A Monte Carlo subsampling method for estimating the distribution of signal-to-noise ratio statistics in nonparametric time series regression models

Francesco Giordano · Pietro Coretto

Accepted: 19 August 2019 / Published online: 29 August 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract
Signal-to-noise ratio (SNR) statistics play a central role in many applications. A common situation where SNR is studied is when a continuous time signal is sampled at a fixed frequency with some noise in the background. While estimation methods exist, little is known about its distribution when the noise is not weakly stationary. In this paper we develop a nonparametric method to estimate the distribution of an SNR statistic when the noise belongs to a fairly general class of stochastic processes that encompasses both short and long-range dependence, as well as nonlinearities. The method is based on a combination of smoothing and subsampling techniques. Computations are only operated at the subsample level, and this allows to manage the typical enormous sample size produced by modern data acquisition technologies. We derive asymptotic guarantees for the proposed method, and we show the finite sample performance based on numerical experiments. Finally, we propose an application to electroencephalography data.

Keywords  Random subsampling · Nonparametric smoothing · Kernel regression · Time series data · Stochastic processes

1 Introduction

Signal-to-noise ratio (SNR) statistics are widely used to describe the strength of the variations of the signal relative to those expressed by the noise. SNR statistics are used to quantify diverse aspects of models where an observable quantity $Y$ is decom-
posed into a predictable or structural component $s$, often called signal or model, and a stochastic component $\varepsilon$, called noise or error. Although the definition of SNR is rather general in this paper we focus on a typical situation where one assumes a sequence $\{Y_i\}_{i \in \mathbb{Z}}$ is determined by

$$ Y_i := s(t_i) + \varepsilon_i, \quad (1) $$

where $i$ is a time index, $s(\cdot)$ is a smooth function of time evaluated at the time point $t_i$ with $i \in \mathbb{Z}$, and $\{\varepsilon_i\}_{i \in \mathbb{Z}}$ is some random sequence. Assume $t_i \in (0, 1)$, however if a time series is observed at time points $t_i \in (a, b)$, these can be rescaled onto the interval $(0, 1)$ without changing the results of this paper.

Equation (1) is a popular model in many applications that range from physical sciences to engineering, biosciences, social sciences, etc. (see Parzen 1966, 1999 and references therein). Although we use the conventional term “noise” for $\varepsilon_i$, this term may have a rich structure well beyond what we would usually consider noise. Some of the terminology here originates from physical sciences where the following concepts have been first explored. Consider a non stochastic signal $s(t)$ defined on the time interval $(0, 1)$, and assume that $s(t)$ has a zero average level (that is $\int_0^1 s(t)dt = 0$). The average “variation” (or magnitude) of the signal is quantified as

$$ P_{\text{signal}} := \int_0^1 s^2(t)dt. \quad (2) $$

In physical science terminology (2) is the average power of the signal, that is the “energy” contained in $s(\cdot)$ per time unit (if the reference time interval is $(a, b)$ the integral in (2) is divided by $(b - a)$). If the average signal level is not zero, $s(t)$ is centered on its mean value, and then (2) is computed. The magnitude, or the “power”, of the noise component is given by $P_{\text{noise}} := \text{Var}[\varepsilon_i]$. The SNR of the process is the ratio

$$ \text{SNR} := 10 \log_{10} \frac{P_{\text{signal}}}{P_{\text{noise}}}, \quad (3) $$

expressed in decibels unit. The SNR can also be defined as the ratio $(P_{\text{signal}}/P_{\text{noise}})$, however the decibel scale is more common. Low SNR implies that the strength of the random component of (1) makes the signal $s(\cdot)$ barely distinguishable from the observation of $Y_i$. On the other hand, high SNR means that the sampling about $Y_i$ will convey enough information about the predictable/structural component $s(\cdot)$.

In many analysis, SNR is a crucial parameter to be known. In radar detection applications (Richards 2014), speech recognition (Loizou 2013), audio and video applications of signal processing (Haykin and Kosko 2001), it is crucial to build filtering algorithms that are able to reconstruct $s(\cdot)$ with the largest possible SNR. In neuroscience there is strong interest in quantifying the SNR of signals produced by neurons activity. In fact, the puzzle is that single neurons seem to have low SNR meaning that they emit “weak signals” that are still processed so efficiently by the brain system (Czanner...
et al. 2015). In medical diagnostics, a physiological activity is measured and digitally sampled (e.g. fMRI, EEG, etc) with methods and devices that need to guarantee the largest SNR possible (Ullsperger and Debener 2010). The historic discovery of the first detection of a gravitational wave announced on 11 February 2016 has been made possible because of decades of research efforts in designing instruments and measurement methods able to work in an extremely low SNR environment (Kalogera 2017). These are just a few examples of the relevance of the SNR concept. The main goal of this paper is to define a SNR statistic, and provide an estimator for its distribution with proven statistical guarantees under general assumptions on the elements of (1).

Let $Y_n := \{y_1, y_2, \ldots, y_n\}$ be sample values of $Y$ observed at equally spaced time points $t_i = i/n$ for $i = 1, 2, \ldots, n$, with $t_i \in (0, 1)$. That is, in this work we focus on situations where $Y$ is sampled at constant sampling rate (also known as fixed frequency sampling, or uniform design), although the theory developed here can be extended to non-constant sampling rates. Let $\hat{s}(\cdot)$ and $\hat{\epsilon}$ be estimated quantities based on $Y_n$. Consider the observed SNR statistic

$$\hat{SNR} := 10 \log_{10} \left( \frac{\frac{1}{n} \sum_{i=1}^{n} \hat{s}_i^2 (\frac{i}{n})}{\frac{1}{m} \sum_{i=1}^{m} (\hat{\epsilon}_i - \bar{\hat{\epsilon}})^2} \right), \quad \text{with} \quad \bar{\hat{\epsilon}} = \frac{1}{m} \sum_{i=1}^{m} \hat{\epsilon}_i, \quad (4)$$

for some choice of an appropriate sequence $\{m\}$ such that $m \to \infty$ and $m/n \to 0$ as $n \to \infty$. In this paper we propose a subsampling strategy that consistently estimates the quantiles of the distribution of $\tau_m(\hat{SNR} - SNR)$ for an appropriate sequence $\{\tau_m\}$ (see Theorem 3). These quantiles are used to construct simple confidence intervals for the SNR parameter.

In most applications the observed $\{Y_i\}_{i \in \mathbb{Z}}$ is treated as “stable” enough so that smoothing methods (typically linear filtering) are applied to get $\hat{s}(\cdot)$ and the error terms, $\hat{\epsilon}_i$, $i = 1, 2, \ldots, n$. Therefore, the general practice is to divide the observed data stream into sequential blocks of overlapping observations of some length (time windowing), and for each block $\hat{SNR}$ is computed (see Haykin and Kosko 2001; Weinberg 2017). These SNR measurements are then used to construct its distribution to make inference statements about the underlying SNR. These windowing methods implicitly assume some sort of local stationarity and uncorrelated noise, but there is a lack of theoretical justification. However, data structures often exhibit strong time-variations and other complexities not consistent with these simplifying assumptions. To our knowledge, a framework and a method for estimating the distribution of SNR statistics like (4) with provable statistical guarantees does not exist in the literature.

The major contribution of this paper is a subsampling method for the approximation of the quantiles of the distribution of the centered statistic $\tau_m(\hat{SNR} - SNR)$. The method is based on the observation that $\tau_m(\hat{SNR} - SNR)$ can be decomposed into the sum of the following components:
\[
-10\tau_m \begin{bmatrix}
\log_{10} \left(1 + \frac{V_m - \sigma^2}{\sigma^2} \right) & -\log_{10} \left(1 + \frac{1}{n} \sum_{i=1}^{n} \hat{s}^2(i/n) - \int s^2(t)dt \right)
\end{bmatrix},
\]

where \( \hat{V}_m = m^{-1} \sum_{i=1}^{m} \left( \hat{\varepsilon}_i - \hat{\varepsilon} \right)^2 \). In (5) the two components reflect the power contribution of the signal estimated based on \( \hat{s}^2(\cdot) \), and the power contribution of and the error term estimated in terms of \( \hat{V}_m \). In practice, the proposed method, formalized in Algorithm 1, works as follows: (i) the observed time series is randomly divided into subsamples, that is random blocks of consecutive observations; (ii) in each subsample the estimates \( \hat{s}^2(\cdot) \) and \( \hat{V}_m \) in (5) are computed; (iii) finally these subsample estimates are used to approximate the distribution of \( \tau_m (SNR - SNR) \) and its quantiles.

Based on Altman (1990), a consistent kernel smoother with an optimal bandwidth estimator is derived. The smoother does not require any further tuning parameters, even though the stochastic structure here is richer than that considered in the original paper by Altman (1990). The subsampling procedure extends the contributions of Politis and Romano (1994) and Politis et al. (2001). The main difference with the classical subsampling is that the method proposed in this paper does not require computations on the entire observed time series. Therefore the kernel smoothing is performed subsample-wise, and this is particularly beneficial in applications when the sample size is significantly large. In sleep studies often an electrophysiological record (EEG) of the brain activity is performed in several positions of the scalp, where each sensor samples an electrical signal for 24h at 100Hz, implying \( n = 8,640,000 \) data points for each sensor (see Kemp et al. 2000). Music is usually recorded at 44.1Khz (ISO 9660), which implies that a stereo song of 5 min produces \( n = 26,460,000 \) data points. This approach has been explored by Coretto and Giordano (2017) for the estimation of the dynamic range of music signals. However, the work of Coretto and Giordano (2017) deals with noise structures less general than those studied here. A further original element of this work is that, although the setup for \( \{\varepsilon_i\}_{i \in \mathbb{Z}} \) does not exclude long memory regimes, the methods proposed do not require the identification of any long memory parameter.

The rest of the paper is organized as follows. In Sect. 2 we define and discuss the reference framework for \( \{Y_i\}_{i \in \mathbb{Z}} \). In Sect. 3 the main estimation Algorithm 1 is introduced. The smoothing step of Algorithm 1 is studied in Seciton 4, while the subsampling step is investigated in Sect. 5. In Sect. 6 we show finite sample results of the proposed method based on simulated data, moreover an application to real data is illustrated. Final remarks and conclusions are given in Sect. 8. All proofs are given in the final “Appendix”.

\( \text{Springer} \)
2 Setup and assumptions

The framework (1) underpins a popular strategy to study experiments where a continuous time (analog) signal \( s(\cdot) \) is sampled at fixed time points \( t_i \). The stochastic term \( \{\epsilon_i\}_{i \in \mathbb{Z}} \) represents various sources of randomness. In some cases, the source and the structure of the random component are known, but this does not apply universally. A ubiquitous assumption about \( \{\epsilon_i\}_{i \in \mathbb{Z}} \) is that it is white noise; sometimes the simplification pushes further towards Gaussianity (see Parzen 1966, and references therein). However, in various applications the evidence of departure from this simplicity it is quite rich.

The most elementary source of randomness is the quantization noise, i.e. the added noise introduced by the quantization of the signal. In Music, speech, EEG and many other applications, a voltage amplitude is recorded at fixed time intervals using a limited range of integer numbers. This is the so called Pulse Code Modulation (PCM), which is at the base of digital encoding techniques. The quantization noise is produced by the rounding error of the PCM sampling. Theoretically the quantization noise is a uniform white noise process, however, Gray (1990) showed that the structure of the quantization noise varies a lot across applications and measurement techniques, and often the white noise assumption is too restrictive. Apart from quantization noise, the recorded signal may be affected by a number of disturbances unrelated to the signal. Take for example an EEG acquisition where electrical noise from the power lines is injected into the measuring device. In speech recording microphones capture stray radio frequency energy. Another example is that of wireless signal transmission affected by multi-path interference, that is: waves bounce off of and around surfaces creating unpredictable phase distortions. Complex effects like these happen in radar transmission too, where it is well known that the Gaussian white noise assumption is generally violated (see Conte and Maio 2002, and references therein).

Sometimes the stochastic component does not only include unpredictable external artifacts. There are cases where the structure of \( \{\epsilon_i\}_{i \in \mathbb{Z}} \) is the result of several complex phenomena occurring within the system under study. In their pioneering works Voss and Clarke (1978, 1975) found evidence of \( 1/f \)–noise or similar fractal processes in recorded music. Similar evidence is documented in Levitin et al. (2012). \( 1/f \)–noise is a stochastic process where the spectral density follows the power law \( c|f|^{-\beta} \), where \( f \) is the frequency, \( \beta \) is the exponent, and \( c \) is a scaling constant. \( \beta = 1 \) gives pink noise, that is just an example of such processes. Depending on \( \beta \) these forms of noise are characterized by slowly vanishing serial correlations and/or what is known as long memory. Many electronic devices found in data acquisition instruments introduce \( 1/f \)-type noise (Kogan 1996; Weissman 1988; Kenig and Cross 2014). Evidence of departure from linearity and Gaussianity in the transient components of music recordings was also found in Brillinger and Irizarry (1998) and Coretto and Giordano (2017).

The main goal of this paper is to build an estimation method for the distribution of the SNR that works under the most general setting. Of course achieving universality is impossible, but here we set a model environment that is as rich as possible. Our model is restricted by the following assumptions:
Assumption A1 The function $s(\cdot)$ has a continuous second derivative.

Assumption A2 The sequence $\{\varepsilon_i\}_{i \in \mathbb{Z}}$ fulfills one of the following:

(SRD) $\{\varepsilon_i\}_{i \in \mathbb{Z}}$ is a strictly stationary and $\alpha$-mixing process with mixing coefficients $\alpha(k)$, $\text{E}[\varepsilon_i] = 0$, $\text{E}[\varepsilon_i^2]^{2+\delta} < +\infty$, and $\sum_{k=1}^{+\infty} \alpha^{\delta/(2+\delta)}(k) < \infty$ for some $\delta > 0$.

(LRD) $\varepsilon_i = \sum_{j=0}^{\infty} \psi_j a_{i-j}$ with $\text{E}[a_i] = 0$, $\text{E}[a_i^4] < \infty \ \forall i$, $\{a_i\} \sim i.i.d.$, $\psi_j \sim C_1 j^{-\eta}$ with $\eta = \frac{1}{2}(1 + \gamma_1)$, $C_1 > 0$ and $0 < \gamma_1 \leq 1$.

Assumption A1 reflects a common smoothness requirement for $s(\cdot)$ which does not need further discussion. In most applications, $s(\cdot)$ will represent the sum of possibly many harmonic components, or long term smooth trends. A2 sets a wide range of possible structures for the stochastic component. Two regimes are considered here: short range dependence (SRD) and long range dependence (LRD). SRD is a rather general $\alpha$-mixing assumption that allows to overcome the usual linear process assumption. The latter is essential to model fast decaying energy variations that is typical in some form of noise. Assumptions A1 and A2-SRD are also considered in Coretto and Giordano (2017) for the estimation of the dynamic range of music signals. However, in this paper, we are interested in SNR statistics, and we extend the analysis to the cases where LRD occurs. A2-LRD has the role to capture situations where the noise spectra shows long-range dependence; in practice this assumption accommodates the $1/f$-type noise. The LRD is controlled by $\gamma_1$ which is between zero and 1.

Note that A2-LRD assumes a linear structure while A2-SRD does not. SRD assumption allows for dependence, and the rate at which it vanishes it is controlled by $\delta$. Under SRD, in the infinite future, the terms of $\{\varepsilon_i\}_{i \in \mathbb{Z}}$ act as an independent sequence. Hence SRD can capture many different forms of dependence but not long memory features. Is the linearity structure of LRD a strong assumption for the long-memory cases? The class of long-memory linear processes is well known in the literature, and in most cases, LRD effects are more common to appear with a linear autoregressive structure. Moreover, A2-LRD is compatible with the classical parametric models for LRD, e.g. the well known ARFIMA class, already used to capture the $1/f$-noise phenomenon. One could overcome the linearity assumption in LRD but at the expense of serious technical complications. It is important to stress that we are not interested in identifying SRD-vs-LRD, and we want to avoid the additional estimation of the LRD order. The latter is crucial in most parametric models for LRD. Assumption A2 only defines plausible stochastic structures that can occur in the most diverse applications. Note that A2-LRD does not imply that $\{\varepsilon_i\}_{i \in \mathbb{Z}}$ is a Gaussian process or a function of it, as it is assumed in Jach et al. (2012) and Hall et al. (1998).

3 The smoothing-subsampling procedure

The SNR distribution is estimated performing Algorithm 1. This is a simple smoothing-subsampling procedure where for each subsample $P_{\text{signal}}$ is consistently estimated by $\hat{U}_{n,b,t}$, and $P_{\text{noise}}$ is estimated by $\hat{V}_{n,b_1,t}$ on a secondary subsample taken from the
Algorithm 1: blockwise smoothing

**input**: data \(\{y_1, y_2, \ldots, y_n\}\), constants \(K \in \mathbb{N}, b \in \mathbb{N}\), and \(b_1 = o\left(b^{2/5}\right)\)

**output**: quantiles of the SNR statistic

Draw without replacement and with uniform probability a random sample \(\mathcal{F}_K = \{t_1, t_2, \ldots, t_K\}\) from the set \(\{1, 2, \ldots, n - b + 1\}\).

**for** \(t \in \mathcal{F}_K\) **do**

Take the subsample \(\mathcal{Y}_t = \{y_t, y_{t+1}, \ldots, y_{t+b-1}\}\)

Based on kernel methods estimate \(s(\cdot)\) and the signal power

\[
\hat{U}_{n,b,t} = \frac{1}{b} \sum_{i=t}^{t+b-1} \left[ \hat{s} \left( \frac{i-t+1}{b} \right) \right]^2.
\]

Compute \(\tilde{\epsilon}_i = y_i - \hat{s}(i-t+1)/b\) for \(i = t, t+1, \ldots, t+b-1\).

Estimate the noise variance on \(b_1\) values

\[
\hat{\epsilon}_{b_1,t} = \frac{1}{b_1} \sum_{i=t}^{t+b_1-1} \tilde{\epsilon}_i, \quad \hat{V}_{n,b_1,t} = \frac{1}{b_1} \sum_{i=t}^{t+b_1-1} (\tilde{\epsilon}_i - \hat{\epsilon}_{b_1,t})^2.
\]

Compute the subsample SNR statistic

\[
\widehat{SNR}_{n,b,t} = 10 \log_{10} \left( \frac{\hat{U}_{n,b,t}}{\hat{V}_{n,b_1,t}} \right).
\]

**end**

Based on \(\{\widehat{SNR}_{n,b,t}\}_{t \in \mathcal{F}_K}\) compute quantiles as in (16)

Previous one. Details and theoretical motivation of the procedure will be treated in Sects. 4 and 5. The distribution is constructed for the SNR expressed in decibel scale.

The procedure is called “Monte Carlo” because the subsample selection is randomized. The latter reduces the huge number of subsamples to be explored. Note that here none of the calculations involve computations over the entire observed sample \(\mathcal{Y}_n\). The latter differs from the classical subsampling for time series data introduced in Politis and Romano (1994) and Politis et al. (2001). In the classical subsampling, one would estimate the variance of \(\{\epsilon_i\}_{i \in \mathbb{Z}}\) based on the entire sample. This would require that the estimation of \(s(\cdot)\) is performed globally on \(\mathcal{Y}_n\). In Algorithm 1 both \(s(\cdot)\), and the variance of \(\{\epsilon_i\}_{i \in \mathbb{Z}}\) in (7) are estimated blockwise. This blockwise smoothing strategy, where computations are performed only at the subsample level, has been proposed in Coretto and Giordano (2017). The advantages over the classical subsampling are twofold. First, thanks to the increased data acquisition technology, in most of the applications mentioned in Sect. 1, \(n\) scales in terms of millions or billions of data points. It is well known that kernel and other nonparametric smoothing methods become computationally intractable for such big sample sizes. In Algorithm 1 the computational complexity for the calculation of \(\hat{s}(\cdot)\) is governed by the subsample...
size $b$, which is chosen much smaller than $n$ (see Theorem 2). Second, the kind of signals we want to reconstruct may exhibit strong structural variations along the time axis, therefore, estimation of $s(\cdot)$ on the entire sample would require the use of optimal kernel methods with local bandwidth increasing the computational burden even more. Working on smaller data chunks allows treating the signal locally. Therefore, simpler kernel methods based on global bandwidth within the subsampled block are better suited to capture the local structure of the signal. Optimal estimation of $s(\cdot)$, and the random subsampling part of Algorithm 1 are developed in the next two sections.

4 Optimal signal reconstruction

Unless one has enough information about the shape of $s(\cdot)$, nonparametric estimators of functions with proven statistical properties are natural candidates to reconstruct the underlying signal. Our choice is the classical Priestley-Chao kernel estimator (Priestley and Chao 1972), because it can be easily optimized in regression models where the error is not necessarily uncorrelated. The estimator for $s(\cdot)$ is defined as

$$\hat{s}(t) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{t - i/n}{h}\right) y_i.$$  (9)

The following assumption involving the kernel function $K(\cdot)$ and the bandwidth $h$ is assumed to hold.

Assumption A3 $K(\cdot)$ is a density function with compact support and symmetric about 0. Moreover, $K(\cdot)$ is Lipschitz continuous of some order. The bandwidth $h \in H = [c_1 \Lambda_n^{-1/5}, c_2 \Lambda_n^{-1/5}]$, where $c_1 < c_2$ are two positive constants such that: $c_1$ is arbitrarily small, $c_2$ is arbitrarily large.

Define

$$\Lambda_n := \begin{cases} n & \text{if A2-SRD holds}, \\ \frac{n}{\log n} & \text{if A2-LRD holds with } \gamma_1 = 1, \\ n^{\gamma_1} & \text{if A2-LRD holds with } 0 < \gamma_1 < 1. \end{cases}$$  (10)

Whenever $n \to \infty$ it happens that $h \to 0$ and $\Lambda_n h \to \infty$.

There are a number of possible choices for $K(\cdot)$ satisfying A3, and we will use the Epanechnikov kernel for its well known efficiency properties. Setting an optimal bandwidth in (9) when the error term may be correlated requires special care. Here an optimal choice of $h$ is even more involved due to the fact that $\{\varepsilon_i\}_{i \in \mathbb{Z}}$ may follow either the SRD or the LRD regime. The sequence (10) has a role in managing this added complexity. Altman (1990) developed the Priestley-Chao kernel estimator (9) with dependent additive errors, and showed that under serial correlation standard bandwidth optimality theory does not apply. Altman (1990) proposed to estimate an optimal $h$ based on a cross-validation function accounting for the dependence.
A Monte Carlo subsampling method for estimating… 491

structure of $\{\varepsilon_i\}_{i\in\mathbb{Z}}$. Altman’s contribution deals with errors belonging to the class of linear processes with finite memory. Therefore, Altman’s assumptions do not allow the LRD case. Moreover, we consider the SRD assumption because it is typical for stochastic processes with a nonlinear model representation in time series framework. Finally, Altman (1990) assumes that the true autocorrelation function of $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ is known which is not the case in real world applications.

Let $\hat{\varepsilon}_i = y_i - \hat{s}(i/n)$, and define the cross-validation objective function

$$CV(h) = \left[ 1 - \frac{1}{nh} \sum_{j=-M}^{M} \mathcal{K} \left( \frac{j}{nh} \right) \hat{\rho}(j) \right]^{-2} \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i^2. \quad (11)$$

The optimal bandwidth is estimated by minimizing (11), that is

$$\hat{h} = \arg\min_{h\in H} CV(h).$$

The first term in (11) is the correction factor proposed by Altman (1990), but replacing the true unknown autocorrelations with their sample counterparts $\hat{\rho}(\cdot)$ up to the $M$th order. $M$ is an additional smoothing parameter, but Altman’s contribution does not deal with its choice. Consistency of the optimal bandwidth estimator is obtained if $M$ increases at a rate smaller than the product $nh$. As in Coretto and Giordano (2017) $M$ is chosen so that the following holds.

**Assumption A4** Whenever $n \to \infty$; then $M \to \infty$ and $M = O(\sqrt{nh})$.

Let $\text{MISE}(\hat{s}; h)$ be the mean integrated square error of $\hat{s}(\cdot)$, that is

$$\text{MISE}(\hat{s}; h) = \int_{0}^{1} \text{MSE}(\hat{s}(t); h) \, dt \quad \text{where} \quad \text{MSE}(\hat{s}(t); h) = E[(\hat{s}(t) - s(t))^2].$$

Let $h^*$ be the global minimizer of $\text{MISE}(\hat{s}; h)$. The next result states the optimality of the kernel estimator.

**Theorem 1** Assume A1, A2, A3 and A4. $\hat{h}/h^* \xrightarrow{p} 1$ as $n \to \infty$.

The previous result relates $\hat{h}$ to the optimal global bandwidth for which convergence rate is known, that is $O(\Lambda_n^{-1/5})$. Theorem 1 is equivalent to that given in Coretto and Giordano (2017), however, the difference here is that $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ may well follow LRD. Therefore, proof of Theorem 1 (given in the “Appendix”) needs some further developments.

**Remark 1** Theorem 1 improves the existing literature in several aspects. First of all, the proposed signal reconstruction is optimal (in the MISE sense) under both SRD and LRD. Its key feature is that one does not need to identify the type of dependence, that is SRD vs LRD. There are only two smoothing tunings: $h$ that is estimated optimally, and $M$ fixed according to A4. The SRD regime is already treated in Coretto and Giordano (2017). Regarding LRD, the result should be compared to Hall et al.
The advantages of our approach compared to the latter are: (i) the method is simplified by eliminating a tuning needed to deal with LRD, that is the block length for the leave-\(k\)-out cross-validation in Hall et al. (1995). This is because the Altman’s cross-validation correction in (11) already incorporates the dependence structure via \(\hat{\rho}(\cdot)\), and \(M\) is able to correct (11) without any further step identifying whether LRD or SRD occurs; (ii) here we do not assume existence of higher order moments of \(\{\varepsilon_i\}_{i\in\mathbb{Z}}\).

5 Monte Carlo approximation of the subsampling distribution

In this section we exploit the subsampling procedure underlying Algorithm 1. We call this procedure “Monte Carlo”, because it is based on a random selection of subsamples, and here we provide a Monte Carlo approximation of the subsampling distribution of the statistic of interest. Let us introduce the following quantities:

\[
V_n = \frac{1}{n} \sum_{i=1}^{n} (\varepsilon_i - \bar{\varepsilon})^2, \quad \text{with} \quad \bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i. \tag{12}
\]

Although the random sequence \(\{\varepsilon_i\}_{i\in\mathbb{Z}}\) is not observable, one can work with its estimate. Replace \(\varepsilon_i\) with \(\hat{\varepsilon}_i\) in the previous formula and obtain

\[
\hat{V}_n = \frac{1}{n} \sum_{i=1}^{n} (\hat{\varepsilon}_i - \bar{\hat{\varepsilon}})^2, \quad \text{with} \quad \bar{\hat{\varepsilon}} = \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i.
\]

The distribution of a proper scaled and centered \(\hat{V}_n\) can now be used to approximate the distribution of \(\tau_n(V_n - \sigma_{\varepsilon}^2)\) where \(\tau_n\) is defined in (13) and \(\sigma_{\varepsilon}^2 := \mathbb{E}[\varepsilon_i^2]\). One way to do this is to perform the subsampling as proposed in Politis et al. (1999) and Politis et al. (2001).

That is, for all blocks of observations of length \(b\) (subsample size) compute \(\hat{V}_n\). However the number of possible subsample is huge even for moderate \(n\). Moreover, in typical cases where \(n\) is of the order of millions or billions of samples, the computation of the optimal \(\hat{s}(\cdot)\) would require an enormous computer power. The problem is solved by performing the blockwise smoothing of Algorithm 1 proposed in Coretto and Giordano (2017). Therefore, the signal and the average error are estimated block-wise, so that the computing effort is only driven by \(b\). This allows making the algorithm scalable with respect to \(n\), a very important feature to process data from modern data acquisition systems. Here we investigate the theoretical properties of the estimation Algorithm 1. The formalization is similar to that given in Coretto and Giordano (2017), however, here we deal with a different target statistic, and we face the added complexity of the existence of LRD regimes in \(\{\varepsilon_i\}_{i\in\mathbb{Z}}\).
First define

$$
\tau_n := \begin{cases} 
    n^{1/2} & \text{if A2-SRD holds,} \\
    n^{1/2} & \text{if A2-LRD holds with } 1/2 < \gamma_1 \leq 1, \\
    \left( \frac{n}{\log n} \right)^{1/2} & \text{if A2-LRD holds with } \gamma_1 = 1/2, \\
    n^{\gamma_1} & \text{if A2-LRD holds with } 0 < \gamma_1 < 1/2.
\end{cases}
$$

(13)

At a given time point $t$ consider a block of observations of length $b$, and the statistics computed in Algorithm 1:

$$
V_{n,b,t} = \frac{1}{b} \sum_{i=t}^{t+b-1} (\varepsilon_i - \bar{\varepsilon}_{b,t})^2, \quad \text{and} \quad \hat{V}_{n,b,t} = \frac{1}{b} \sum_{i=t}^{t+b-1} (\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}_{b,t})^2,
$$

with $\bar{\varepsilon}_{b,t} = b^{-1} \sum_{i=t}^{t+b-1} \varepsilon_i$ and $\bar{\hat{\varepsilon}}_{b,t} = b^{-1} \sum_{i=t}^{t+b-1} \hat{\varepsilon}_i$. The empirical distribution functions of $\tau_n(V_n - \sigma_\varepsilon^2)$, based on the true and estimated noise, respectively, are given by

$$
G_{n,b}(x) = \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} \mathbb{I}\{\tau_b(V_{n,b,t} - V_n) \leq x\},
$$

and

$$
\hat{G}_{n,b}(x) = \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} \mathbb{I}\{\tau_b(\hat{V}_{n,b,t} - V_n) \leq x\}.
$$

$\mathbb{I}\{A\}$ denotes the usual indicator function of the set $A$. $\tau_b$ is defined in (13). Lemma 2 and 3 in the “Appendix” state that the subsampling based on statistic (12) is consistent under both A2-SRD and A2-LRD. Notice that results in Politis et al. (2001) can only be used to deal with SRD. The LRD treatment is inspired to Hall et al. (1998) and Jach et al. (2012). However, we improve upon their results in the sense that the Gaussianity assumption for $\varepsilon_t$ is avoided under A2-LRD with $1/2 < \gamma_1 \leq 1$. The quantiles of the subsampling distribution also converges to the quantiles of the asymptotic distribution of $\tau_n(V_n - \sigma_\varepsilon^2)$. This is a consequence of the fact that $\tau_n(V_n - \sigma_\varepsilon^2)$ converges weakly (see Remark 2). For $\gamma_2 \in (0,1)$ the quantities $q(\gamma_2)$, $q_{n,b}(\gamma_2)$ and $\hat{q}_{n,b}(\gamma_2)$ denote respectively the $\gamma_2$-quantiles with respect the distributions $G$, (see Remark 2), $G_{n,b}$ and $\hat{G}_{n,b}$ respectively. We adopt the usual definition that $q(\gamma_2) = \inf\{x : G(x) \geq \gamma_2\}$. Lemma 4 in the “Appendix” states the same consistency for the quantiles. The following remark covers the different cases (A2-SRD and A2-LRD) for the asymptotic distribution of $\tau_n(V_n - \sigma_\varepsilon^2)$.

**Remark 2** By A2 it can be shown that $\tau_n(V_n - \sigma_\varepsilon^2)$ converges weakly to a random variable with distribution, say $G(\cdot)$, where $\sigma_\varepsilon^2 = \mathbb{E}[\varepsilon_t^2]$. Under A2-SRD, $G(\cdot)$ is a Normal distribution. $G(\cdot)$ is still a Normal distribution under A2-LRD with $1/2 < \gamma_1 \leq 1$, which follows from Theorem 4 of Hosking (1996). The same Theorem implies that $G(\cdot)$ is Normal under A2-LRD with $\gamma_1 = 1/2$ when $\varepsilon_t$ is normally distributed. Moreover, $G(\cdot)$ is not Normal under A2-LRD with $0 < \gamma_1 < 1/2$.

© Springer
A variant is to reduce the number of subsamples by introducing a random block selection with $s(\cdot)$ estimated blockwise on subsamples of length $b$. Let $I_i, i = 1, \ldots, K$ be random variables indicating the initial point of every block of length $b$. We draw, without replacement with uniform probabilities, the sequence $\{I_i\}_{i=1}^{K}$ from the set $I = \{1, 2, \ldots, n - b + 1\}$. The empirical distribution function of the subsampling variances of $\hat{\epsilon}_t$ over the random blocks is

$$
\tilde{G}_{n,b}(x) = \frac{1}{K} \sum_{i=1}^{K} \mathbb{I}\left\{\tau_b \left( \hat{V}_{n,b,I_i} - V_n \right) \leq x \right\}.
$$

In order to get the consistency of the subsample procedure both in the SRD and LRD cases, we consider two subsamples. The first one has a length of $b$ and we use it to estimate the signal, that is $s(\cdot)$. Instead, the second subsample, which is a subset of the first, has a length $b_1 = o(b^{4/5})$ and we use this second subsample to estimate the variance and its distribution. The following result states the consistency of $\tilde{G}$ in approximating $G$.

**Theorem 2** Assume A1, A2, A3 and A4. Suppose that $\{a_t\}$, in A2, is Normally distributed when $0 < \gamma_1 \leq 1/2$. Let $\hat{s}(t)$ be the estimate of $s(t)$ on a subsample of length $b$. Let $n \to \infty$, $b \to \infty$, $b/n \to 0$, $b_1 = o(b^{4/5})$ and $K \to \infty$, then

$$
\sup_x \left| \tilde{G}_{n,b_1}(x) - G(x) \right| \overset{p}{\to} 0.
$$

Proof of Theorem 2 is given in the “Appendix”. In analogy with what we have seen before we also establish consistency for the quantiles of $\tilde{G}(\cdot)$. Let $\tilde{q}_{n,b_1}(\gamma_2)$ be the $\gamma_2$-quantile with respect to $\tilde{G}(\cdot)$.

**Corollary 1** Assume A1, A2, A3 and A4. Suppose that $\{a_t\}$, in A2, is Normally distributed when $0 < \gamma_1 \leq 1/2$. Let $\hat{s}(t)$ be the estimate of $s(t)$ on a subsample of length $b$. Let $n \to \infty$, $b \to \infty$, $b/n \to 0$, $b_1 = o(b^{4/5})$ and $K \to \infty$, then $\tilde{q}_{n,b_1}(\gamma_2) \overset{p}{\to} q(\gamma_2)$.

Proof of Corollary 1 is given in the “Appendix”.

**Remark 3** Note that the second subsample of length $b_1$ is a consequence of the optimal rate for the estimation of $s(t)$ subsample-wise.

The next result states the consistency of subsample procedure when $\hat{V}_n$ replaces $V_n$ in $\tilde{G}_{n,b}(x)$. Define

$$
\tilde{G}_{n,b}^0(x) = \frac{1}{K} \sum_{i=1}^{K} \mathbb{I}\left\{\tau_b \left( \hat{V}_{n,b,I_i} - \hat{V}_n \right) \leq x \right\}.
$$

**Corollary 2** Assume the same assumptions as in Theorem 2. Let $\hat{s}(t)$ be the estimate of $s(t)$ on a subsample of length $b$. Let $n \to \infty$, $b \to \infty$, $b/n \to 0$, $b_1 = o(b^{4/5})$ and $K \to \infty$, then

$$
\sup_x \left| \tilde{G}_{n,b_1}^0(x) - G(x) \right| \overset{p}{\to} 0.
$$
Proof of Corollary 2 is given in the “Appendix”.

**Remark 4** Following the same arguments as in the proof of Corollary 2, we have that $V_n - \sigma_e^2 = O_p(\tau_n^{-1})$ and also $\tilde{V}_n - \sigma_e^2 = O_p(\tau_n^{-1})$ in the cases of SRD and LRD with $5/8 \leq \gamma_1 \leq 1$. Moreover, in the proof of Theorem 2 the key point for the consistency is $\tau_{b1} \Lambda_b^{-4/5} \rightarrow 0$ when $n \rightarrow \infty$. Therefore, if we set $b_1 = b$ the results of Theorem 2 and Corollary 2 hold again in the cases of SRD and LRD with $5/8 \leq \gamma_1 \leq 1$. In fact, $\tau_b \Lambda_b^{-4/5} \rightarrow 0$ and $\tau_b \tau_n^{-1} \rightarrow 0$ when $n \rightarrow \infty$.

Now, by using the previous results, we can state that the subsample strategy is consistent to estimate the asymptotic distribution of $\tau_m(\hat{SNR} - SNR)$ where $\hat{SNR}$ is defined in (4). The statistic $\hat{SNR}$ has the numerator and denominator depending on $n$ and $m$, respectively. The latter is mimicked in the subsample procedure. In fact, a subsample of length $b$ is used for the estimation of the signal power, while a subsample of length $b_1$ is used to estimate the variance of the error term.

**Theorem 3** Let $\mathcal{I}_n := \{y_1, y_2, \ldots, y_n\}$ be a sampling realization of $\{Y_i\}_{i \in \mathbb{Z}}$. Assume A1, A2, A3 and A4. Suppose that $\{a_i\}$, in A2, is Normally distributed when $0 < \gamma_1 \leq 1/2$. Assume $n \rightarrow \infty$, $b \rightarrow \infty$, $b/n \rightarrow 0$, $b_1 = o(b^{2/5})$, $m = o(n^{2/5})$, $b_1/m \rightarrow 0$ and $K \rightarrow \infty$, then

$$\sup_x |\mathbb{Q}_n(x) - \mathbb{Q}(x)| \xrightarrow{p} 0$$

where

$$\mathbb{Q}_n(x) := \frac{1}{K} \sum_{i=1}^{K} \mathbb{I} \left\{ \tau_{b1}(\hat{SNR}_{n,b,l_i} - SNR) \leq x \right\},$$

with $\hat{SNR}_{n,b,l_i} := 10 \log_{10} \left( \frac{\hat{U}_{n,b,l_i}}{\hat{V}_{n,b_l,l_i}} \right)$

and $\mathbb{Q}(x)$ is the asymptotic distribution of $\tau_m(\hat{SNR} - SNR)$.

Proof of Theorem 3 is given in the appendix. Note that in Theorem 3 we need $b_1 = o(b^{2/5})$ instead of $b_1 = o(n^{4/5})$ found in previous results. The reason for this is that the statistical functional $\hat{SNR}$ is more complex than $\hat{V}_n$ and a different relative speed for the secondary block size $b_1$ is required.

Theorem 3 provides the theoretical justification for the consistency of the subsample procedure with respect to the statistic $\hat{SNR}$. Let $q^Q(\gamma_2)$ and $\tilde{q}^Q_{n,b_1}(\gamma_2) \equiv \tilde{q}^Q_{n,b_1}(\gamma_2|\tau_{b1})$ be the quantiles with respect to $\mathbb{Q}(x)$ and $\mathbb{Q}_n(x)$, respectively. Note that we write $\tilde{q}^Q_{n,b_1}(\gamma_2|\tau_{b1})$ to highlight the dependence on the scaling factor $\tau_{b1}$ as in Sect. 8 of Politis et al. (1999). The main goal is to do inference for $SNR$ without estimating the long memory parameter, and without using the sample statistic $\hat{SNR}$. In this way, we do not need to fix or estimate $m$. To do this, we use Lemma 8.2.1 in Politis et al. (1999). First, $\mathbb{Q}(x)$ always has a strictly positive density function, at least, in a subset.
of real line (see Hosking (1996) and references therein). So, by Lemma 8.2.1 in Politis et al. (1999), and using the same arguments as in the proof of Corollary 1, we have that

$$\tilde{q}_{n,b_1}^Q (\gamma_2 | \tau_{b_1}) = q^Q (\gamma_2) + o_p(1). \quad (14)$$

Following the same lines as in Sect. 8 of Politis et al. (1999), we have that

$$\tilde{q}_{n,b_1}^Q (\gamma_2 | 1) = \frac{\tilde{q}_{n,b_1}^Q (\gamma_2 | \tau_{b_1})}{\tau_{b_1}} + \widehat{SNR}. \quad (15)$$

Note that $\tilde{q}_{n,b_1}^Q (\gamma_2 | 1)$ is the quantile with respect to the empirical distribution function $1/K \sum_{i=1}^{K} I (\frac{\hat{V}_n}{\tau_{b_1}} \leq \frac{\hat{V}_n}{\hat{SNR}}).$ Therefore, by (14) and (15) it follows that

$$\tilde{q}_{n,b_1}^Q (\gamma_2 | 1) \tau_{b_1} = \tau_{b_1} \widehat{SNR} + q^Q (\gamma_2) + o_p(1).$$

Since $\widehat{SNR} = SNR + O_p(\tau_m^{-1})$ and $\tau_{b_1}/\tau_m \to 0$ when $n \to \infty$, we have that

$$\tilde{q}_{n,b_1}^Q (\gamma_2 | 1) = SNR + \frac{q^Q (\gamma_2)}{\tau_{b_1}} + o_p(\tau_{b_1}^{-1}).$$

Therefore, a confidence interval for $SNR$ with a nominal level of $\gamma_2$ is given by

$$\left[ \tilde{q}_{n,b_1}^Q (\gamma_2/2 | 1), \tilde{q}_{n,b_1}^Q (1 - \gamma_2/2 | 1) \right]. \quad (16)$$

It is possible to consider the methods of self-normalization as in Jach et al. (2012), and the estimation of the scaling factor $\tau$ as in Politis et al. (1999). These methods would lead to more efficient confidence bands, in the sense that these would be first order correct with a rate of $\tau_m^{-1}$ instead of $\tau_{b_1}^{-1}$. However, this would require the estimation of the unknown constants as in Jach et al. (2012).

**Remark 5** In Theorems 2, Corollary 2 and Theorem 3 the definition of $\tilde{G}(\cdot), \tilde{G}^0(\cdot)$ and $Q_n(\cdot)$ depend on quantities ($V_n, \hat{V}_n$ and $\hat{SNR}$) computed on the whole sample. On the other hand, these results give the theoretical framework for computing confidence interval as in (16), and these calculations will not require any calculation on the entire sample. In other words, $V_n, \hat{V}_n$, and $\hat{SNR}$ are needed to center the involved distributions, but not needed to approximate the quantiles as in (16). Therefore, in this work these quantities only have a theoretical role to show that the subsample procedure does not produce degenerate asymptotic distributions.

### 6 Numerical experiments

In this section we present numerical experiments on simulated data. The assumptions given in this paper are rather general, and it is not possible to design a computer
experiment that can be considered representative of all the kind of structures consistent with A1–A4. Here we assess the performance of Algorithm 1 under different scenarios for the structure of the noise term. In order to do this we keep the structure of true signal fixed, and we investigate three variations of the noise data generating process. Data are sampled at fixed sampling frequency set at $F_s = 44100$Hz, a common value in audio applications. Let $[0, T]$ be the data acquisition time interval, where $T$ is the duration of the simulated signal in seconds. The signal is sampled at time $t_i = (i - 1)/F_s$ for $i = 1, 2, \ldots, T \times F_s$, as follows

$$y_i = A_s \sin(2\pi \times 50 t_i) + \epsilon_i,$$

and $t_i = i/(T \times F_s)$. Therefore, the signal consists of a sinusoidal wave that produces energy at 50Hz. The signal power is equal to $A_s^2/2$, where $A_s$ is a scaling constant properly tuned to achieve a given true SNR. We set $T = 30$sec (implying $n = 1, 323, 000$), and we consider the following three cases for the noise.

**AR**  The noise is generated from an AR(1) process with independent normal innovations. This produces serial correlation in the error term and represents a case for SRD. In particular $\epsilon_i = -0.7\epsilon_{i-1} + u_i$, where $\{u_i\}$ is an i.i.d. sequence with distribution Normal$(0, A_{\epsilon})$, where $A_{\epsilon}$ is set to achieve a certain SNR.

**P1**  The random sequence $\{\epsilon_i\}$ has power spectrum equal to $P(f) = A_{\epsilon}/f^\beta$, where $P(f)$ is the power spectral density at frequency $f$Hz. Here $\beta = 0.2$ which induces some moderate LRD in $\{\epsilon_i\}$. The scaling constant $A_{\epsilon}$ is set to achieve the desired SNR.

**P2**  same as P1 but with $\beta = 0.6$. This design introduces a much stronger LRD.

In **P1** and **P2** the noise has a so-called $1/f^\beta$-“power law” where $\beta$ controls the amount of long range dependence. Larger values $\beta$ implies slower rate of decays for the serial correlations. For $\beta = 1$ pink noise is obtained. Values of $\beta \in [0, 1]$ give a behavior between the white noise and the pink noise. In the case **P1**, $\gamma_1 = 1 - \beta = 0.8$ in A2-LRD. So, the asymptotic distribution of $\tau_n(V_n - \sigma_{\epsilon}^2)$ is Normal. Whereas, in the case **P2**, $\gamma_1 = 1 - \beta = 0.4$. This implies that the asymptotic distribution of $\tau_n(V_n - \sigma_{\epsilon}^2)$ is not Normal (see Remark 2).

**P1** and **P2** are simulated based on the algorithm of Timmer and König (1995) implemented in the tuneR software of Ligges et al. (2016). For each of the three sampling designs we consider two values for the true SNR: 10dB and 6dB. In most applications an SNR = 6dB is considered a rather noisy situations. We recall that at 6dB the signal power is circa only four times the variance of the noise, and 10dB means that the signal power is ten times the noise variance. There is a challenging aspect of these designs. The case with **P2** noise and SNR = 6dB, is particularly difficult for our method. In fact, **P2** puts relatively large amount of variance (power) at low frequencies around 50Hz, so that the signal is not well distinguished from some spectral components of the noise. The two parameters of Algorithm 1 are $b$ and $K$.

We consider three settings for the subsample window: $b = 10$ms = 441 samples, $b = 15$ms = 662 samples, and $b$ estimated based on the method proposed in Götze and Račkauskas (2001). In the latter case the optimal $b$ is computed over a grid ranging from $b = 2$ms to $b = 20$ms. In many applications is not easy to fix a value for $b$. 
Table 1  Monte Carlo averages for the Mean Square Error (MSE) of the estimated signal power

| Error model | True SNR | $b$ (ms) | MSE       |
|-------------|----------|----------|-----------|
| AR          | 6        | 10       | 0.0016 (0.000010) |
| AR          | 6        | 15       | 0.0008 (0.000005) |
| AR          | 6        | Opt      | 0.0016 (0.000030) |
| AR          | 10       | 10       | 0.0046 (0.000024) |
| AR          | 10       | 15       | 0.0022 (0.000012) |
| AR          | 10       | Opt      | 0.0049 (0.000074) |
| P1          | 6        | 10       | 0.0012 (0.000008) |
| P1          | 6        | 15       | 0.0020 (0.000024) |
| P1          | 6        | Opt      | 0.0012 (0.000008) |
| P1          | 10       | 10       | 0.0051 (0.000030) |
| P1          | 10       | 15       | 0.0037 (0.000024) |
| P1          | 10       | Opt      | 0.0049 (0.000058) |
| P2          | 6        | 10       | 0.0024 (0.000018) |
| P2          | 6        | 15       | 0.0020 (0.000015) |
| P2          | 6        | Opt      | 0.0025 (0.000027) |
| P2          | 10       | 10       | 0.0060 (0.000038) |
| P2          | 10       | 15       | 0.0051 (0.000036) |
| P2          | 10       | Opt      | 0.0060 (0.000056) |

Standard errors for the Monte Carlo averages are given in parenthesis. The block length $b$ is expressed in milliseconds [ms]. The case when $b = \text{opt}$ corresponds to the estimated optimal block length reported in Table 2.

However, in certain situations researchers have an idea about the structure of the signal, and the time series is windowed with blocks of a certain length. In applications where the underlying signal is expected to be composed by harmonic components, the usual practice is to take blocks of size approximately equal to the period of the harmonic component with the lowest expected frequency. The rationale is to take the smallest window size so that each block is still expected to carry some information about the low frequency components. For example for speech data usually blocks of 10ms are normally considered (Haykin and Kosko 2001), whereas for music data 50ms is a common choice (Weihs et al. 2016). Note that the artificial data here have an harmonic component at 50Hz with a period of 20ms, and we consider the fixed alternatives $b = 10$ms and $b = 15$ms as a robustness check.

We set $K = 200$, of course larger values of $K$ would ensure less subsample induced variability. The $b_1$, i.e. the window length of the secondary subsample needed to estimate the distribution of the sampling variance, is set according to Theorem 2. This is achieved by setting $b_1 = \left[b^{2/5}\right]$. For each combination of noise type, SNR, and $b$ we considered 500 Monte Carlo replica and we computed statistics to assess the performance of the procedure. Two aspects of the method are investigated corresponding to the two main contributions of the paper.

The first contribution of the paper is Theorem 1, where optimality and consistency of the Priestley-Chao kernel estimator is established under rather general assumptions
Table 2  Monte Carlo averages for the optimal \( b \) estimated using the method proposed in Götte and Račkauskas (2001)

| Error model | True SNR | Optimal \( b \) [ms] |
|-------------|----------|----------------------|
| AR          | 6        | 9.44 (0.193)         |
| AR          | 10       | 8.93 (0.164)         |
| P1          | 6        | 8.18 (0.168)         |
| P1          | 10       | 11.05 (0.162)        |
| P2          | 6        | 9.36 (0.198)         |
| P2          | 10       | 10.80 (0.172)        |

Standard errors for the Monte Carlo averages are given in parentheses. The block length \( b \) is expressed in milliseconds [ms].

on the error term. The kernel smoothing is used in Algorithm 1 to estimate the signal power in the numerator of (8). In Table 1 we report the Monte Carlo averages for the Mean Square Error (MSE) of the estimated signal power. Going from the simplest AR to the complex P2 noise model there is an increase in MSE as expected. The longer \( b = 15 \text{ ms} \) subsample window always produced better results. The apparently counterintuitive evidence is that for larger amount of noise (lower SNR), the signal power is slightly better estimated. In order to understand this, note that the noise (in all three cases) produces most of its power in a low frequency region containing the signal frequency (i.e. 50Hz). In the lower noise case there is still a considerable amount of noise acting at low frequency that the adaptive nature of the kernel smoother is not able to recognize properly. In Table 2 we report the estimated average \( b \) with its Monte Carlo standard error. The estimated \( b \) is always near 10ms, and the latter produced results that are only slightly worse than those obtained for fixed \( b = 15 \text{ ms} \).

The second contribution of the paper is the consistency result (see Theorem 3 and related results) for the distribution of the SNR statistic. In order to measure the quality of method one needs to define the ground truth in terms of the sampling distribution of the target SNR statistic. The derivation of an expression for such a distribution would be an analytically intractable. Therefore, we computed the quantiles of the true SNR statistic based on Monte Carlo integration, and in Table 3 we report the average absolute differences between estimated quantiles and the true counterpart. Based on Corollary 1 the convergence of the distribution of the SNR is mapped into its quantiles, therefore this makes sense. Comparison involves five different quantile levels to assess the behavior of the procedure both in the tails and in the center of the distribution. The average deviations of Table 3 are computed in decibels. Overall the method can capture the center of the distribution pretty well in all cases. The estimation error increases in the tails of the distribution as one would expect. The right tail is estimated better than the left tail. In all cases the performance in the tails of the SNR distribution is better captured with a \( b = 15 \text{ ms} \) window, although in the center of the distribution the differences implied by different values of \( b \) are much smaller. Going from \( \text{SNR} = 6 \) to \( \text{SNR} = 10 \) results are clearly better on the left tail of the distribution especially in the case P2. Again the estimated version of \( b \) pushes the corresponding results towards the \( b = 10 \text{ ms} \) case.

Every method has its own tunings, and the evidence here is that \( b \) has some effects on the proposed method. The major impact of \( b \) is about the tails of the SNR distribution.
### Table 3  Monte Carlo averages for the absolute deviation of the estimated quantiles of the SNR distribution from the true counterpart

| Error model | True SNR | Quantile level [ms] | 0.1 | 0.25 | 0.5 | 0.75 | 0.90 |
|-------------|----------|---------------------|-----|------|-----|------|-----|
| AR 6        | 10       | 4.99 (0.013)        | 2.99 (0.022) | 0.28 (0.010) | 1.50 (0.008) | 2.09 (0.007) |
| AR 6        | 15       | 1.99 (0.008)        | 1.30 (0.011) | 0.22 (0.007) | 0.80 (0.006) | 1.13 (0.006) |
| AR 6        | Opt      | 4.29 (0.125)        | 2.19 (0.059) | 0.28 (0.010) | 0.94 (0.017) | 1.38 (0.025) |
| AR 10       | 10       | 5.00 (0.012)        | 3.04 (0.021) | 0.28 (0.010) | 1.49 (0.008) | 2.08 (0.007) |
| AR 10       | 15       | 2.01 (0.007)        | 1.30 (0.011) | 0.19 (0.007) | 0.81 (0.006) | 1.13 (0.006) |
| AR 10       | Opt      | 4.70 (0.113)        | 2.43 (0.053) | 0.30 (0.011) | 1.01 (0.014) | 1.45 (0.021) |
| P1 6        | 10       | 5.37 (0.019)        | 3.09 (0.024) | 0.34 (0.011) | 1.87 (0.007) | 2.52 (0.006) |
| P1 6        | 15       | 2.41 (0.015)        | 1.26 (0.013) | 0.37 (0.010) | 1.18 (0.007) | 1.68 (0.007) |
| P1 6        | Opt      | 4.58 (0.094)        | 2.13 (0.047) | 0.39 (0.013) | 1.27 (0.015) | 1.76 (0.019) |
| P1 10       | 10       | 5.60 (0.022)        | 3.07 (0.026) | 0.39 (0.012) | 1.93 (0.009) | 2.56 (0.007) |
| P1 10       | 15       | 2.77 (0.025)        | 1.32 (0.015) | 0.49 (0.013) | 1.35 (0.007) | 1.86 (0.008) |
| P1 10       | Opt      | 4.46 (0.087)        | 2.20 (0.048) | 0.45 (0.012) | 1.54 (0.012) | 2.05 (0.013) |
| P2 6        | 10       | 5.19 (0.031)        | 2.36 (0.026) | 0.69 (0.016) | 2.52 (0.011) | 3.58 (0.012) |
| P2 6        | 15       | 2.83 (0.028)        | 0.83 (0.019) | 0.99 (0.015) | 2.17 (0.012) | 3.05 (0.013) |
| P2 6        | Opt      | 4.00 (0.082)        | 1.41 (0.043) | 0.81 (0.017) | 2.02 (0.018) | 2.89 (0.020) |
| P2 10       | 10       | 5.16 (0.035)        | 2.02 (0.031) | 1.03 (0.017) | 2.65 (0.011) | 3.56 (0.012) |
| P2 10       | 15       | 2.82 (0.038)        | 0.64 (0.018) | 1.39 (0.016) | 2.37 (0.011) | 3.07 (0.011) |
| P2 10       | Opt      | 4.29 (0.082)        | 1.46 (0.049) | 1.12 (0.020) | 2.32 (0.014) | 3.05 (0.015) |

SNR is expressed in decibels. The true quantiles of the SNR distribution are computed based on Monte Carlo integration. Standard errors for the Monte Carlo averages are given in parenthesis. The block length $b$ is expressed in milliseconds [ms]. The case when $b = \text{opt}$ corresponds to the estimated optimal block length reported Table 2.

The selection of $b$ based on the method proposed by Götze and Račkauskas (2001) deliver a fully satisfying solution that does not require any prior knowledge on the data structure. The only drawback of estimating $b$ is that the overall algorithm needs to be executed for several candidate values of $b$. As final remark we want to stress that the method proposed here is designed to cope with much larger values of $n$. In this experiment the sampling is repeated a number of times to produce Monte Carlo estimates, therefore we had to choose an $n$ compatible with reasonable computing times according to the available hardware. A limited number of trials with $T$ up to several minutes (which implies that $n$ goes up to several millions) have been successfully tested without changing the final results. Therefore, we can conclude that the algorithm scales well with the sample size.

### 7 Application to EEG data

In this section we illustrate an application of the proposed methodology to electroencephalography (EEG) data obtained from the PhysioNet repository (Goldberger et al.).
Fig. 1 Time series plots of the amplitude of the EEG signals recorded in positions P8-02 and T8-P8 for two distinct subjects in the experiment. The duration of the fragment is 5 sec, and it starts at 12 min from the beginning of the recording.

In particular we considered the “CHB-MIT Scalp EEG Database” available at https://physionet.org/pn6/chbmit/. The database contains EEG traces recorded at the Children’s Hospital Boston on pediatric subjects with intractable seizures. Subjects were monitored for several days after the withdrawal of anti-seizure medication before the final decision about the surgical intervention. 22 subjects were traced during the experiment for several days using the international 10–20 EEG system. The latter is a standard that specifies electrode positions and nomenclature. Therefore, for each subject 21 electrodes have been placed in certain positions of the scalp, each of these electrodes produced an electric signal sampled at 256 Hz and measured with 16 bit precision. This means that each day (24 h), the EEG machine produced 21 time series each containing $n = 22 \times 118 \times 400$ data points for a total of 464,486,400 amplitude measurements for each subject in the experiment. A description of the “CHB-MIT Scalp EEG Database”, as well as details about the data acquisition is given in Shoeb (2009).

EEG signals have complex structures. Various sources of noise can be injected in the measurement chain, therefore it is always of interest to understand the behavior of the SNR. For this application we considered data for the first 3 subjects of the database, and we considered two electrode positions labeled P8-02 and T8-P8 in the 10–20 EEG system. The P8-02 electrode is placed on the parietal lobe responsible for integrating sensory information of various types. The T8-P8 electrode is placed on the temporal lobe which transforms sensory inputs into meanings retained as visual memory, language comprehension, and emotion association. An example of these traces is given in Fig. 1.

The method proposed here has been applied to obtain confidence intervals for the SNR. An SNR $\geq 10$ dB can be considered a requirement for a favorable noise floor in these applications. In order to assess the robustness of the procedure with respect to the...
choice of the subsampling window $b$, for each case we considered windows of fixed size $b = \{3\text{sec}, 5\text{sec}, 7\text{sec}\}$ which means $b = \{768, 1280, 1792\}$, plus the estimated $b$ with the method proposed by Götze and Račkaukas (2001). The estimation of $b$ is performed on a grid of equispaced points between 2sec and 10sec. The literature about EEG signals doesn’t tell us whether the processes involved have a clear time scale, but 5sec is considered approximately the time length needed to identify interesting cerebral activities. For each $b$ the corresponding $b_1$ is set according to $b_1 = [b^{2/5}]$ as for the numerical experiments. In Table 4 lower and upper limits for 90% and 95% confidence intervals of the SNR are reported.

Overall the results with estimated $b$ are comparable to those with fixed $b$. The upper limits of these confidence intervals is never smaller than 10dB. The lower limits is negative in all cases, which means that for all cases there is a chance that the power of the stochastic component dominates that of the deterministic component in model (1). While the upper limit of these confidence intervals is rather stable across units for the same $b$ value, larger differences are observed in terms of the lower limit. All this is a clear indication of the asymmetry of the SNR statistic. But this is expected since the two tails of the SNR statistic reflects two distinct mechanisms. In fact, a negative value for the SNR statistic (left tail) corresponds to situations where the dynamic of the observed time series is driven by the error term of equation (1). On the other hand a positive value of the SNR statistic (right tail) corresponds to situations where the dynamic is driven by the smooth changes induces by $s(\cdot)$. Ceteris paribus, going from 90% level to 95% does not change the results dramatically. Note that in this kind of applications a 3dB difference is not considered a large difference. Regarding data recorded in the P8-02 position the length of the confidence interval, going from 90% to 95% changes between 1.28dB to 3.6dB, where the maximum variation is measured for Subject 3 when $b = 7$sec. For the T8-P8 case the length of the confidence interval, going from 90% to 95%, changes between 1.52dB to 3.83dB, and here the maximum variation is measured for Subject 2 when $b = 7$sec.

Some pattern is observed across experimental units. For given confidence level and $b$, overall Subject 1 reports the shortest confidence intervals. Subject 2 reports the longest intervals for records in position P8-02. Subject 3 reports the longest intervals in position T8-P8. The variations across values of $b$, with all else equal, are not dramatic. The settings with $b = 3$, and $b = 5$, produced longer intervals if compared with $b = 7$sec and the optimal $b$. The data-driven method of Götze and Račkaukas (2001) produced an estimated $b$ in the range [3sec, 7sec] for P8-02 data, and [6sec, 8sec] for T8-P8 data. These values are comparable with the rule of thumb that 5sec is a reasonable time scale for the kind of signals involved here. The general conclusion is that in absence of relevant information, the method of Götze and Račkaukas (2001) gives a useful data-driven choice of $b$.

8 Conclusions and final remarks

In this paper we developed an estimation method that consistently estimates the distribution of a SNR statistic in the context of time series data with errors belonging to a rich class of stochastic processes. We restricted the model to the case where the signal
Table 4  Lower and upper limit of the confidence interval for the SNR (expressed in dB) of the EEG records in positions P8-02 and T8-P8

| Confidence level (%) | Subject | b [s] | P8-02 | T8-P8 |
|----------------------|---------|-------|-------|-------|
|                      |         |       | Lower | Upper | Lower | Upper |
| 90                   | 1       | 3     | -0.93 | 11.18 | -2.85 | 11.92 |
|                      | 5       | 1.14  | 11.26 | -3.41 | 11.92 |
|                      | 7       | 0.86  | 12.25 | -3.04 | 12.76 |
|                      | Opt     | -2.24 | 12.52 | -2.56 | 14.70 |
| 51.54                | 5       | 3     | -9.04 | 14.93 | -8.38 | 14.00 |
|                      | 7       | -8.29 | 15.45 | -7.41 | 12.53 |
|                      | Opt     | -8.91 | 15.68 | -7.57 | 12.42 |
| 70.86                | 3       | 3     | -8.30 | 13.06 | -7.14 | 16.50 |
|                      | 5       | -8.30 | 13.41 | -5.37 | 16.97 |
|                      | 7       | -5.81 | 13.01 | -5.01 | 15.38 |
|                      | Opt     | -7.24 | 13.49 | -6.70 | 15.66 |
| 50.50                | 1       | 3     | -1.36 | 12.19 | -3.81 | 13.19 |
|                      | 5       | 0.50  | 13.00 | -4.30 | 12.93 |
|                      | 7       | 0.49  | 15.45 | -3.63 | 14.81 |
|                      | Opt     | -3.11 | 14.36 | -3.48 | 15.30 |
| 70.49                | 2       | 3     | -9.65 | 15.60 | -9.54 | 15.02 |
|                      | 5       | -10.47| 16.46 | -8.98 | 14.02 |
|                      | 7       | -8.96 | 16.17 | -8.96 | 14.81 |
|                      | Opt     | -10.62| 16.81 | -8.16 | 13.49 |
| 70.49                | 3       | 3     | -9.60 | 15.04 | -8.09 | 18.22 |
|                      | 5       | -10.66| 14.44 | -6.80 | 18.68 |
|                      | 7       | -7.46 | 14.96 | -6.58 | 16.07 |
|                      | Opt     | -9.75 | 14.23 | -7.94 | 17.05 |

The block length $b$ is expressed in seconds. The notation $b = \text{opt}$ denotes the case when $b$ is estimated.

is a smooth function of time. The theory developed here can be easily adapted to more general time series additive regression models. The reference model for the observed data, and the theory developed here adapts to many possible applications that will be the object of a distinct paper. In this work we concentrated on the theoretical guarantees of the proposed method. The estimation is based on a random subsampling algorithm that can cope with massive sample sizes. Both the smoothing, and the subsampling techniques at the earth of Algorithm 1 embodies original innovations compared to the existing literature on the subject. Numerical experiments described in Sect. 6 showed that the proposed algorithm performs well in finite samples.

Acknowledgements  We thank the editor and two anonymous reviewers for their constructive comments, which helped to improve the manuscript.
Appendix

In this section we report the proofs of statements and some useful technical lemmas. First, we state a lemma to evaluate the MISE ($\hat{s}; h$).

**Lemma 1** Assume A1, A2 and A3. For $t \in I_h = (h, 1 - h)$

\[
MISE(\hat{s}; h) = \frac{h^4 R(s'')}{4} d_K + \sigma_\varepsilon^2 N_K \frac{S_{\rho}^*}{\Lambda_n h} + o\left(\frac{1}{\Lambda_n h} + h^4\right),
\]

where $\hat{s}$ is the kernel estimator in (9), $R(s'') = \int_{I_h} [s''(t)]^2 dt$, $d_K = \int u^2 \mathcal{X}(u) du$, $N_K = \int \mathcal{X}^2(u) du$, $\sigma_\varepsilon^2 = \mathbb{E}[\varepsilon_t^2]$, $\Lambda_n$ is defined in (10) and

\[
S_{\rho}^* := \begin{cases} 
\lim_{n \to \infty} \sum_{j=1}^n \rho(j) & \text{if A2-SRD holds} \\
\lim_{n \to \infty} \frac{1}{\log n} \sum_{j=1}^n \rho(j) & \text{if A2-LRD holds with } \gamma_1 = 1 \quad (17) \\
\lim_{n \to \infty} \frac{1}{n^{1-\gamma_1}} \sum_{j=1}^n \rho(j) & \text{if A2-LRD holds with } 0 < \gamma_1 < 1.
\end{cases}
\]

**Proof** By A3 it follows that conditions A–C of Altman (1990) are satisfied. Now, let

\[
\rho_n(j) := \begin{cases} 
\rho(j) & \text{if A2-SRD holds} \\
\frac{1}{\log n} \rho(j) & \text{if A2-LRD holds with } \gamma_1 = 1 \\
\frac{1}{n^{1-\gamma_1}} \rho(j) & \text{if A2-LRD holds with } 0 < \gamma_1 < 1.
\end{cases}
\]

For the cases SRD and LRD with $\gamma_1 = 1$ the conditions D and E of Altman (1990) are still satisfied with $\rho_n(j)$. Following the same arguments as in the proof of Theorem 1 of Altman (1990) the result follows. Finally, in the last case, $\rho_n(j)$ satisfies condition D but not condition E of Altman (1990). So, we have

\[
\sum_{j=1}^n j \rho_n(j) = O(n).
\]

Therefore, using Lemma A.4 in Altman (1990), it follows that

\[
\text{Var}[\hat{s}] = \sigma_\varepsilon^2 N_K \frac{S_{\rho}^*}{nh} + 2\sigma_\varepsilon^2 N_K \frac{S_{\rho}^*}{\Lambda_n h} + o\left(\frac{1}{\Lambda_n h}\right).
\]

The latter completes the proof. \(\square\)
The AMISE($\hat{s}; h$) is the asymptotic MISE, the main part of the MISE. Note that Lemma 1 gives a similar formula to (2.8) in Theorem 2.1 of Hall et al. (1995). However, differently from Hall et al. (1995) our approach does not need to introduce an additional parameter to capture SRD and LRD. Also notice that taking $h \in H$ as in A3, implies that $\text{MISE}(\hat{s}; h) = O \left( \Lambda_n^{-4/5} \right)$, which means that the kernel estimator achieves the global optimal rate.

Proof of Theorem 1 Lemma 1 holds under A1, A2 and A3. Let $\hat{\gamma}(j) = \frac{1}{n} \sum_{i=1}^{n} \left( s(i/n) - \hat{s}(i/n) \right) \left( s((i+j)/n) - \hat{s}((i+j)/n) \right)$ be the estimator of the autocovariance $\gamma(j)$ with $j = 0, 1, \ldots$. By A3 $r_n = \frac{1}{\Lambda_n h} + h^4 = \Lambda_n^{-4/5}$, and by Markov inequality

$$P \left( \frac{1}{n} \sum_{i=1}^{n} (s(i/n) - \hat{s}(i/n))^2 > \eta \right) \leq \frac{1}{\eta} \frac{1}{n} \sum_{i=1}^{n} E \left( (\hat{s}(i/n) - s(i/n))^2 \right) = O \left( \text{MISE}(\hat{s}; h) \right),$$

(18)

for some $\eta > 0$ and when $n \to \infty$.

It means that $\frac{1}{n} \sum_{i=1}^{n} (s(i/n) - \hat{s}(i/n))^2 = \text{AMISE}(\hat{s}; h) + o_p(r_n)$. Rewrite $\hat{\gamma}(j)$ as

$$\hat{\gamma}(j) = -\frac{1}{n} \sum_{i=1}^{n} \left( s(i/n) - \hat{s}(i/n) \right) \left( s((i+j)/n) - \hat{s}((i+j)/n) \right) + \frac{1}{n} \sum_{i=1}^{n} \left( s((i+j)/n) - \hat{s}((i+j)/n) \right) \varepsilon_i + \frac{1}{n} \sum_{i=1}^{n} \left( s(i/n) - \hat{s}(i/n) \right) \varepsilon_{i+j} + \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \varepsilon_{i+j} = \text{I+II+III+IV}.$$

(19)

By (18) and Cauchy–Schwartz inequality it results that term I= $O_p(r_n)$ in $\hat{\gamma}(j)$. Consider term III in (19). Without loss of generality assume $s(t) \neq 0$. By Chebyshev inequality

$$P \left( |\hat{s}(t) - s(t)| > \eta \right) \leq \frac{\text{MSE}(\hat{s}; h)}{\eta^2},$$

for some $\eta > 0$. By using the same arguments as in the proof of Lemma 1, it follows that $\text{MSE}(\hat{s}; h) = O \left( r_n \right)$ so that $\hat{s}(t) = s(t)(1 + O_p(r_n^{1/2}))$. Therefore, it is sufficient to investigate the behaviour of

$$\frac{1}{n} \sum_{i=1}^{n} s(i/n) \varepsilon_{i+j}.$$
\[ \sum_{j} \rho(j) = O(\log n) \text{ under LRD with } \gamma_1 = 1, \text{ and } \sum_{j} \rho(j) = O(n^{1-\gamma_1}) \text{ under LRD with } 0 < \gamma_1 < 1. \]

By A1, and applying Chebyshev inequality, it happens that

\[ \text{III} = O_p(\Lambda_n^{-1/2}). \]

Based on similar arguments one has that term II = \( O_p(\Lambda_n^{-1/2}) \). Now consider last term of (19), and notice that it is the series of products of autocovariances.

Theorem 3 in Hosking (1996) is used to conclude that the series is convergent under SRD and LRD with \( 1/2 < \gamma_1 \leq 1 \), while it is divergent under LRD with 0 < \( \gamma_1 \leq 1/2 \). Based on this, direct application of Chebyshev inequality to term IV implies that

\[ \text{IV} = o_p(\Lambda_n^{-1/2}). \]

Now consider last term of (19), and notice that it is the series of products of autocovariances.

Theorem 3 in Hosking (1996) is used to conclude that the series is convergent under SRD and LRD with \( 1/2 < \gamma_1 \leq 1 \), while it is divergent under LRD with 0 < \( \gamma_1 \leq 1/2 \). Based on this, direct application of Chebyshev inequality to term IV implies that

\[ \text{IV} = o_p(\Lambda_n^{-1/2}). \]

Then

\[ \hat{\gamma}(j) = \gamma(j) + o_p(r_n) + o_p(\Lambda_n^{-1/2}) + o_p(j/n), \]

where the \( o_p(r_n) \) is due to the bias of \( \hat{\gamma}(j) \). This means that \( \hat{\gamma}(j) = \gamma(j) + o_p(r_n) + o_p(\Lambda_n^{-1/2}) + o_p(j/n) \). Since \( \mathscr{K}(\cdot) \) is bounded then one can write

\[ \frac{1}{nh} \sum_{j=-M}^{M} \mathscr{K} \left( \frac{j}{nh} \right) \rho(j) = \frac{1}{nh} \sum_{j=-M}^{M} \mathscr{K} \left( \frac{j}{nh} \right) \rho(j) + o_p(r_n). \]

Consider

\[ Q_1 = \left| \frac{1}{nh} \sum_{j=-[nh/2]}^{[nh/2]} \mathscr{K} \left( \frac{j}{nh} \right) \rho(j) - \frac{1}{nh} \sum_{j=-M}^{M} \mathscr{K} \left( \frac{j}{nh} \right) \hat{\rho}(j) \right|, \]

and by (20) it follows that

\[ Q_1 = \left| 2 \frac{1}{nh} \sum_{j=M+1}^{[nh/2]} \mathscr{K} \left( \frac{j}{nh} \right) \rho(j) \right| + o_p(r_n). \]

By A2, A3 and A4,

\[ \frac{1}{nh} \sum_{j=M+1}^{[nh/2]} \mathscr{K} \left( \frac{j}{nh} \right) \rho(j) \sim \rho(nh - M) = o(r_n), \]

which implies that \( Q_1 = o_p(r_n) \). It means that the CV function, as defined in (22) of Altman (1990) with the estimated correlation function, has an error rate of \( o_p(r_n) \) with respect to

\[ \mathscr{K} \text{ Springer} \]
A Monte Carlo subsampling method for estimating... 507

Now, we can apply the classical bias correction and based on (14) in Altman (1990), we have that

\[ CV(h) = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 + \text{MISE}(\hat{\sigma}; h) + o_p(r_n) = \sigma^2 + \text{AMISE}(\hat{\sigma}; h) + o_p(r_n) \]

Since AMISE(\hat{\sigma}; h) = O(r_n), it follows that \( \hat{h} \), the minimizer of CV(h), is equal to \( h^* \), the minimizer of MISE(\hat{\sigma}; h), asymptotically in probability. By Lemma 1, it follows that \( h^* \) is the same minimizer with respect to AMISE(\hat{\sigma}; h) asymptotically.

The subsequent Lemmas are needed to show Theorem 2 and Corollary 1.

**Lemma 2** Assume A2. Suppose that \( \{a_i\} \), in A2, is Normally distributed when \( 0 < \gamma_1 \leq 1/2 \). Then \( n \to \infty, b \to \infty \) and \( b/n \to 0 \) implies \( \sup_x |G_n,b(x) - G(x)| \overset{p}{\to} 0 \), and \( q_n,\gamma_2 \overset{p}{\to} q(\gamma_2) \) for all \( \gamma_2 \in (0,1) \).

**Proof** Under A2-SRD, Theorems 4.1 and 5.1 of Politis et al. (2001) hold and the results follow. The rest of the proof deals with the LRD case. Since \( G(x) \) is continuous (Hosking 1996), we follow proof of Theorem 4 of Jach et al. (2012). Fix \( G_{n,b}^0(x) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I} \{ \tau_b (V_{n,b,i} - \sigma^2) \leq x \} \) with \( N = n - b + 1 \). It is sufficient to show that \( \text{Var}[G_{n,b}^0(x)] \to 0 \) as \( n \to \infty \). Apply Theorem 2 in Hosking (1996) to conclude that \( \tau_n (V_n - \sigma^2) \) has the same distribution as \( \tau_n (V_n^1) \), where \( V_n^1 = \frac{1}{n} \sum_{i=1}^{n} (\varepsilon_i^2 - \sigma^2) \). Therefore, we have to show that \( \text{Var}[G_{n,b}^1(x)] \to 0 \) as \( n \to \infty \), where

\[ G_{n,b}^1(x) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I} \{ \tau_b V_{n,b,i}^1 \leq x \} \quad \text{with} \quad V_{n,b,i}^1 = \frac{1}{b} \sum_{j=1}^{b} (\varepsilon_{j+i-1}^2 - \sigma^2). \]

Using the stationarity of \( \{\varepsilon_i\}_{i \in \mathbb{Z}} \), it follows that \( \text{Var}[G_{n,b}^1(x)] = \mathbb{E}[(G_{n,b}^1(x) - G_b^1(x))^2] \), where \( G_b^1(x) = \mathbb{P} (\tau_b V_b^1 \leq x) \). By Hall et al. (1998) the Hermite rank of the square function is 2. Then, based on the same arguments as in the proof of Theorem 2.2 of Hall et al. (1998) with \( q = 2 \), we can write

\[ \text{Var}[G_{n,b}^1(x)] \leq \frac{2b+1}{N} G_b^1(x) + \frac{2}{N} \sum_{i=b+1}^{N-1} \mathbb{P} \{ \tau_b V_{n,b,i}^1 \leq x, \tau_b V_{n,b,i+1}^1 \leq x \} - \left[ G_b^1(x) \right]^2. \]

\[ (21) \]

\( \square \) Springer
Consider

$$\text{Cov}[\tau_b V_{n,b,1}^1, \tau_b V_{n,b,N}^1] = \frac{\tau_b^2}{b^2} E \left[ \sum_{i=1}^{b} (\varepsilon_i^2 - \sigma_\varepsilon^2) \cdot \sum_{i=1}^{b} (\varepsilon_{i+N-1}^2 - \sigma_\varepsilon^2) \right].$$

After some algebra, we obtain

$$\text{Cov}[\tau_b V_{n,b,1}^1, \tau_b V_{n,b,N}^1] = \frac{\tau_b^2}{b} \sum_{k=-(b-1)}^{b-1} \left( 1 - \frac{|k|}{b} \right) \phi_2(k + N), \quad (22)$$

where for \( k = 1, 2, \ldots \), \( \phi_2(k) \) are the autocovariances of \( \{\varepsilon_t^2\}_{t \in \mathbb{Z}} \). For \( k \to \infty \), A2-LRD with \( 0 < \gamma_1 \leq 1 \) implies that \( \phi_2(k) = O(k^{-2\gamma_1}) \) by Theorem 3 of Hosking (1996). Take (22) and note that

$$\left| \text{Cov}[\tau_b V_{n,b,1}^1, \tau_b V_{n,b,N}^1] \right| \leq \frac{\tau_b^2}{b} \sum_{k=-(b-1)}^{b-1} |\phi_2(k + N)| \leq \frac{\tau_b^2}{b} \left( \frac{1}{N-b+1} \right)^{2\gamma_1} C_b,$$

where

$$C_b := \begin{cases} O(1) & 1/2 < \gamma_1 \leq 1, \\ O(\log b) & \gamma_1 = 1/2, \\ O(b^{1-2\gamma_1}) & 0 < \gamma_1 < 1/2. \end{cases}$$

The latter implies that for \( n \to \infty \), (22) converges to zero. Therefore, \( \tau_b V_{n,b,1}^1 \) and \( \tau_b V_{n,b,N}^1 \) are asymptotically independent. The latter can be argued based on asymptotic normality when \( 1/2 \leq \gamma_1 \leq 1 \). For the case \( 0 < \gamma_1 < 1/2 \) the asymptotic independence can be obtained by using Theorem 2.3 of Hall et al. (1998). Thus, right hand side of (21) converges to zero as \( n \to \infty \) by Cesaro Theorem. The latter shows that \( \sup_x |G_{n,b}(x) - G(x)| \overset{p}{\to} 0 \).

Following the same arguments as in Theorem 5.1 of Politis et al. (2001), and by using the first part of this proof one shows that \( q_{n,b}(\gamma_2) \overset{p}{\to} q(\gamma_2) \). The latter completes the proof.

**Lemma 3** Assume \( A1, A2, A3 \) and \( A4 \). Suppose that \( \{a_t\} \), in \( A2 \), is Normally distributed when \( 0 < \gamma_1 \leq 1/2 \). Let \( \hat{s}(t) \) be the estimate of \( s(t) \) computed on the entire sample (of length \( n \)). Then \( n \to \infty \) and \( b = o(n^{4/5}) \) implies \( \sup_x |\hat{G}_{n,b}(x) - G(x)| \overset{p}{\to} 0 \).

**Proof** Denote \( r_n = \frac{1}{\Lambda_n} + h^4 \). By Lemma 1 and \( A3 \), \( r_n = \Lambda_n^{-\frac{1}{2}} \). \( \hat{s}(t) \) is computed on the whole time series. By Lemma 2, we can use the same approach as in Lemma 1, part (i) of Coretto and Giordano (2017). We have only to verify that \( \tau_b r_n \to 0 \) as \( n \to \infty \) which is always true if \( b = o(n^{4/5}) \). □
Lemma 4 Assume A1, A2, A3 and A4. Suppose that \(\{a_t\}\), in A2, is Normally distributed when \(0 < \gamma_1 \leq 1/2\). Let \(\hat{s}(t)\) be the estimate of \(s(t)\) computed on the entire sample (of length \(n\)). Then \(n \to \infty\) and \(b = o(n^{4/5})\) implies \(\hat{q}_{n,b}(\gamma_2) \overset{P}{\longrightarrow} q(\gamma_2)\) for any \(\gamma_2 \in (0, 1)\).

Proof Using the same arguments as in Lemma 3 we have that \(\hat{G}_{n,b}(x) - G_{n,b}(x) = o_p(1)\) for each point \(x\). By the continuity of \(G(x)\) at all \(x\) we have that \(q_n,b(\gamma_2) \overset{P}{\longrightarrow} q(\gamma_2)\) by Lemma 2. Therefore \(\hat{q}_{n,b}(\gamma_2) \overset{P}{\longrightarrow} q(\gamma_2)\). Note that assumption \(b = o(n^{4/5})\) is needed to deal with A2-LRD, however for A2-SRD only we would only need \(b = o(n)\). \(\square\)

Proof of Theorem 2 Let \(P^*(X)\) and \(E^*(X)\) be the conditional probability and the conditional expectation of a random variable \(X\) with respect to a set \(\chi = \{Y_1, \ldots, Y_n\}\). Let \(\hat{G}_{n,b_1}(x)\) be the same as \(\hat{G}_{n,b}(x)\), but now \(\hat{s}(t)\) is estimated on each subsample of length \(b\), and the variance of the error term is computed on the same subsample of length \(b_1 < b\). Without loss of generality, we consider the first observation with \(t = 1\) as in Algorithm 1. Then,

\[
\frac{1}{b} \sum_{i=1}^{b} (\hat{\epsilon}_i - \epsilon_i)^2 \overset{P}{\longrightarrow} \text{MISE}(\hat{s}; h) = O_p\left(\Lambda_b^{-4/5}\right),
\]

using Lemma 1 as in the proof of Lemma 3. Let \(b_1 = o(b^{4/5})\).

Let \(Z_i(x) = \mathbb{I}\left\{\tau_{b_1} \left(\hat{V}_{n,b_1,i} - V_n\right) \leq x\right\}\) and \(Z_i^*(x) = \mathbb{I}\left\{\tau_{b_1} \left(\hat{V}_{n,b_1,i} - V_n\right) \leq x\right\}\). \(I_i\) is a Uniform random variable on \(I = \{1, 2, \ldots, n-b+1\}\). \(P(Z_i^*(x) = \hat{G}_{n,b_1}(x) = \frac{1}{K} \sum_{i=1}^{K} Z_i^*(x)\), it follows that

\[
E^*(\hat{G}_{n,b_1}(x)) = \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} Z_i(x) = \hat{G}_{n,b_1}(x) \overset{P}{\longrightarrow} G(x),
\]

as \(n \to \infty\), the latter is implied by by Lemma 3, and the fact that \(\tau_{b_1} \Lambda_b^{-4/5} \to 0\) when \(0 < \gamma_1 \leq 1\) in assumption A2.

Since \(\{I_i\}\) is the set of uniform random variables sampled without replacement, we can apply Corollary 4.1 of Romano (1989).

Therefore it follows that \(\hat{G}_{n,b_1}(x) - \hat{G}_{n,b_1}(x) \overset{P}{\longrightarrow} 0\) as \(K \to \infty\) and \(n \to \infty\). Applying the delta method approach

\[
\hat{G}_{n,b_1}(x) - G(x) = \left(\hat{G}_{n,b_1}(x) - \hat{G}_{n,b_1}(x)\right) + \left(\hat{G}_{n,b_1}(x) - G(x)\right) \overset{P}{\longrightarrow} 0,
\]

as \(K \to \infty, n \to \infty\) and \(\forall x\). Since \(G(x)\) is continuous, the convergence is uniform because of the argument of the last part of the proof of Theorem 2.2.1 in Politis et al. (1999). This concludes the proof. \(\square\)

Proof of Corollary 1 The results follow from the proof of Lemma 4 by replacing Lemma 3 with Theorem 2. \(\square\)
**Proof of Corollary 2** We can write $\tilde{G}_{n,b_1}^0(x)$ as

$$\tilde{G}_{n,b_1}^0(x) = \frac{1}{K} \sum_{i=1}^{K} \mathbb{I} \left\{ \tau_{b_1} \left( \hat{V}_{n,b_1,i} - \bar{V}_n + \left( V_n - \hat{\bar{V}}_n \right) \right) \leq x \right\}.$$ 

So, it is sufficient to use Theorem 2 to show that $\sup_x |\tilde{G}_{n,b_1}^0(x) - G(x)| \xrightarrow{p} 0$ and we only need to show that $\tau_{b_1} \left( V_n - \hat{\bar{V}}_n \right) \xrightarrow{p} 0$.

Following the same arguments as in the proof of Theorem 1, we have that $V_n - \sigma^2 = O_p \left( \tau^{-1}_n \right)$ and

$$\hat{V}_n - \sigma^2 = \begin{cases} O_p \left( \tau^{-1}_n \right) & \text{SRD and LRD with } 5/8 \leq \gamma_1 \leq 1 \\ O_p \left( \Lambda_n^{-4/5} \right) & \text{LRD with } 0 < \gamma_1 < 5/8. \end{cases}$$

So, $\tau_{b_1} \left( V_n - \hat{\bar{V}}_n \right) = \tau_{b_1} \left( V_n - \sigma^2 \right) - \tau_{b_1} \left( \hat{V}_n - \sigma^2 \right)$.

Since $b_1 = o(b^{4/5})$, we have that $\tau_{b_1} \left( V_n - \sigma^2 \right) = O_p \left( \tau_{b_1} \tau_n^{-1} \right) = o_p(1)$ and

$$\tau_{b_1} \left( \hat{V}_n - \sigma^2 \right) = \begin{cases} O_p \left( \tau_{b_1} \tau_n^{-1} \right) \quad \text{SRD and LRD with } 5/8 \leq \gamma_1 \leq 1 \\ O_p \left( \tau_{b_1} \Lambda_n^{-4/5} \right) \quad \text{LRD with } 0 < \gamma_1 < 5/8. \end{cases}$$

In both cases, it follows that $O_p \left( \tau_{b_1} \tau_n^{-1} \right) = o_p(1)$ and $O_p \left( \tau_{b_1} \Lambda_n^{-4/5} \right) = o_p(1)$, respectively. Finally, we can conclude that the result follows.

**Proof of Theorem 3** By (4) we have that

$$\hat{SNR} = 10 \log_{10} \left( \frac{\frac{1}{n} \sum_{i=1}^{n} \hat{s}^2(t_i)}{\frac{1}{m} \sum_{i=1}^{m} \left( \hat{\varepsilon}_i - \hat{\bar{\varepsilon}} \right)^2} \right)$$

and $SNR = 10 \log_{10} \left( \sigma^{-2}_\varepsilon \int s^2(t)dt \right)$. First, we analyze the quantity $\tau_m(\hat{SNR} - SNR)$. So we can write

$$\tau_m(\hat{SNR} - SNR) = 10\tau_m \left[ \log_{10} \left( \frac{\frac{1}{n} \sum_{i=1}^{n} \hat{s}^2(t_i)}{\int s^2(t)dt} \right) - \log_{10} \left( \frac{\hat{V}_m}{\sigma^2_\varepsilon} \right) \right]$$

$$= -10\tau_m \left[ \log_{10} \left( 1 + \frac{\hat{V}_m}{\sigma^2_\varepsilon} \right) - \log_{10} \left( 1 + \frac{\frac{1}{n} \sum_{i=1}^{n} \hat{s}^2(t_i)}{\int s^2(t)dt} \right) \right]$$

$$= -10\tau_m (I - II).$$
Using the same arguments as in the proof of Theorem 1, it follows that \( \hat{V}_m - \sigma_\epsilon^2 = O_p(\tau_m^{-1}) \). Expanding \( \log_{10}(1 + x) \) in Taylor’s series, we have that

\[
\tau_m I = \frac{\tau_m}{\sigma_\epsilon^2} (\hat{V}_m - \sigma_\epsilon^2) + o_p(1)
\]

and

\[
\tau_m II = \frac{\tau_m}{\int s^2(t)dt} \left( \frac{1}{n} \sum_{i=1}^{n} \hat{s}^2(t_i) - \int s^2(t)dt + o_p(\Lambda_n^{-2/5}) \right) = o_p(1). \tag{23}
\]

Now, we show the last result. From the proof of Theorem 1 and by assumption A3, we have that

\[
\hat{s}(t) = s(t) \left( 1 + O_p(\Lambda_n^{-2/5}) \right). \tag{24}
\]

Now, we can write

\[
\frac{1}{n} \sum_{i=1}^{n} \hat{s}^2(t_i) - \int s^2(t)dt = \left( \frac{1}{n} \sum_{i=1}^{n} \hat{s}^2(t_i) - \frac{1}{n} \sum_{i=1}^{n} s^2(t_i) \right) + \left( \frac{1}{n} \sum_{i=1}^{n} s^2(t_i) - \int s^2(t)dt \right) = I_s + II_s.
\]

By using the convergence of the quadrature of a bounded and continuous function to its integral, it follows that \( II_s = O(n^{-1}) \). By (24), we have that

\[
I_s = \frac{1}{n} \sum_{i=1}^{n} \left( \hat{s}^2(t_i) - s^2(t_i) \right) = O_p(\Lambda_n^{-2/5}).
\]

Since \( m = o(n^{2/5}) \), it follows that \( \tau_m \Lambda_n^{-2/5} \to 0 \) as \( n \to \infty \). So, (23) is shown.

Hence, we can conclude that \( \tau_m (SNR - SNR) \) has the same asymptotic distribution as \( \frac{\tau_m}{\sigma_\epsilon^2} (\hat{V}_m - \sigma_\epsilon^2) \) by the Slutsky’s Theorem. Therefore, assumption 3.2.1 of Politis et al. (1999) is verified by Theorem 2.

Consider the SNR evaluated at a given point, namely \( SNR_i = 10 \log_{10} \left( \frac{s^2(t_i)}{\sigma_\epsilon^2} \right) \), and write \( \tau_{b_1} (\hat{SNR}_{n,b,I} - SNR) \) in \( \mathcal{Q}_n(\chi) \) as

\[
\tau_{b_1} (\hat{SNR}_{n,b,I} - SNR) = \tau_{b_1} (\hat{SNR}_{n,b,I} - SNR_i) - \tau_{b_1} (SNR_i - SNR) + \tau_{b_1} (SNR_i - SNR) = S_1 - S_2 + S_3,
\]

for a given subsample starting at \( I_i \). By using the first part of this proof, it follows that \( S_2 = O_p(\tau_{b_1}/\tau_m) = o_p(1) \) since \( b_1/m \to 0 \) when \( n \to \infty \). Now, in order to deal
with the quantity $S_1$, we need to show that

$$\frac{1}{b} \sum_{j=i}^{i+b-1} \left[ s \left( \frac{j - i + 1}{b} \right) \right]^2 \to s^2(t_i) \quad \text{as} \quad n \to \infty, \quad (25)$$

where $t_i$ is the initial point in the block of $b$ values. By using again the convergence of the quadrature of a bounded and continuous function to its integral, we have that

$$\frac{1}{b} \sum_{j=i}^{i+b-1} \left[ s \left( \frac{j - i + 1}{b} \right) \right]^2 \to \int_0^1 (s^b(t))^2 \, dt \quad \text{as} \quad n \to \infty, \quad b \to \infty \quad \text{and} \quad b/n \to 0.\quad (25)$$

The quantity $s^b_i(\cdot)$ denotes the portion of the signal in the block of $b$ values in $(0, 1)$ with $i$ the index for the initial point. Note that $b/n \to 0$, and by the mean value theorem $\int_0^1 (s^b_i(t))^2 \, dt \to s^2(t_i)$. By using, again, the first part of this proof and by (25), we have that

$$\tau_{b_1} \left( \frac{S \bar{N} R_{n,b,i} - SNR_i}{\sigma^2} \right) \text{ has the same asymptotic distribution as } \frac{\tau_{b_1}}{\sigma^2} \left( \tilde{V}_{n,b_1,i} - \sigma^2 \right).$$

Now we study the quantity $S_3$. First, we show that

$$\frac{1}{n - b + 1} \sum_{i=1}^{n-b+1} I \{ (SNR_i - SNR) > x \} \to 0 \quad (26)$$

when $n \to \infty$ with some $x > 0$. Since $SNR_i - SNR = 10 \log_{10} \left( \frac{s^2(t_i)}{\int s^2(t) \, dt} \right)$, the equation in (26) becomes

$$\frac{1}{n - b + 1} \sum_{i=1}^{n-b+1} I \left\{ \frac{s^2(t_i)}{\int s^2(t) \, dt} > 10^{\frac{x}{\tau_{b_1}}} \right\}. \quad (26)$$

We have that

$$\frac{1}{n - b + 1} \sum_{i=1}^{n-b+1} s^2(t_i) = \int s^2(t) \, dt + O(n^{-1}). \quad (27)$$

Moreover, $\frac{s^2(t_i)}{\int s^2(t) \, dt} > 10^{\frac{x}{\tau_{b_1}}} \text{ can be written as}$

$$\tau_{b_1} \left( \frac{s^2(t_i)}{\int s^2(t) \, dt} - 1 \right) > \tau_{b_1} \left( 10^{\frac{x}{\tau_{b_1}}} - 1 \right).$$

Summing over the index $i$ and dividing by $n - b + 1$, we can write

$$\tau_{b_1} \left( \frac{\frac{1}{n-b+1} \sum_{i=1}^{n-b+1} s^2(t_i)}{\int s^2(t) \, dt} - 1 \right) > \tau_{b_1} \left( 10^{\frac{x}{\tau_{b_1}}} - 1 \right). \quad (26)$$
Since \( \tau_{b_1} \left( 10^{ \frac{x}{\tau_{b_1}} } - 1 \right) \rightarrow c > 0 \) when \( b_1 \rightarrow \infty \), by using equation (27) we obtain
\[
\tau_{b_1} \left( \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} \frac{s^2(t_i)}{s^2(t)} dt - 1 \right) = O(\tau_{b_1} n^{-1}) \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.
\]

Therefore \( \frac{N^b_n}{n-b+1} \rightarrow 0 \) as \( n \rightarrow \infty \), where \( N^b_n = \sum_{i=1}^{n-b+1} I \{ \frac{s^2(t_i)}{s^2(t)} dt > 10^{ \frac{x}{\tau_{b_1}} } \} \).

Then, (26) is shown.

As in the proof of Slutsky’s Theorem, we split \( Q_n(x) \) as the sum of three empirical distribution function computed over \( S_1, S_2 \) and \( S_3 \) respectively. Here the random variables \( I_i \) are treated as in the proof of Theorem 2. Based on the argument above only the component of \( Q_n(x) \) computed over \( S_1 \) has a non degenerate limit distribution, and this will the same as the asymptotic distribution of the estimator for the variance of the error term. The proof is now completed. □

References

Altman NS (1990) Kernel smoothing of data with correlated errors. J Am Stat Assoc 85(411):749–759
Brillinger DR, irizarry RA (1998) An investigation of the second-and higher-order spectra of music. Sig Process 65(2):161–179
Conte E, Maio AD (2002) Adaptive radar detection of distributed targets in non-gaussian noise. In: RADAR 2002. IEE
Coretto P, Giordano F (2017) Nonparametric estimation of the dynamic range of music signals. Aust N Z J Stat 59(4):389–412
Czanner G, Sarma SV, Ba D, Eden UT, Wu W, Eskandar E, Lim HH, Temereanca S, Suzuki WA, Brown EN (2015) Measuring the signal-to-noise ratio of a neuron. Proc Natl Acad Sci 112(23):7141–7146
Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 101(23):e215–e220
Götz F, Račkauskas A (2001) Adaptive choice of bootstrap sample sizes. In: State of the art in probability and statistics. Institute of Mathematical Statistics, pp 286–309
Gray R (1990) Quantization noise spectra. IEEE Trans Inf Theory 36(6):1220–1244
Hall P, Jing BY, Lahiri SN (1998) On the sampling window method for long-range dependent data. Stat Sin 8:1189–1204
Hall P, Lahiri SN, Polzehl J (1995) On bandwidth choice in nonparametric regression with both short- and long-range dependent errors. Ann Stat 23(6):1921–1936
Haykin SS, Kosko B (2001) Intelligent signal processing. Wiley-IEEE Press
Hosking JRM (1996) Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. J Econom 73:261–284
Jach A, McElroy T, Politis DN (2012) Subsampling inference for the mean of heavy-tailed long-memory time series. J Time Ser Anal 33:96–111
Kalogera V (2017) Too good to be true? Nat Astron 1(0112):1–4
Kay SM (1993) Fundamentals of statistical signal processing, volume 1. Estimation theory. Prentice Hall, Englewood Cliffs
Kemp B, Zwiederman A, Tuk B, Kamphuisen H, Oberye J (2000) Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. IEEE Trans Biomed Eng 47(9):1185–1194
Kenig E, Cross MC (2014) Eliminating 1/f noise in oscillators. Phys Rev E 89(4):0429011–0429017
Kogan S (1996) Electronic noise and fluctuations in solids. Cambridge University Press, Cambridge
Levitin DJ, Chordia P, Menon V (2012) Musical rhythm spectra from bach to joplin obey a 1/f power law. Proc Natl Acad Sci 109(10):3716–3720
Ligges U, Krey S, Mersmann O, Schnackenberg S (2016) tuneR: analysis of music. CRAN
Loizou PC (2013) Speech enhancement: theory and practice, 2nd edn. CRC Press, Boca Raton
Parzen E (1966) Time series analysis for models of signal plus white noise. Technical report, Department of Statistics, Stanford University
Parzen E (1999) Stochastic processes (classics in applied mathematics). Society for Industrial and Applied Mathematics
Politis DN, Romano JP (1994) Large sample confidence regions based on subsamples under minimal assumptions. Ann Stat 22(4):2031–2050
Politis DN, Romano JP, Wolf M (1999) Subsampling. Springer, New York
Politis DN, Romano JP, Wolf M (2001) On the asymptotic theory of subsampling. Stat Sin 11(4):1105–1124
Priestley MB, Chao MT (1972) Nonparametric function fitting. J R Stat Soc 34:385–392
Richards MA (2014) Fundamentals of Radar signal processing, second edition (McGraw-Hill professional engineering). McGraw-Hill Education, New York
Romano JP (1989) Bootstrap and randomization tests of some nonparametric hypotheses. Ann Stat 17(1):141–159
Shoeb AH (2009) Application of machine learning to epileptic seizure onset detection and treatment. Ph.D. thesis, Massachusetts Institute of Technology
Timmer J, König M (1995) On generating power law noise. Astron Astrophys 300:707
Ullsperger M, Debener S (2010) Simultaneous EEG and fMRI: recording, analysis, and application. Oxford University Press, Oxford
Voss RF, Clarke J (1975) “1/f noise” in music and speech. Nature 258:317–318
Voss RF, Clarke J (1978) “1/f noise” in music: music from 1/f noise. J Acoust Soc Am 63:258
Weihs C, Jannach D, Vatolkin I, Rudolph G (2016) Music data analysis: foundations and applications. Chapman and Hall/CRC, New York
Weinberg G (2017) Radar detection theory of sliding window processes. CRC Press, Boca Raton
Weissman MB (1988) 1fnoise and other slow, nonexponential kinetics in condensed matter. Rev Mod Phys 60(2):537–571

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.