-1} \sum_{t=1}^{n-1} [q(X_t; \hat{\eta})]^2 \) have the same limit in probability. Therefore,

\[
\frac{1}{n-1} \sum_{t=1}^{n-1} [q(X_t; \hat{\eta})]^2 \xrightarrow{p} E[\sigma^4(X_t)],
\]

when \( n \to \infty \). We can conclude that \( \hat{m}_{4\varepsilon} \xrightarrow{p} m_{4\varepsilon} \). The proof is complete. \( \square \)

**B.2. Consistency for the functional of variance**

Using model (S.1), we can write the functional of variance, \( \mathbb{V}_{\omega_{I_x}} \), as

\[
\mathbb{V}_{\omega_{I_x}} = C_2 \int_{I_x} [\sigma^4(u)] \, d\omega_{I_x}(u) \, (m_{4\varepsilon} - 1). \tag{S.7}
\]

Similarly, we can write its estimator as

\[
\hat{\mathbb{V}}_{\omega_{I_x}} = \frac{C_2 \sum_{i=1}^{n^*} [q(z_i; \hat{\eta})]^2 / n^*}{\sum_{t=1}^{n} \mathbb{I}(X_t \in I_x)/n} \, (\hat{m}_{4\varepsilon} - 1). \tag{S.8}
\]
The points \( \{z_1, z_2, \ldots, z_{n^*}\} \) are uniformly spaced values from the interval \( I_x \), with \( n^* = O(n) \).

**Proposition 2** Using the same conditions as in Proposition 1, then \( \hat{V}_{\omega I_x} \), defined in (S.8), with \( I_x \subset \mathbb{R} \) and \( n^* = O(n) \), is consistent in the sense that:

- If \( Q_1(n) \to 0 \) as \( n \to \infty \), then
  \[
  \hat{V}_{\omega I_x} \xrightarrow{p} V_{\omega I_x} \quad n \to \infty
  \]
- If, in addition, \( Q_2(n) \to 0 \) for some \( \lambda > 0 \), then
  \[
  \hat{V}_{\omega I_x} \xrightarrow{a.s.} V_{\omega I_x} \quad n \to \infty.
  \]

**Proof:** As in the previous proofs, we analyze the convergence in probability since the almost sure convergence is straightforward.

By Corollary 1, it follows that \( \hat{m}_{4e} \xrightarrow{p} m_{4e} \), when \( n \to \infty \). Using the ergodicity of the stochastic process \( \{X_t\} \), we have that 
\[
\sum_{t=1}^{n^*} \mathbb{I}(X_t \in I_x)/n \xrightarrow{a.s.} \mu_X(I_x), \quad n \to \infty.
\]
Since \( C_2 \) is a known quantity, we need only to show that
\[
\frac{1}{n^*} \sum_{i=1}^{n^*} [q(z_i; \hat{\eta})]^2 / n^* \xrightarrow{p} \int_{I_x} [\sigma^4(u)] \, du
\]
where the points \( \{z_1, z_2, \ldots, z_{n^*}\} \) are deterministic and uniformly spaced values from the interval \( I_x \).

By Lemma 1 and Lemma 3, we know that
\[
\inf_{s \in \mathcal{F}_n} \int_{\mathbb{R}} (s(x) - \sigma^2(x))^2 \, d\mu_X(x) \to 0 \quad n \to \infty \quad (S.9)
\]
\[
\sup_{s \in \mathcal{F}_n} \left\{ \frac{1}{n} \sum_{k=1}^{n-1} |s(X_k) - X_{k+1}^2|^2 - E \left\{ |s(X_t) - X_{t+1}^2|^2 \right\} \right\} \xrightarrow{P} 0 \quad n \to \infty. \quad (S.10)
\]

Both (S.9) and (S.10) refer to the Neural Network estimator with respect to the stochastic process \( \{X_t\} \). Instead, we consider some points which are not drawn by the process. So, in this case, we have to show that (S.9) and (S.10) hold.

By assumption (a5), we have that
\[
\inf_{s \in \mathcal{F}_n} \int_{\mathbb{R}} (s(x) - \sigma^2(x))^2 \, d\mu_X(x) \geq \inf_{s \in \mathcal{F}_n} \int_{I_x} (s(x) - \sigma^2(x))^2 \, f_X(x) \, dx \geq
\]
\[
\geq C_f \inf_{s \in \mathcal{F}_n} \int_{I_x} (s(x) - \sigma^2(x))^2 \, dx
\]
where \( C_f := \min_{x \in I} \{ f_X(x) \} \), with \( 0 < C_f < \infty \). By (S.9), it follows that
\[
\inf_{s \in \mathcal{F}_n} \int_{I_n} (s(x) - \sigma^2(x))^2 \, dx \to 0 \quad n \to \infty.
\]
Thus, we have proved that (S.9) is also true with respect to the points \( \{ z_1, z_2, \ldots, z_{n^*} \} \) uniformly spaced from the interval \( I_x \). Since \( n^* = O(n) \), we can consider asymptotically \( n \) instead of \( n^* \).

Put \( z_i = \bar{X}_i + (z_i - \bar{X}_i) \), where \( \bar{X}_i \in \{ X_1, X_2, \ldots, X_n \}, i = 1, 2, \ldots, n^* \). For every \( \epsilon > 0 \) and \( z_i \), we choose a \( \bar{X}_i \) such that \( |z_i - \bar{X}_i| < \epsilon \). Now, we have to show that such a \( \bar{X}_i \) exists with probability 1.

By assumption (a5) and based on Proposition A1.7 in [6], every non null compact set is a "small set" with respect to the Lebesgue measure for the Markov process in (S.1). But every set of radius \( \epsilon \), which contains \( z_i \) is non null compact set using the Lebesgue measure. Therefore, there exists a \( n_0 \) such that for each \( n > n_0 \) we can find at least a \( \bar{X}_i \in \{ X_1, X_2, \ldots, X_n \} \) with probability 1.

Define \( d_i := (z_i - \bar{X}_i) \). Then \( |d_i| < \epsilon \) with probability 1, when \( n \to \infty \), \( \forall i \).

Define \( Z_i := (X_i, d_i) \). The bi-dimensional random variables \( Z_i \) retain the property of exponentially \( \alpha \)-mixing because we have only deterministic variables \( z_i \) and random variables \( X_i \) which are exponentially \( \alpha \)-mixing. Since \( n^* = O(n) \), we can write, asymptotically,
\[
\sup_{s \in \mathcal{F}_n} \left| \sum_{k=1}^{n-1} \left| s(X_k) - X_{k+1}^2 \right|^2 - E \left\{ \left| s(X_t) - X_{t+1}^2 \right|^2 \right\} \right| 
\leq \sup_{s \in \mathcal{F}_n} \left| \sum_{k=1}^{n-1} \left| s(Z_k) - \bar{X}_{k+1}^2 \right|^2 - E \left\{ \left| s(Z_t) - \bar{X}_{t+1}^2 \right|^2 \right\} \right|
\]

because, using the proof of Theorem 3.2 from [2], the upper bounds for the \( \sup \) depend on \( d_n \), \( \Delta_n \), and the dimension of the input variables. But this dimension is 1 in (S.10) and 2 if we use \( Z_i \) as input variables, that is the uniformly spaced values in \( I_x \). Therefore, these upper bounds are the same when \( n \to \infty \). So, it follows that
\[
\sup_{s \in \mathcal{F}_n} \left| \sum_{k=1}^{n-1} \left| s(Z_k) - \bar{X}_{k+1}^2 \right|^2 - E \left\{ \left| s(Z_t) - \bar{X}_{t+1}^2 \right|^2 \right\} \right| \overset{p}{\to} 0 \quad n \to \infty.
\]

In this way, we can apply Lemma 3 in the case of the uniformly spaced values in \( I_x \). Then we have that \( \sum_{i=1}^{n^*} \left[ q(z_i; \hat{\eta}) \right]^2 / n^* \) and \( \sum_{i=1}^{n^*} (\sigma^4(z_i)) / n^* \) have the same limit in probability, when \( n \to \infty \).

Finally, the result follows. \( \square \)

References

[1] Giordano, F. and Parrella, M. (2014). Global Adaptive Smoothing Regression. Electronic Journal of Statistics 8, 2848–2878.

[2] Franke, J. and Diagne, M. (2006). Estimating market risk with neural networks. Statistics & Decision 24, 233–253.
[3] Györfi, L., Kohler, M., Krzyzak, A. and Walk, H. (2002). A Distribution-Free Theory of Nonparametric Regression, Springer-Verlag, Heidelberg.

[4] Härdle, W. and Tsybakov, A. (1997). Local polynomial estimation of the volatility function in nonparametric autoregression. *J. of Econometrics* **81**, 223–242.

[5] Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks* **4**, 251–257.

[6] Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford.