University of Cyprus

DEPARTMENT OF MATHEMATICS AND STATISTICS

TOPICS IN STATISTICAL INFERENCE FOR LINEAR AND NONLINEAR TIME SERIES

DOCTOR OF PHILOSOPHY DISSERTATION

MARIA FRAGKESKOU

2017

University of Cyprus

DEPARTMENT OF MATHEMATICS AND STATISTICS

TOPICS IN STATISTICAL INFERENCE FOR LINEAR AND NONLINEAR TIME SERIES

MARIA FRAGKESKOU

A Dissertation Submitted to the University of Cyprus in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

DECEMBER 2017

DECLARATION OF DOCTORAL CANDIDATE

The present doctoral dissertation was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy of the University of Cyprus. It is a product of original work of my own, unless otherwise mentioned through references, notes, or any other statements.

Maria Fragkeskou

To my parents,

Andreas and Stavry,

for their unconditional support and endless love.

Abstract

The aim of this thesis is twofold. First, to investigate the problem of estimating the rescaled fourth order cumulant of the unobserved innovations of linear time series which is an important parameter for statistical inference. Second, to propose two modifications of the autoregressive-sieve respectively of the autoregressive bootstrap.

For the first problem, an existing nonparametric estimator is first discussed and its asymptotic properties are derived. In particular, it is shown how the autocorrelation structure of the underlying process affects the behavior of the estimator. Based on these findings and on a discovered and important invariance property of the parameter of interest with respect to linear filtering, a pre-whitening based nonparametric estimator of the same parameter is proposed. The aforementioned invariance property implies that the parameter of interest can be estimated using the residuals obtained by applying the linear filter to the time series at hand and an inverse-transformation is not needed. It is shown that if the filter chosen to pre-whiten the time series is such that the filtered time series is less correlated than the original one, then the new estimator has several advantages.

The asymptotic properties of the new estimator based on a simple autoregressive filter are investigated and its superiority is theoretically established for large classes of stochastic processes. It is shown that for the particular estimation problem considered, pre-whitening not only reduces the variance of the estimator but it can also lead to gains in terms of bias. The finite sample performance of the existing and of the new estimator is investigated and compared by means of several simulations. As an application, we show that the new estimator allows for a simple modification of the multiplicative frequency domain bootstrap which considerable extends its range of validity. Furthermore, the problem of testing hypotheses about the rescaled fourth order cumulant of the unobserved innovations is also considered. In this context, a simple test for Gaussianity is proposed and some real-life data applications are presented.

Concerning the two modifications of the autoregressive-sieve respectively of the autoregressive bootstrap proposed, the first replaces the classical i.i.d. resampling scheme applied to the residuals of the autoregressive fit by a generation of i.i.d. wild pseudo-innovations that appropriately mimic the first and the second order moments as well as the rescaled fourth order cumulant of the true innovations driving the underlying linear process. This modification, uses the estimator of the fourth order cumulant presented in the first part of the thesis and extends the range of validity of the autoregressive-sieve bootstrap to classes of statistics for which the classical, residual-based autoregressive-sieve bootstrap, fails. The second modification, is a version of the autoregressive bootstrap which is applied to an appropriately transformed time series. This, together with a dependent-wild type generation of pseudo-innovations, delivers a bootstrap procedure which is valid for large classes of statistics and for stochastic processes that satisfy quite general weak dependence conditions. A fully data-driven selection of the tuning parameters involved in both bootstrap modifications is proposed, while extensive simulations, including comparisons with alternative bootstrap methods, show a good finite sample performance of the proposed bootstrap procedures.

Περίληψη

Η παρούσα διατριβή έχει δύο κυρίως στόχους. Ο πρώτος, να ασχοληθεί με το πρόβλημα της εκτίμησης του ανακλιμακώμενου συσσωρευτή τέταρτης τάξης (rescaled fourth order cumulant) του μη παρατηρηθέντος λευκού θορύβου μιας γραμμικής στοχαστικής ανέλιξης, ο οποίος αποτελεί μία σημαντική παράμετρο στην στατιστική συμπερασματολογία χρονοσειρών. Ο δεύτερος στόχος είναι να προτείνει δύο τροποποιήσεις της αυτοπαλινδρομικής διαδικασίας bootstrap, οι οποίες επεκτείνουν σημαντικά το εύρος των εφαρμογών και της ασυμπτωτικής συνέπειας των αντίστοιχων μεθόδων.

Όσον αφορά το πρώτο πρόβλημα εκτίμησης, μία υφιστάμενη μη παραμετρική εκτιμήτρια εξετάζεται και οι ασυμπωτικές ιδιότητες της ερευνούνται. Τα ασυμπτωτικά αποτελέσματα δείχνουν με ποιό τρόπο η αυτοσυσχέτιση της στοχαστικής ανέλιξης επηρεάζει την συμπεριφορά της εκτιμήτριας. Βασισμένοι σε αυτά τα ευρήματα και σε μια σημαντική ιδιότητα αναλλοίωτου της παραμέτρου που μας ενδιαφέρει ως προς τα γραμμικά φιλτραρίσματα της χρονοσειράς, προτείνεται μία καινούργια μη παραμετρική εκτιμήτρια η οποία βασίζεται σε μία λευκοθορυβοποίηση (pre-whitening) της υφιστάμενης χρονοσειράς. Η προαναφερθείσα ιδιότητα του αναλλοίωτου συνεπάγεται ότι η παράμετρος που μας ενδιαφέρει μπορεί να εκτιμηθεί χρησιμοποιώντας τα υπόλοιπα που προκύπτουν εφαρμόζοντας ένα γραμμικό φίλτρο μετασχηματισμού της χρονοσειράς και ότι δεν απαιτείται η χρήση οποιουδήποτε αντίστροφου μετασχηματισμού. Αν η φιλτραρισμένη χρονοσειρά είναι λιγότερο συσχετισμένη σε σχέση με την παρατηρηθείσα, τότε η νέα εκτιμήτρια έχει αρκετά πλεονεκτήματα συγκρινόμενη με την εκτιμήτρια που χρησιμοποιεί την αρχική χρονοσειρά.

Οι ασυμπωτικές ιδιότητες της καινούργιας εκτιμήτριας βασισμένη σε ένα απλό αυτοπαλινδρομικό φίλτρο ερευνούνται και η ανωτερότητα της νέας εκτιμήτριας για μία μεγάλη κλάση στοχαστικών διαδικασιών αποδεικνύεται . Όπως προκύπτει για το συγκεκριμένο πρόβλημα εκτίμησης η λευκοθορυβοποίηση μπορεί να μειώσει σημαντικά όχι μόνο τη διασπορά αλλά και τη μεροληψία της εκτιμήτριας. Μέσω προσομοιώσεων διερευνούμε την συμπεριφορά των δύο εκτιμητριών σε δείγματα πεπερασμένου μεγέθους. Μεταξύ

των εφαρμογών που έχει η καινούργια εκτιμήτρια είναι και μία απλή τροποποίηση της πολλαπλασιαστικής διαδικασίας bootstrap βασισμένης στη φασματική πυκνότητα η οποία και επεκτείνει σημαντικά το εύρος των εφαρμογών της. Επιπρόσθετα εξετάζουμε και το πρόβλημα ελέγχου υποθέσεων σχετικά με τον ανακλιμακώμενο συσσωρευτή τέταρτης τάξης. Στο πλαίσιο αυτό, προτείνεται ένα απλός έλεγχος κανονικότητας (Gaussianity) και παρουσιάζονται πολλές εφαρμογές σε πραγματικά δεδομένα.

Οσον αφορά τις δύο τροποποιήσεις της αυτοπαλινδρομικής διαδικασίας bootstrap, η πρώτη αντικαθιστά το κλασσικό σχήμα αναδειγματολειψίας που αφορά σε ανεξάρτητες και ισόνομες τυχαίες μεταβλητές και εφαρμόζεται στα υπόλοιπα του αυτοπαλινδρομικού μοντέλου. Η νέα διαδικασία αναδειγματολειψίας βασίζεται στη δημιουργία ανεξάρτητων και ισόνομων άτακτων (wild) ψευδο-υπολοίπων τα οποία μιμούνται κατάλληλα την ροπή πρώτης και δεύτερης τάξης και τον ανακλιμακώμενο συσσωρευτή τέταρτης τάξης (rescaled fourth order cumulant) των ανεξάρτητων και ισόνομων τυχαίων μεταβλητών (innovations) που επισέρχονται στη δημιουργία της γραμμικής στοχαστικής ανέλιξης. Αυτή η τροποποίηση η οποία χρησιμοποιεί την εκτιμήτρια του ανακλιμακώμενου συσσωρευτή τέταρτης τάξης που παρουσιάστηκε στο πρώτο μέρος της διατριβής, επεκτείνει την εγκυρότητα της αυτοπαλινδρομικής διαδικασίας bootstrap, σε κλάσεις στατιστικών συναρτήσεων για τις οποίες η κλασσική αυτοπαλινδρομική διαδικασία bootstrap αποτυγχάνει. Στη δεύτερη τροποποίηση, προτείνεται μια αυτοπαλινδρομική διαδικασία bootstrap εφαρμοσμένη όχι στην χρονοσειρά που παρατηρείται, αλλά σε ένα κατάλληλο μετασχηματισμό της. Μαζί με τη δημιουργία εξαρτημένων-άτακτων (dependent-wild) ψευδο-υπολοίπων ορίζεται μία διαδικασία bootstrap η οποία είναι συνεπής για μία μεγάλη κλάση στατιστικών συναρτήσεων και για στοχαστικές διαδικασίες οι οποίες ικανοποιούν αρκετά γενικές ασθενείς συνθήκες εξάρτησης. Τέλος, προτείνεται μία διαδικασία αυτόματης επιλογής των παραμέτρων που εμφανίζονται στις δύο διαδικασίες bootstrap που προτείνουμε, ενώ εκτεταμένες προσομοιώσεις, περιλαμβανομένων συγκρίσεων με εναλλακτικές μεθόδους bootstrap, δείχνουν την καλή συμπεριφορά των νέων διαδικασιών bootstrap σε δείγματα πεπερασμένου μεγέθους.

Acknowledgements

This dissertation would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study. It is a pleasure to convey my gratitude to all of them, in my humble acknowledgment.

First of all, I would like to express my gratitude to my supervisor, Professor Efstathios Paparoditis. His guidance and encouragement, his belief in me and his patience, from the initial point of my postgraduate studies, made the completion of my PhD studies possible. He enabled me to develop a deep understanding, not only of time series analysis but of several other related topics in the area of statistics and of mathematics, and inspired me to contribute to the field in my own way, having always in mind our long, thought-provoking and constructive discussions.

I gratefully thank Professor Konstantinos Fokianos, Professor Theofanis Sapatinas, Professor Jens-Peter Kreiss and Professor Jürgen Franke for their constructive comments and suggestions on this dissertation. I am thankful, that in the midst of all their activities, they accepted to be members of the examining committee.

Special thanks to all the academic and administrative staff of the Department of Mathematics and Statistics of the University of Cyprus for their support and guidance during my undergraduate and postgraduate studies.

At last, I would like to express my most sincere gratitude and love to my family for their unshakable faith in me. Without their encouragement and support the completion of this dissertation would be impossible. Specifically, I immensely thank my parents for all these years that they supported me emotionally and financially to succeed in constructing my personality, education and career as well as, my future husband Stelios for his support, understanding and patience.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Statistical inference for linear processes is a well developed area in time series analysis. Linear stochastic processes are generated by applying a linear filter to a sequence of unobserved innovations $\{\varepsilon_t, t \in \mathbb{Z}\}$, which are assumed to be independent and identically distributed (i.i.d.) with mean zero and finite variance $\sigma_\varepsilon^2 = E(\varepsilon_1^2)$. If we denote by $\mathbf{X} = \{X_t, t \in \mathbb{Z}\}$ the linear process and by $\{\psi_j, j \in \mathbb{Z}\}$,

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty,$$

the coefficients of the linear filter, then the generating equation of the X_t's is given by

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \qquad \psi_0 = 1. \tag{1.1}$$

Statistical inference for such processes is typically based on an observed stretch $X_1, X_2, \ldots, X_n$ stemming from $\mathbf{X}$. There are situations, however, where the outcome of an inference procedure depends also on a certain higher order moments of ε_1. An important and particularly difficult case in this context occurs, when this dependence refers to the rescaled, fourth order cumulant of the innovations, that is, to the parameter

$$\eta_{4,\varepsilon} = \frac{E(\varepsilon_1^4) - 3\sigma_\varepsilon^4}{\sigma_\varepsilon^4}.$$

To elaborate on such inference situations, denote by $I_X(\lambda)$ the periodogram of the time series $X_1, X_2, \ldots, X_n$, i.e.,

$$I_X(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} X_t e^{-it\lambda} \right|^2, \quad \lambda \in \mathbb{R},$$

and consider the following general class of statistics, also known as spectral means,

$$S_n = \int_{-\pi}^{\pi} \varphi(\lambda) I_X(\lambda) d\lambda, \tag{1.2}$$

where φ is some function $\varphi : [-\pi, \pi] \to \mathbb{R}$. Class (1.2) is large enough and includes, as special cases, several interesting examples of commonly used statistics in time series analysis. For instance, for $\varphi(\lambda) = \cos(\lambda h)$, with h some integer $0 \leq h < n$, S_n is the sample autocovariance

$$S_n = \frac{1}{n} \sum_{t=1}^{n-|h|} X_t X_{t+|h|},$$

which is an estimator of the autocovariance

$$\gamma_X(h) = E(X_t X_{t+h}).$$

For $\varphi(\lambda) = \mathbf{1}_{[0,x]}(\lambda)$, $x \in [0, 1]$, S_n is the empirical spectral distribution function

$$S_n = \widehat{F}_n(\lambda) = \int_0^x I_X(\lambda)\, d\lambda,$$

which is an estimator of the spectral distribution function

$$F_X(x) = \int_0^x f_X(\lambda) d\lambda.$$

Here,

$$f_X(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_X(h) \cos(\lambda h)$$

denotes the spectral density of $\mathbf{X}$, which by the assumption that $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, exists and is continuous. It is well known, cf. Dahlhaus (1985) that under certain regularity conditions,

$$\sqrt{n}\big(S_n - \int_{-\pi}^{\pi} \varphi(\lambda) f_X(\lambda) d\lambda\big) \xrightarrow{D} N(0, \sigma_\varphi^2),$$

as $n \to \infty$, where

$$\sigma_\varphi^2 = 2\pi \int_{-\pi}^{\pi} \varphi^2(\lambda) f_X^2(\lambda) d\lambda + \eta_{4,\varepsilon} \left(\int_{-\pi}^{\pi} \varphi(\lambda) f_X(\lambda) d\lambda \right)^2 \tag{1.3}$$

and " $\xrightarrow{D}$ " denotes weak convergence. As equation (1.3) shows, the variance σ_φ^2 of the limiting Gaussian distribution depends on the unknown rescaled fourth order cumulant $\eta_{4,\varepsilon}$ of the i.i.d. innovations ε_t. Thus, implementation of the above asymptotic result for estimating the variance of S_n or for the construction of confidence intervals for

$$\int_{-\pi}^{\pi} \varphi(\lambda) f_X(\lambda) d\lambda,$$

requires estimation of $\eta_{4,\varepsilon}$.

Another situation where interest in estimating the parameter $\eta_{4,\varepsilon}$ occurs, appears when one deals with frequency domain bootstrap methods for the periodogram. Notice first that the periodogram $I_X(\lambda)$ is commonly calculated at the so-called Fourier frequency $\lambda_j \in \mathcal{F}_n$, where

$$\mathcal{F}_n = \left\{ \lambda_j = \frac{2\pi j}{n} : j = -\left[\frac{n-1}{2}\right], \ldots, \left[\frac{n}{2}\right] \right\}.$$

It is well known that for linear processes, periodogram ordinates are asymptotically exponential distributed and obey the following (approximative) multiplicative expression

$$I_X(\lambda_j) = f_X(\lambda_j) I_e(\lambda_j) + O_P\left(\frac{1}{\sqrt{n}}\right), \tag{1.4}$$

where the $O_P\left(1/\sqrt{n}\right)$ is uniform in the frequencies λ_j and

$$I_e(\lambda_j) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} e_t \exp\{-it\lambda_j\} \right|^2,$$

denotes the periodogram of the standardized innovations $e_1, e_2, \ldots, e_n$, $e_t = \varepsilon_t/\sigma_\varepsilon$; see Brockwell and Davis (1991), Ch. 10, Theorem 10.3.1. Furthermore, for $0 < \lambda_j \neq \lambda_k < \pi$, it yields that

$$Cov(I_X(\lambda_j), I_X(\lambda_k)) = \frac{1}{n}\eta_{4,\varepsilon} f_X(\lambda_j) f_X(\lambda_k) + o_P\left(\frac{1}{n}\right);$$

see Paparoditis (2002). That is, periodogram ordinates at different frequencies are asymptotically uncorrelated, although they appear to be weakly n^{-1}-dependent in finite samples. Moreover, it can be shown that for any number of fixed frequencies $0 < \lambda_1 < \lambda_2 < ... < \lambda_m < \pi$, the corresponding periodogram ordinates are asymptotically independent; see for instance Brockwell and Davis (1991), Theorem 10.3.2. Expression (1.4) together with the aforementioned asymptotic independence of the periodogram ordinates have been used by some researchers to bootstrap the periodogram; see Franke and Härdle (1992) and Dahlhaus and Janas (1996). Such approaches work by ignoring the $O_P\left(1/\sqrt{n}\right)$ term in (1.4) and by generating independent pseudo-periodogram ordinates, $I_X^*(\lambda_j)$, which are obtained by replacing the unknown spectral density f_X in (1.4) by some estimator and by generating independent pseudo-random variables U_j^*, designed in a way to mimic the behavior of $I_e(\lambda_j)$. However, because such bootstrap approaches neglect the weak dependence structure of the periodogram ordinates $I_X(\lambda_j)$, they can not be successfully applied to statistics, the distribution of which is affected by this weak dependence. Spectral means defined in

(1.2) are examples of such statistics. As a matter of fact, the second term in the limiting variance σ_φ^2 of S_n exhibited in (1.3), is due to the weak and asymptotically vanishing dependence of the periodogram ordinates; see Dahlhaus and Janas (1996) and Paparoditis (2002) for an extensive discussion. This failure of frequency domain bootstrap methods which generate independent pseudo-periodogram ordinates, motivated many researches to develop bootstrap procedures for the periodogram that imitate correctly also its weak, n^{-1}-vanishing dependence structure; see Janas and Dahlhaus (1994) and Kreiss and Paparoditis (2012). Such bootstrap approaches need, however, a consistent estimator of the rescaled fourth order cumulant $\eta_{4,\varepsilon}$.

In the second chapter of the thesis, we consider the problem of estimating the rescaled fourth order cumulant of the i.i.d innovations for linear processes. We derive the asymptotic distribution of an existing lag-window type estimator as well as bias and variance expressions. Based on these findings, a new estimator is proposed which is based on the idea of pre-whitening. Due to an invariance property of the parameter of interest with respect to linear filtering, an inverse transformation is not required. The asymptotic distribution and the bias and variance properties of the new estimator are derived. Its theoretical superiority is established for large classes of stochastic processes. In addition, approximations of the asymptotic mean square error of the new estimator are obtained which build the basis for some practical rules to select the smoothing parameters involved in the estimation procedure.

Notice that if the underlying time series is not linear, the second expression in (1.3) is typically replaced by certain integrals of the fourth order cumulant spectral density of the underlying process; see Dahlhaus (1985). In this case, consistent estimators of the corresponding expressions based on functions of finite Fourier transforms of $X_1, X_2, \ldots, X_n$ have been considered by Taniguchi (1982). Keenan (1987) derived asymptotic results for more general class of estimators of such quantities. However, the inference problem considered in this chapter is different since we are concerned with linear processes and we focus on the case where the parameter of interest is the rescaled fourth order cumulant $\eta_{4,\varepsilon}$ of the unobserved innovations ε_t and not the fourth order cumulant density of the underlying process $\{X_t, t \in \mathbb{Z}\}$ or integrals thereof.

Bootstrap is a powerful tool for statistical inference in time series. This is mainly due to the fact that for time series and for many statistics of interest, asymptotic derivations are not only quite involved but the results obtained are also difficult to implement in practise. Developing appropriate bootstrap methods for time series is a challenging and difficult task and many approaches already exist in the literature.

A basic problem faced by bootstrap procedures for time series is that, in order to be successfully applied to some statistics of interest, they have to imitate (at least to the necessary extent) the (in principle complicated) dependence structure of the underlying stochastic process. This problem has been addressed by the different bootstrap proposals for time series in a different way, which depends on the kind of weak dependence assumptions imposed on the process generating the observed time series and on the structure of the particular statistic of interest; see Bühlmann (2002), Politis (2003) and Kreiss and Paparoditis (2011) for an overview.

Among the different bootstrap methods proposed in the literature, the autoregressive (AR) and the autoregressive-sieve (AR-sieve) bootstrap, are quite popular due to their easy implementation and their potential applicability to a variety of situations. The basic idea is to generate new pseudo-time series by using an estimated autoregressive model driven by pseudo-innovations generated by means of i.i.d. resampling of the estimated (and centered) residuals of the autoregressive fit. While in practice and for a given time series $X_1, X_2, \ldots, X_n$ at hand, the procedure is the same for the AR and for the AR-sieve bootstrap, from a theoretical point of view, the two methods are quite different. The AR bootstrap assumes that the underlying process follows a linear, finite order autoregressive model, while the AR-sieve bootstrap considers the autoregression fitted to the observed time series as an approximation of the more complicated autocovariance structure of the underlying process. In order to capture the entire autocovariance structure of the process, the AR-sieve bootstrap requires, therefore, that the order p of the autoregression fitted increases to infinity (at an appropriate rate) as the sample size increases to infinity.

An interesting question is, of course, for which stochastic processes and for what kind of statistics are autoregressive bootstrap methods valid. While this seems to be clear for the AR bootstrap, the situation is more involved for the AR-sieve bootstrap. This question becomes even more interesting when one takes into account that a so-called autoregressive representation exists for a wide class of strictly stationary stochastic processes. To elaborate, recall the well-known Wold representation according to which every purely nondeterministic, stationary and zero mean stochastic process $\mathbf{X} = \{X_t, t \in \mathbb{Z}\}$ obeys a unique, infinite order moving average representation

$$X_t = \sum_{j=0}^{\infty} c_j e_{t-j}, \tag{1.5}$$

where

$$c_0 = 1, \quad \sum_{j=0}^{\infty} c_j^2 < \infty$$

and $\{e_t, t \in \mathbb{Z}\}$ is a white noise process, i.e., the e_t's are uncorrelated, zero mean random variables with variance $0 < \sigma_e^2 < \infty$. If $\mathbf{X}$ also possesses a spectral density which is continuous and bounded away from zero everywhere in the interval $[0, \pi]$, then X_t also obeys a so-called, autoregressive representation. That is, X_t can be alternatively expressed as

$$X_t = \sum_{j=1}^{\infty} a_j X_{t-j} + e_t, \tag{1.6}$$

where the innovations e_t are identical to those appearing in Wold's representation (1.5) and the coefficients a_j satisfy the stronger condition

$$\sum_{j=1}^{\infty} |a_j| < \infty;$$

see Pourahmadi (2001) and Kreiss et al. (2011). Notice that expressions (1.5) and (1.6), do not describe a model but just two alternative representations of X_t. The first is a representation of X_t in terms of the history of the innovations $\{e_t, t \leq 1\}$ and the second in terms of the history of the process X_t itself. From a statistical point of view, representation (1.6) seems to be more appealing since it allows for an estimation of the coefficients a_j based on the observed part of the process.

Although originally proposed for infinite order linear autoregressive processes, see Kreiss (1988), Paparoditis and Streitberg (1991) and Bühlmann (1997), the question about the range of validity of the AR-sieve bootstrap has been recently discussed in Kreiss et al. (2011) in the context of general (strictly) stationary processes obeying representation (1.6). Since the AR-sieve bootstrap succeeds in mimicking correctly the second order structure of the underlying process $\mathbf{X}$, it has been shown that if the (asymptotic) distribution of a statistic of interest depends only on second order characteristics, then the AR-sieve bootstrap will be asymptotically valid even if $\mathbf{X}$ is nonlinear. Sample mean and nonparametric estimators of the spectral density are two examples of such statistics; see Kreiss et al. (2011) for details.

For nonlinear processes and for more general statistics, however, like for instance, for the important class of generalized means given by

$$T_n = f\left(\frac{1}{n-m+1}\sum_{t=1}^{n-m+1} g(X_t, X_{t+1}, \ldots, X_{t+m-1})\right), \tag{1.7}$$

$m < n$, with f and g appropriate functions, see Künsch (1989), the AR-sieve bootstrap fails. Class (1.7) includes many statistics of interest in time series analysis, as special cases. In the following, we discuss some examples.

(i) For $m = 1$ and f=g be the identity function, then T_n is the sample mean

$$T_n = \frac{1}{n} \sum_{t=1}^{n} X_t.$$

(ii) For $m < n$

$$g(x_1, x_2, ..., x_m) = (x_1 \cdot x_1, x_1 \cdot x_2, ..., x_1 \cdot x_m),$$

and f be the indicator function, then

$$T_n = \left(\frac{1}{n} \sum_{t=1}^{n-k} X_t X_{t+k}, \quad k = 0, 1, ..., m-1 \right),$$

that is, T_n is an estimator of the autocovariances $(\gamma_X(0), \gamma_X(1), ..., \gamma_X(m-1))$.

(iii) For the function g defined as in (ii) and f as

$$f(y_0, y_1, ..., y_{m-1}) = \left(\frac{y_1}{y_0}, \frac{y_2}{y_0}, ..., \frac{y_{m-1}}{y_0} \right), \quad y_0 \neq 0,$$

them T_n is an estimator of the autocorrelations $(\rho_X(1), \rho_X(1), ..., \rho_X(m-1))$, $\rho_X(h) = \gamma_X(h) / \gamma_X(0)$, given by

$$T_n = \left(\frac{\sum_{t=1}^{n-k} X_t X_{t+k}}{\sum_{t=1}^{n} X_t^2}, \quad k = 1, 2, ..., m-1 \right),$$

(iv) For the same specification of the function g as in (ii) and $f : \mathbb{R}^m \to \mathbb{R}^{m-1}$ defined as

$$f(y_0, y_1, ..., y_{m-1}) = W_{m-1}^{-1} w_{m-1}$$

where W_{m-1} is the non-singular $(m-1) \times (m-1)$ matrix given by $W_{m-1} = \left(y_{|i-j|} \right)_{i,j=1,2,..,m-1}$ and w_{m-1} the $(m-1)-$dimensional vector $w_{m-1} = (y_0, y_1, ..., y_{m-1})^\top$, then T_n is given by

$$T_n \equiv (T_{1,n}, T_{2,n}, ..., T_{m-1,n})^\top = \widehat{\Gamma}_{m-1}^{-1} \widehat{\gamma}_{m-1}.$$

Here, $\widehat{\Gamma}_{m-1} = (\widehat{\gamma}_X(i-j))_{,j=1,2,..,m-1}$ and $\widehat{\gamma}_{m-1} = (\widehat{\gamma}_X(1), \widehat{\gamma}_X(2), ..., \widehat{\gamma}_X(m-1))^\top$. Recall from Brockwell and Davis (1991), Proposition 5.1.1, that $\widehat{\gamma}_X(0) > 0$ and $\widehat{\Gamma}_k = (\widehat{\gamma}_X(i-j))_{i,j=1,2,..,k}$, is non-singular for every $k \in \mathbb{N}$. For $m = p+1$, T_n is the Yule-Walker estimator of $a_p = (a_{1,p}, a_{2,p}, ..., a_{p,p})^\top$, where a_p are the coefficients of the best (in the mean square sense) linear approximation of X_t based on it past values $X_{t-1}, ..., X_{t-p}$, that is,

$$E\Big(X_t - \sum_{j=1}^{p} a_{j,p} X_{t-j}\Big)^2 = \min_{r_1,\dots r_p} E\Big(X_t - \sum_{j=1}^{p} r_j X_{t-j}\Big)^2.$$

Furthermore, notice that since $T_{m-1,n}$ is an estimator of the lag $(m-1)$ partial autocorrelation; see Brockwell and Davis (1991), Definition 3.4.3, the class of statistics (1.7) includes sample partial autocorrelations as special cases. The failure of the AR-sieve in this case is due to fact that the uncorrelated innovations e_t appearing in representation (1.6) are replaced by i.i.d. pseudo-innovations, i.e., in contrast to (1.6), the process generated by the AR-sieve bootstrap is a linear process driven by i.i.d. innovations. What is striking, however, is the fact that the AR-sieve bootstrap may fail for the above class of statistics even if the underlying process $\mathbf{X}$ is linear, that is, if X_t is generated as

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \tag{1.8}$$

where $\{\varepsilon_t, t \in \mathbb{Z}\}$ consists of i.i.d. random variables with zero mean, variance $0 < \sigma_\varepsilon^2 < \infty$, finite fourth moments $E(\varepsilon_t^4) < \infty$ and the coefficients ψ_j satisfy

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

Notice that the last summability condition implies that $\mathbf{X}$ possess a continuous spectral density. If this spectral density is also everywhere positive, then the linear process $\mathbf{X}$ also obeys the autoregressive representation (1.6). For the AR-sieve bootstrap, however, the important point is that the stochastic properties of the two innovation processes, that is of $\{\varepsilon_t\}$ in representation (1.8) and of $\{e_t\}$ in representation (1.6), could be different. For instance, the fourth order moments of e_t and of ε_t, may be different. This implies, that if the (limiting) distribution of some statistic of interest depends on the fourth order moments of the innovations, then the AR-sieve bootstrap may fail even if the underlying process is linear. Empirical autocovariances are a prominent example of such statistics; see Kreiss et al. (2011) for details. Notice that for linear processes, and apart from the Gaussian case, the two innovation sequences in representations (1.6) and (1.8) are identical if the underlying process is causal and invertible, that is, if

$$\psi_j = 0 \quad \text{for} \quad j < 0$$

and the power series

$$\Psi(z) \equiv 1 + \sum_{j=1}^{\infty} \psi_j z^j$$

has no roots for $|z| \leq 1$. Since in this specific case $\Psi(z)$ is invertible, X_t can be expressed as

$$X_t = \sum_{j=1}^{\infty} a_j X_{t-j} + \varepsilon_t,$$

where

$$A(z) \equiv 1 - \sum_{j=1}^{\infty} a_j z^j = \Psi^{-1}(z).$$

This representation is identical to the AR-representation (1.6) of X_t, i.e., $\{e_t\}$ and $\{\varepsilon_t\}$ are, in this case, identical innovation processes. $\mathbf{X}$ is then a so-called linear, infinite order autoregressive process, see Brockwell and Davis (1991). It is well-known that for this specific class of processes, the AR-sieve bootstrap works for the entire class of generalized means (1.7); see also Bühlmann (1997).

In the third chapter of the thesis, we proceed with a re-examination of the well known AR-sieve bootstrap for time series. Motivated by some limitations of this bootstrap procedure in the case of linear processes, we propose a modification of the standard AR-seive bootstrap procedure which is based on a modification of the resampling procedure used to generate the i.i.d pseudo-innovations. The difference is that the pseudo-innovations are generated in a way which correctly captures (asymptotically) the fourth order moment structure of the true innovations. For this, the estimation results of the rescaled fourth order cumulant discussed in Chapter 2 are used. This procedure is called the AR-seive with i.i.d. wild innovations and it is shown that this modification extends the range of validity of the standard AR-seive bootstrap.

Furthermore, another AR-bootstrap procedure has been proposed, which is applied to approximate the distribution of a large class of statistics, the class of so-called generalized means given by (1.7). The basic idea is to apply the AR-bootstrap not to the obsereved time series itself but to the thransformed time series which is used in the estimator of the parameter of interest. This novel idea is investigated more closely in the Chapter 3 of the thesis and a bootstrap procedure for the class of statistics considered is proposed. The procedure is called the AR-bootstrap with dependent wild innovations and it is proven that this procedure ia asymptotically valid for general classes of stochastic processes.

Chapter 2

Inference for the Fourth Order Innovation Cumulant in Linear Time Series

2.1 Introduction

In this chapter we investigate more closely the problem of estimating the parameter $\eta_{4,\varepsilon}$. We first consider a nonparametric estimator of this parameter, the origins of which go back to Grenander and Rosenblatt (1957). This nonparametric estimator has been also used by Janas and Dahlhaus (1994), while a simpler and computationally more tractable version has been applied in Kreiss and Paparoditis (2012). We derive the asymptotic properties of this estimator and show how the entire autocorrelation structure of the underlying linear process affects the behavior of the estimator. Our theoretical deviations show that, the more correlated the time series is the worse is the estimator, both in terms of bias and of variance. Motivated by these findings, a new nonparametric estimator of $\eta_{4,\varepsilon}$ is proposed, which exploits a basic invariance property of the parameter of interest with respect to linear filtering of the time series. The new estimator is based on pre-whitening the time series by means of an autoregressive filter, the coefficients of which are determined in a way that reduces the correlation of the time series at hand. The aforementioned invariance property implies that the parameter of interest can be estimated using the residuals obtained by applying the linear filter to the time series at hand and an inverse-transformation is not needed.

The asymptotic properties of the new estimator are investigated and its superiority for large classes of stochastic processes is established. Our derivations allow also for some interesting applications. For instance, we propose a modification of the multiplicative periodogram bootstrap investigated by Franke and Härdle (1992) which is able to imitate the weak dependence structure of the periodogram. We also consider the problem of testing hypotheses about the parameter $\eta_{4,\varepsilon}$. In this context, a simple bootstrap-based test for the important null hypothesis

$$H_0 : \eta_{4,\varepsilon} = 0$$

is proposed. Notice that rejection of this hypothesis can be interpreted as rejection of a hypothesized Gaussianity of the time series. Finally, simulations show a clearly improved performance of the new estimator compared to the existing estimator, where the gains in terms of variance and bias reduction, especially for correlated time series, could be very impressive.

The remaining of the chapter is organized as follows. Section 2.2 discusses the nonparametric estimator for $\eta_{4,\varepsilon}$ proposed so far in the literature and investigates its asymptotic behavior. Section 2.3 introduces the new nonparametric estimator and derives its asymptotic properties. Theoretical comparisons are made in Section 2.4, while applications of the results obtained for bootstrapping the periodogram and for testing hypotheses about the parameter $\eta_{4,\varepsilon}$ are discussed in Section 2.5. The issue of the practical choice of the filtering and of the smoothing parameters, involved in the estimation procedure, is addressed in Section 2.6. This section presents also several simulations that verify our theoretical findings and demonstrate the superiority of the new nonparametric estimator proposed. Some interesting applications to real-life data are also discussed. All technical proofs are deferred to the Section 2.7.

2.2 Nonparametric Estimation

2.2.1 Assumptions and Estimators

Throughout the chapter we assume that the underlying stochastic process $\mathbf{X} = \{X_t : t \in \mathbb{Z}\}$ which generates the observed time series $X_1, X_2, ..., X_n$ follows equation (1.1) and that the following condition is satisfied.

Assumption 2.1. *$\sum_{j\in\mathbb{Z}} j^2\psi_j^2 < \infty$, $\psi_0 = 1$ and $\{\varepsilon_t, t \in \mathbb{Z}\}$ consists of i.i.d. random variables with $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma_\varepsilon^2$ and $E(\varepsilon_t^8) < \infty$.*

The requirement of finiteness of eight moments of the innovations seems in-avoidable since our derivations include calculations of the variance of estimators which are functions of moments up to fourth order of the ε_t's.
Assumption 2.1 implies that

$$\sum_{h=-\infty}^{\infty} h^2 \left|\gamma_X(h)\right| = \sigma_\varepsilon^2 \sum_{h=-\infty}^{\infty} h^2 \left|\sum_{j=-\infty}^{\infty} \psi_j \psi_{j+|h|}\right| \leq \sigma_\varepsilon^2 \sum_{j=-\infty}^{\infty} \left|\psi_j\right| \sum_{h=-\infty}^{\infty} h^2 \left|\psi_{j+|h|}\right| < \infty,$$

and, therefore, the process $\mathbf{X}$ possesses a spectral density f_X given by

$$f_X(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_X(h) \cos(\lambda h), \quad \lambda \in (-\pi, \pi],$$

which is twice continuously differentiable. Denote by

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}$$

the autocorrelation at lag $h \in \mathbb{Z}$ of $\mathbf{X}$.

Recall that our aim is to estimate the parameter

$$\eta_{4,\varepsilon} = \frac{\kappa_{4,\varepsilon}}{\sigma_\varepsilon^4}$$

where

$$\kappa_{4,\varepsilon} = \text{cum}_4(\varepsilon_t) = E(\varepsilon_t^4) - 3\sigma_\varepsilon^4$$

is the fourth order cumulant of ε_t. Following Grenander and Rosenblatt (1957), Ch. 6.5, the covariance $\gamma_{2,X}(h) \equiv \text{Cov}(X_t^2, X_{t+h}^2)$, is given by

$$\gamma_{2,X}(h) = (E\varepsilon_t^4 - 3(E\varepsilon_t^2)^2) \sum_{j=-\infty}^{\infty} \psi_j^2 \psi_{h+j}^2 + 2\gamma_X^2(h), \tag{2.1}$$

hence

$$\sum_{h=-\infty}^{\infty} \gamma_{2,X}(h) = \kappa_{4,\varepsilon} \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j^2 \psi_{h+j}^2 + 2 \sum_{h=-\infty}^{\infty} \gamma_X^2(h). \tag{2.2}$$

Using the equality

$$\sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j^2 \psi_{h+j}^2 = \frac{\gamma_X^2(0)}{\sigma_\varepsilon^4},$$

we get the basic expression

$$\eta_{4,\varepsilon} = \frac{1}{\gamma_X^2(0)} \sum_{h=-\infty}^{\infty} \left(\gamma_{2,X}(h) - 2\gamma_X^2(h)\right), \tag{2.3}$$

which relates the parameter $\eta_{4,\varepsilon}$ of interest to the autocovariances of $\mathbf{X}$ and to that of the squared process $\mathbf{X}^2 = \{X_t^2, t \in \mathbb{Z}\}$. Based on (2.3), Kreiss and Paparoditis (2012) proposed a lag-window estimator of $\eta_{4,\varepsilon}$, given by

$$\breve{\eta}_{4,\varepsilon} = \frac{1}{\widehat{\gamma}_X^2(0)} \sum_{h=-(n-1)}^{n-1} w\left(\frac{h}{M_n}\right)\left(\widehat{\gamma}_{2,X}(h) - 2\widehat{\gamma}_X^2(h)\right). \tag{2.4}$$

Here

$$\widehat{\gamma}_X(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(X_t - \overline{X}_n)(X_{t+|h|} - \overline{X}_n),\ \widehat{\gamma}_{2,X}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(X_t^2 - \overline{X}_{2,n})(X_{t+|h|}^2 - \overline{X}_{2,n}),$$

$$\overline{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t \quad \text{and} \quad \overline{X}_{2,n} = \frac{1}{n}\sum_{t=1}^{n} X_t^2,$$

are sample estimates of the corresponding unknown quantities appearing in (2.3). Furthermore, w is a so-called lag-window and $M_n = M(n) < n$ is a truncation parameter. The lag window and the truncation parameter are assumed to satisfy the following conditions.

Assumption 2.2.

(i) $w : [-1, 1] \to \mathbb{R}$ *is a symmetric, non-negative and continuous function and satisfies*

$$w(x) = \int_{-\infty}^{\infty} K(u) e^{-iux} du,$$

where K is a non-negative kernel function. Furthermore,

$$\int_{-\infty}^{\infty} w^2(u) du < \infty.$$

(ii) $\frac{1}{M_n} + \frac{M_n^4}{n} \to 0$ *as* $n \to \infty$.

Notice that Grenander and Rosenblatt (1957) proposed a different but asymptotically equivalent estimator of the parameter $\eta_{4,\varepsilon}$ which is based on equation (2.3). This estimator is obtained using the relations

$$\sum_{h=-\infty}^{\infty} \gamma_X^2(h) = 2\pi \int_{-\pi}^{\pi} f_X^2(\lambda)\, d\lambda, \quad \sum_{h=-\infty}^{\infty} \gamma_{2,X}(h) = 2\pi f_{X^2}(0),$$

and

$$\gamma_X(0) = \int_{-\pi}^{\pi} f_X(\lambda)\, d\lambda,$$

where

$$f_{X^2}(\lambda) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} \gamma_{2,X}(h)\cos(\lambda h)$$

is the spectral density of the squared process $\mathbf{X}^2$. It is then easily seen that

$$\eta_{4,\varepsilon} = \frac{2\pi f_{X^2}(0) - 4\pi \int_{-\pi}^{\pi} f_X^2(\lambda)d\lambda}{\left(\int_{-\pi}^{\pi} f_X(\lambda)d\lambda\right)^2},$$

which leads to the following frequency domain version of $\breve{\eta}_4$,

$$\overline{\eta}_{4,\varepsilon} = \frac{2\pi \widehat{f}_{X^2}(0) - 4\pi \int_{-\pi}^{\pi} \widehat{f}_X^2(\lambda)d\lambda}{\left(\int_{-\pi}^{\pi} \widehat{f}_X(\lambda)d\lambda\right)^2}.$$

In the above expression,

$$\widehat{f}_X(\lambda) = \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} w\left(\frac{h}{M_n}\right) \widehat{\gamma}_X(h)\cos(\lambda h),$$

and

$$\widehat{f}_{X^2}(\lambda) = \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} w\left(\frac{h}{M_n}\right) \widehat{\gamma}_{2,X}(h)\cos(\lambda h),$$

are lag window estimators of the spectral densities $f_X(\lambda)$ and $f_{X^2}(\lambda)$ respectively; see also Janas and Dahlhaus (1994).

2.2.2 Asymptotic Properties

Kreiss and Paparoditis (2012) established consistency of the estimator $\breve{\eta}_{4,\varepsilon}$. The following theorem extends this result by establishing several additional properties of this estimator.

Theorem 2.2.1. *Suppose that Assumption 2.1 and Assumption 2.2 are satisfied. Then, as $n \to \infty$,*

(i) $M_n^2(E(\breve{\eta}_{4,\varepsilon}) - \eta_{4,\varepsilon}) \to \int_{-1}^{1} x^2 w(x)dx \frac{\kappa_{4,\varepsilon}}{\sigma_\varepsilon^2 \gamma_X^2(0)} 2\pi f_{\widetilde{X}}''(0),$

(ii) $\frac{n}{M_n} Var\left(\breve{\eta}_{4,\varepsilon}\right) \to \tau_X^2 = 2\int_{-1}^{1} w^2(x)dx \Big(\eta_{4,\varepsilon} + 2\sum_{h=-\infty}^{\infty} \rho_X^2(h)\Big)^2,$

(iii) $\sqrt{\frac{n}{M_n}}\left(\breve{\eta}_{4,\varepsilon} - E\left(\breve{\eta}_{4,\varepsilon}\right)\right) \xrightarrow{D} N\left(0, \tau_X^2\right).$

Here, $f_{\widetilde{X}}''$ denotes the second derivative of the spectral density of the process $\widetilde{\mathbf{X}} = \left\{\widetilde{X}_t : t \in \mathbb{Z}\right\}$, where $\widetilde{X}_t = \sum_{j=-\infty}^{\infty} \psi_j^2 \varepsilon_{t-j}$.

The results of Theorem 2.2.1 highlight several interesting features of $\breve{\eta}_{4,\varepsilon}$. First of all, the variance of this estimator, given in part (ii) of the above theorem, has a simple form and depends, apart from the parameter $\eta_{4,\varepsilon}$ itself, also on the entire (squared) autocorrelation structure of the underlying process $\mathbf{X}$. In fact, the larger $\sum_{h=1}^{\infty} \rho_X^2(h)$ is, the larger is the variability of $\breve{\eta}_{4,\varepsilon}$. Furthermore, and since

$$2\pi f''_{\widetilde{X}}(0) = -\sum_{h=-\infty}^{\infty} h^2 \gamma_{\widetilde{X}}(h),$$

with $\gamma_{\widetilde{X}}(\cdot)$ the autocovariance function of the process $\widetilde{\mathbf{X}}$, the behavior of the bias given in part (i) depends, among other things, also on the autocovariance structure of $\widetilde{\mathbf{X}}$. Interestingly, for $\eta_4 = 0$ the bias term disappears, that is,

$$M_n^2(E(\breve{\eta}_{4,\varepsilon}) - \eta_{4,\varepsilon}) \to 0 \quad \text{as} \quad n \to \infty.$$

Finally, if $\mathbf{X}$ is an i.i.d. process, that is, if $X_t = \varepsilon_t$, then

$$M_n^2(E(\breve{\eta}_{4,\varepsilon}) - \eta_{4,\varepsilon}) \to 0 \quad \text{and} \quad \frac{n}{M_n} Var\,(\breve{\eta}_{4,\varepsilon}) \to 2\int_{-1}^{1} w^2(x)dx\Big(\eta_{4,\varepsilon} + 2\Big)^2 \qquad (2.5)$$

as $n \to \infty$ where the last limit above is the lower bound that the variance of the nonparametric estimator $\breve{\eta}_{4,\varepsilon}$ can achieve over all linear processes $\mathbf{X}$ which have i.i.d. innovations ε_t with given first, second and fourth order moments.

2.3 An Improved Nonparametric Estimator

According to the results presented in the previous section, the behavior of the estimator $\breve{\eta}_{4,\varepsilon}$ is seriously affected by the autocorrelation structure of the underlying process $\mathbf{X}$, that is, the more correlated the time series is, the worse is the estimator $\breve{\eta}_{4,\varepsilon}$. In this section we propose another estimator of the parameter of interest. Toward this goal, the following two observations are important.

First, the target parameter $\eta_{4,\varepsilon}$ is invariant with respect to linear transformations of the time series. To elaborate, let B be the shift operator, i.e.,

$$B^k X_t = X_{t-k} \quad \text{for} \quad k \in \mathbb{Z},$$

and write

$$X_t = \Psi(B)\varepsilon_t, \quad \text{where} \quad \Psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j.$$

Let

$$A(B) = \sum_{j\in\mathbb{Z}} a_j B^j$$

be a linear filter. Then the filtered process $\mathbf{Y} = \{Y_t, t \in \mathbb{Z}\}$ with

$$Y_t = A(B)X_t = A(B)\Psi(B)\varepsilon_t$$

is obviously linear and both processes $\mathbf{X}$ and $\mathbf{Y}$ are driven by the same i.i.d. innovations $\{\varepsilon_t, t \in \mathbb{Z}\}$. Thus, applying the estimator $\breve{\eta}_{4,\varepsilon}$ to a time series $Y_1, Y_2, ..., Y_n$ stemming from the filtered process $\mathbf{Y}$, or to a time series $X_1, X_2, ..., X_n$ stemming from the original process $\mathbf{X}$, estimates the same target parameter $\eta_{4,\varepsilon}$.

Second, given the above invariance property, the behavior of the estimator can be improved by reducing the correlation of the time series at hand by means of applying an appropriate linear filter $A(B)$. Ideally, such a filter will transform the observed time series to an uncorrelated sequence. For this purpose, different filtering approaches can be considered. For instance, following Brockwell and Davis (1988) we can apply a moving average filter by means of the innovations algorithm. Computationally more attractive is a pre-whitening approach which is based on applying an autoregressive filter. This approach has a long history in time series analysis, see Press and Tukey (1956). It transforms the time series by fitting an autoregressive process, that is, by considering the process $\mathbf{U} = \{U_{t,p}, t \in \mathbb{Z}\}$, where

$$U_{t,p} = X_t - \sum_{j=1}^{p} a_{j,p} X_{t-j}.$$

Here $(a_{1,p}, a_{2,p}, ..., a_{p,p})^\top$ are the coefficients of the best (in the mean square sense) approximation of X_t by means of a linear combination of the past p values $X_{t-1}, ..., X_{t-p}$. It is well known that, under Assumption 2.1, the coefficients $a_p = (a_{1,p}, a_{2,p}, \ldots, a_{p,p})^\top$ are uniquely determined and given by

$$a_p = \Gamma_p^{-1}\gamma_p,$$

where

$$\Gamma_p = (\gamma_X(i-j))_{i,j=1,2,\ldots,p} \quad \text{and} \quad \gamma_p = (\gamma_X(j), j = 1, 2, \ldots, p)^\top,$$

cf. Proposition 5.1.1 of Brockwell and Davis (1991). Notice that because of the invariance of the parameter $\eta_{4,\varepsilon}$ with respect to linear filtering, we get, along the same lines as in obtaining (2.3),

$$\eta_{4,\varepsilon} = \frac{1}{\gamma_U^2(0)} \sum_{h=-\infty}^{\infty} \big(\gamma_{2,U}(h) - 2\gamma_U^2(h)\big), \tag{2.6}$$

where $\gamma_U(h) = Cov(U_{t,p}, U_{t+h,p})$ and $\gamma_{2,U}(h) = Cov(U_{t,p}^2, U_{t+h,p}^2)$ denote the autocovariance of the filtered process $\{U_{t,p}, t \in \mathbb{Z}\}$ and of the squared filtered process $\{U_{t,p}^2, t \in \mathbb{Z}\}$, respectively.

Summarizing the observations made so far, the following procedure to estimate the parameter $\eta_{4,\varepsilon}$ is suggested. Let

$$\widehat{U}_{t,p} = X_t - \sum_{j=1}^{p} \widehat{a}_{j,p} X_{t-j}, \quad t = p+1, p+2, \ldots, n,$$

where $\widehat{a}_{j,p}$, $j = 1, 2, ..., p$ are the Yule-Walker estimators of $a_{j,p}$, $j = 1, 2, \ldots, p$, obtained by replacing the autocovariances $\gamma_X(h)$ in the expression $a_p = \Gamma_p^{-1}\gamma_p$ by the corresponding sample autocovariances $\widehat{\gamma}_X(h)$. Then, the alternative nonparametric estimator of $\eta_{4,\varepsilon}$ we propose is given by

$$\widehat{\eta}_{4,\varepsilon} = \frac{1}{\widehat{\gamma}_U^2(0)} \sum_{h=-(N-1)}^{N-1} w\left(\frac{h}{M_n}\right)\left(\widehat{\gamma}_{2,U}(h) - 2\widehat{\gamma}_U^2(h)\right), \tag{2.7}$$

where $N = n - p$,

$$\widehat{\gamma}_U(h) = \frac{1}{N} \sum_{t=p+1}^{n-|h|} \left(\widehat{U}_{t,p} - \overline{U}_n\right)\left(\widehat{U}_{t+|h|,p} - \overline{U}_n\right), \quad \overline{U}_n = \frac{1}{N} \sum_{t=p+1}^{n} \widehat{U}_{t,p},$$

$$\widehat{\gamma}_{2,U}(h) = \frac{1}{N} \sum_{t=p+1}^{n-|h|} \left(\widehat{U}_{t,p}^2 - \overline{U}_{2,n}\right)\left(\widehat{U}_{t+|h|,p}^2 - \overline{U}_{2,n}\right) \text{ and } \overline{U}_{2,n} = \frac{1}{N} \sum_{t=p+1}^{n} \widehat{U}_{t,p}^2.$$

We stress here the fact that we do not assume that the time series at hand stems from an autoregressive process. We rather use the autoregressive fit solely as a filtering approach in order to reduce the correlation of the observations which will be used in the nonparametric estimator of the parameter of interest. In view of the results obtained in Theorem 2.2.1, we expect that such a filtering will improve the estimator. In the following theorem we first summarize the asymptotic properties of the new estimator $\widehat{\eta}_{4,\varepsilon}$. Comparisons with $\breve{\eta}_{4,\varepsilon}$ are given in the next section.

Theorem 2.3.1. *Suppose that Assumption 2.1 and Assumption 2.2 are satisfied and let $p \in \mathbb{N}$ be fixed. Then, as $n \to \infty$*

(i) $M_n^2(E(\widehat{\eta}_{4,\varepsilon}) - \eta_{4,\varepsilon}) \to \int_{-1}^{1} x^2 w(x)dx \frac{\kappa_{4,\varepsilon}}{\sigma_\varepsilon^2 \gamma_U^2(0)} 2\pi f_{\widetilde{U}}''(0)$,

(ii) $\frac{n}{M_n} Var\left(\widehat{\eta}_{4,\varepsilon}\right) \to \tau_U^2 = 2\int_{-1}^{1} w^2(x)dx \Big(\eta_{4,\varepsilon} + 2\sum_{h=-\infty}^{\infty} \rho_U^2(h)\Big)^2$,

(iii) $\sqrt{\frac{n}{M_n}}\left(\widehat{\eta}_{4,\varepsilon} - E\left(\widehat{\eta}_{4,\varepsilon}\right)\right) \xrightarrow{D} N\left(0, \tau_U^2\right)$,

where $\rho_U(h) = \gamma_U(h)/\gamma_U(0)$ *is the autocorrelation of the process* $\{U_{t,p}, t \in \mathbb{Z}\}$ *and* $f''_{\widetilde{U}}(\lambda)$ *denotes the second derivative of the spectral density of the process* $\widetilde{\mathbf{U}} = \left\{\widetilde{U}_{t,p} : t \in \mathbb{Z}\right\}$*, where* $\widetilde{U}_{t,p} = \sum_{j=-\infty}^{\infty} c_{j,p}^2 \varepsilon_{t-j}$ *and the coefficients* $\{c_{j,p}, j \in \mathbb{Z}\}$ *are given by* $\sum_{j=-\infty}^{\infty} c_{j,p} z^j = (1 - \sum_{j=1}^{p} a_{j,p} z^j)(\sum_{j=-\infty}^{\infty} \psi_j z^j)$*.*

2.4 Comparisons

In this section we compare the estimators $\breve{\eta}_{4,\varepsilon}$ and $\widehat{\eta}_{4,\varepsilon}$ based on the asymptotic results obtained in the previous sections. For this, we first impose the following condition on the underlying process $\mathbf{X}$.

Assumption 2.3. *The spectral density of* $\mathbf{X}$ *satisfies*

$$\inf_{\lambda \in [0,\pi]} f_X(\lambda) > 0.$$

Assumption 2.3 restricts slightly the class of linear processes considered since it excludes processes for which the power series

$$\Psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j, \quad z \in \mathbb{C},$$

has zeros on the unit disc. Recall that by Assumption 2.1 the spectral density $f_X(\lambda)$ of $\mathbf{X}$ is continuous. This together with Assumption 2.3 above, implies that the process $\mathbf{X}$ obeys a so-called autoregressive (AR-) representation, that is, X_t can be expressed as

$$X_t = \sum_{j=1}^{\infty} a_j X_{t-j} + v_t, \tag{2.8}$$

where the a_j's are defined as the coefficients of the best (in the mean square sense) linear approximation of X_t based on its infinite past $\{X_{t-1}, X_{t-2}, \ldots\}$. Furthermore, $\{v_t, t \in \mathbb{Z}\}$ is a white noise innovation process, i.e.,

$$E(v_t) = 0, \quad Var(v_t) = \sigma_v^2, \quad Cov(v_t, v_s) = 0$$

for $t \neq s$ and which may be different from the i.i.d. innovation process $\{\varepsilon_t\}$ appearing in (1.1); see Pourahmadi (2001) for details. We stress the fact that the autoregressive representation (2.8) of X_t with respect to the white noise innovations $\{v_t, t \in \mathbb{Z}\}$ is different and should not be confused with the linear AR(∞) representation

$$X_t = \sum_{j=1}^{\infty} \pi_j X_{t-j} + \varepsilon_t$$

with respect to the i.i.d. innovation process ε_t, which exists if $\mathbf{X}$ is a causal and invertible linear process. That is, if the coefficients ψ_j in (1.1) satisfy

$$\psi_j = 0 \quad \text{for} \quad j < 0 \quad \text{and} \quad \Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j \neq 0 \quad \text{for} \quad |z| \leq 1;$$

cf. Brockwell and Davis (1991). In fact, representation (2.8) of a purely stationary stochastic process, possessing a continuous and positive spectral density f_X, is an autoregressive analog to the well-known, moving average Wold representation, see Brockwell and Davis (1991), Chapter 5.7. To give an example, the non-invertible, first order moving average process

$$X_t = \theta\varepsilon_{t-1} + \varepsilon_t \quad \text{with} \quad |\theta| > 1,$$

does not possess a linear AR(∞) representation with respect to the i.i.d. innovations ε_t. However, it has the AR-representation

$$X_t = \sum_{j=1}^{\infty} a_j X_{t-j} + v_t,$$

with respect to the white noise series $\{v_t, t \in \mathbb{Z}\}$, where

$$a_j = -(-1/\theta)^j, \; j = 1, 2, \ldots \quad \text{and} \quad v_t = \varepsilon_t + (1-\theta^2) \sum_{j=1}^{\infty} (-\theta)^{-j} \varepsilon_{t-j};$$

see also Kreiss et al. (2011).

Consider first the variance of the two estimators $\breve{\eta}_{4,\varepsilon}$ and $\widehat{\eta}_{4,\varepsilon}$. For this it suffices to compare the two series

$$\sum_{h=-\infty}^{\infty} \rho_X^2(h) \quad \text{and} \quad \sum_{h=-\infty}^{\infty} \rho_U^2(h).$$

Interpreting,

$$\|\rho_X\|_2 = \sqrt{\sum_{h=-\infty}^{\infty} \rho_X^2(h)} \quad \text{and} \quad \|\rho_U\|_2 = \sqrt{\sum_{h=-\infty}^{\infty} \rho_U^2(h)},$$

as global measures of correlation of the processes $\mathbf{X} = \{X_t : t \in \mathbb{Z}\}$ and $\mathbf{U} = \{U_{t,p} : t \in \mathbb{Z}\}$ respectively, it yields that if $\mathbf{U}$ is less correlated than $\mathbf{X}$, that is, if

$$\|\rho_U\|_2 < \|\rho_X\|_2,$$

then the estimator $\widehat{\eta}_{4,\varepsilon}$ will be (asymptotically) more efficient than the estimator $\breve{\eta}_{4,\varepsilon}$.

The following theorem shows that filtering indeed achieves this goal and that the advantages of filtering in terms of variance reduction are uniform over the order p of the autoregressive filter used, provided this order exceeds some given value p_0, which depends on characteristics of the underlying process $\mathbf{X}$. Furthermore, allowing for the order p of the autoregressive filter used, to increase to infinity at an appropriate rate, the pre-whitened estimator $\widehat{\eta}_{4,\varepsilon}$ becomes asymptotically efficient, that is, $Var(\widehat{\eta}_{4,\varepsilon})$ achieves the lower bound given in (2.5) which corresponds to the case where the underlying process is uncorrelated.

Theorem 2.4.1. *Suppose that Assumption 2.1, Assumption 2.2 and Assumption 2.3 are satisfied.*

(i) There exist $p_0 \in \mathbb{N}$ (p_0 depends on the process $\mathbf{X}$), such that for all $p \in \mathbb{N}$ with $p \geq p_0$, it yields that

$$\lim_{n\to\infty} \frac{Var\left(\widehat{\eta}_{4,\varepsilon}\right)}{Var\left(\breve{\eta}_{4,\varepsilon}\right)} < 1. \tag{2.9}$$

(ii) If $p = p(n) \to \infty$ such that $p = o((n/log(n))^{1/4})$, then

$$\lim_{n\to\infty} \frac{n}{M_n} Var(\widehat{\eta}_{4,\varepsilon}) = 2\int_{-1}^{1} w^2(x)dx\big(2+\eta_{4,\varepsilon}\big)^2. \tag{2.10}$$

Notice that p_0 could be as small as $p_0 = 1$. For instance, for the first order moving average process

$$X_t = \varepsilon_t + \theta\varepsilon_{t-1} \quad \text{with} \quad \theta \neq 0,$$

it is true that for any $p \geq 1$, the variance of the estimator $\widehat{\eta}_{4,\varepsilon}$ is strictly smaller than that of the originally proposed estimator $\breve{\eta}_{4,\varepsilon}$. For example, by simple algebra and using $p = 1$ one gets

$$\sum_{h=-\infty}^{\infty} \rho_U^2(h) = 2\left\{\frac{\rho_X^4(1)+\rho_X^2(1)}{\left(1-\rho_X^2(1)\right)^2}\right\}\rho_X^2(1) + 1 < 2\rho_X^2(1) + 1 = \sum_{h=-\infty}^{\infty} \rho_X^2(h),$$

where the last inequality follows because

$$\frac{\rho_X^4(1)+\rho_X^2(1)}{(1-\rho_X^2(1))^2} < 1,$$

due to the fact that, for any linear first order moving average process it is true that

$$|\rho_X(1)| \leq \frac{1}{2}.$$

We next consider the biases of the estimators $\breve{\eta}_{4,\varepsilon}$ and $\widehat{\eta}_{4,\varepsilon}$. A comparison of the bias expressions given in part (ii) of Theorem 2.2.1 and Theorem 2.3.1 respectively, requires, among other things, the comparison of the second derivatives at frequency zero of the spectral densities of the processes $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{U}}$ respectively. Although it seems difficult to derive results for the most general case, such results can be established for the important case where the underlying linear process $\mathbf{X}$ is causal and invertible.

Assumption 2.4. *The process* $\mathbf{X}$ *generated as in (1.1) satisfies*

$$\psi_j = 0 \quad \textit{for} \quad j < 0 \quad \textit{and} \quad \Psi(z) = 1 + \sum_{j=1}^{\infty} \psi_j z^j \neq 0 \quad \textit{for} \quad |z| \leq 1.$$

As previously mentioned, Assumption 2.4 is stronger than Assumption 2.3. In fact if Assumption 2.4 is true, then X_t can be expressed as a linear AR(∞) process with respect to the i.i.d. innovations ε_t, i.e.,

$$X_t = \sum_{j=1}^{\infty} c_j X_{t-j} + \varepsilon_t,$$

where the c_j, $j = 1, 2, \ldots$ are uniquely determined as the coefficients of z^j, $j = 1, 2, \ldots$ of $\Psi^{-1}(z)$. The next theorem summarizes our findings regarding the comparison of the biases of the two nonparametric estimators considered.

Theorem 2.4.2. *Suppose that Assumption 2.1, Assumption 2.2 and Assumption 2.4 are satisfied.*

(i) There exist $p_0 \in \mathbb{N}$ *(*p_0 *depends on the process* $\mathbf{X}$*), such that for all* $p \in \mathbb{N}$ *with* $p \geq p_0$*, it yields that*

$$\lim_{n\to\infty} M_n^2 \left(|E(\widehat{\eta}_{4,\varepsilon}) - \eta_{4,\varepsilon}| - |E(\breve{\eta}_{4,\varepsilon}) - \eta_{4,\varepsilon}|\right) < 0. \tag{2.11}$$

(ii) If $p = p(n) \to \infty$ *such that* $p = o((n/log(n))^{1/4})$*, then*

$$M_n^2(E(\widehat{\eta}_{4,\varepsilon}) - \eta_{4,\varepsilon}) = o(1). \tag{2.12}$$

As the above theorem shows, if $\mathbf{X}$ belongs to the important class of causal and invertible linear processes, then the advantages of pre-whitening are extended also to the bias of the new nonparametric estimator $\widehat{\eta}_{4,\varepsilon}$.

2.5 Applications

2.5.1 Periodogram Bootstrap

As mentioned in the Introduction, the multiplicative periodogram bootstrap procedure investigated by Franke and Härdle (1992) and Dahlhaus and Janas (1996), consists of generating pseudo periodogram ordinates as

$$I_X^*(\lambda_j) = \widehat{f}_X(\lambda_j) U_j^*, \quad \lambda_j = \frac{2\pi j}{n}, \quad j = 1, 2, \dots \left[\frac{n}{2}\right],$$

where the U_j^*'s are independent pseudo-random variables designed to mimic the behavior of $I_e(\lambda_j)$ which is the periodogram of the standardized innovations $e_t = \varepsilon_t/\sigma_\varepsilon$. The independence of the U_j^*'s and, consequently, of the periodogram ordinates $I_X^*(\lambda_j)$, restrict the range of validity of this frequency domain bootstrap procedure to statistics the distribution of which does not depend on the fourth order moments of ε_1.

Using the improved estimator $\widehat{\eta}_{4,\varepsilon}$ of $\eta_{4,\varepsilon}$ proposed in this chapter, we can easily modify this bootstrap procedure to overcome this limitation; see also Kreiss and Paparoditis (2012). This can be achieved as follows: Generate i.i.d. random variables $e_1^+, e_2^+, \dots, e_n^+$ such that

$$E(e_1^+) = 0, \quad E(e_1^+)^2 = 1 \quad \text{and} \quad E(e_1^+)^4 = \widehat{\eta}_{4,\varepsilon} + 3.$$

Let

$$I_e^+(\lambda_j) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} e_t^+ \exp\{-it\lambda_j\} \right|^2,$$

be the periodogram of $e_1^+, e_2^+, \dots, e_n^+$ and define bootstrap pseudo periodogram ordinates $I_X^+(\lambda_j)$, $j = 1, 2, \dots, [n/2]$, as

$$I_X^+(\lambda_j) = \widehat{f}_X(\lambda_j) I_e^+(\lambda_j).$$

It is easily seen that for $0 < \lambda_j \neq \lambda_k < \pi$,

$$Cov(I_e^+(\lambda_j), I_e^+(\lambda_k)) = \frac{\widehat{\eta}_{4,\varepsilon}}{4\pi^2 n},$$

that is, $I_e^+(\lambda_j)$ mimics the covariance structure of the periodogram $I_e(\lambda_j)$. The pseudo-periodogram ordinates $I_X^+(\lambda_j)$ can now be used to approximate the distribution of some statistic which is based on the periodogram ordinates $I_X(\lambda_j)$. For instance, consider the important class of spectral means (1.2). For this, we can define

$$S_n^+ = \frac{2\pi}{n} \sum_{\lambda_j \in \mathcal{F}_n} \varphi(\lambda_j) I_X^+(\lambda_j),$$

with

$$I_X^+(\lambda_j) = I_X^+(-\lambda_j), \quad \text{for} \quad \lambda_j < 0,$$

as a discrete bootstrap analogue of S_n and use the distribution of

$$L_n^+ = \sqrt{n}(S_n^+ - \frac{2\pi}{n} \sum_{\lambda_j \in \mathcal{F}_n} \varphi(\lambda_j) \widehat{f}_X(\lambda_j)),$$

to approximate the distribution of

$$L_n = \sqrt{n}(S_n - \int_{[-\pi,\pi]} \varphi(\lambda) f_X(\lambda) d\lambda).$$

Asymptotic validity of this bootstrap proposal can be established along the same lines as in the proof of Theorem 3.1 in Kreiss and Paparoditis (2012).

2.5.2 Testing Hypotheses

Apart from the estimation problem considered so far, the results presented in this chapter allow also for the construction of tests of hypotheses about the parameter $\eta_{4,\varepsilon}$. A special case in this context concerns the test

$$H_0 : \eta_{4,\varepsilon} = 0 \quad \text{vs.} \quad H_1 : \eta_{4,\varepsilon} \neq 0. \tag{2.13}$$

This case is of particular interest for several reasons. First of all, and as we already have seen in the Introduction, the case $\eta_{4,\varepsilon} = 0$ simplifies considerably statistical inference. Furthermore, $\eta_{4,\varepsilon} = 0$ occurs if $\mathbf{X}$ is a Gaussian time series, that is, if X_t obeys the causal representation

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

with innovations ε_t that are i.i.d. Gaussian random variables with mean zero and variance σ_ε^2; see Fan and Yao (2003), Proposition 2.1. In this context, rejection of H_0 can be interpreted as rejection of a hypothesized Gaussianity of the underlying time series. Now, to test hypotheses (2.13), one can exploit the results of Theorem 2.3.1 and use as test statistic the studentized quantity

$$t_n = \frac{\sqrt{n}(\widehat{\eta}_{4,\varepsilon} - E_{H_0}(\widehat{\eta}_{4,\varepsilon}))}{\sqrt{M_n}\,\widehat{\tau}_U}.$$

Here $E_{H_0}(\widehat{\eta}_{4,\varepsilon})$ denotes the expected value of $\widehat{\eta}_{4,\varepsilon}$ under the null hypothesis and $\widehat{\tau}_U$ denotes a consistent estimator of the standard deviation $\tau_U = \sqrt{\tau_U^2}$. Notice that if $\eta_{4,\varepsilon} = 0$ and $M_n = O\left(n^{1/5}\right)$, then

$$\sqrt{\frac{n}{M_n}} E_{H_0}(\widehat{\eta}_{4,\varepsilon}) \to 0,$$

which implies that for testing the null hypothesis (2.13), the test statistic simplifies to

$$t_n = \frac{\sqrt{n}\widehat{\eta}_{4,\varepsilon}}{\sqrt{M_n}\widehat{\tau}_U}.$$

Such a test will reject H_0 whenever

$$|t_n| > z_{1-\alpha/2},$$

where $z_{1-\alpha/2}$ denotes the upper $\alpha/2$-percentage point of the standard Gaussian distribution.

A simpler way to implement a test of hypotheses (2.13), however, is the following. Notice, first that under the null hypothesis, the limiting distribution of

$$\sqrt{\frac{n}{M_n}}(\widehat{\eta}_{4,\varepsilon} - \eta_{4,\varepsilon})$$

depends only on the autocorrelation structure of the filtered process $\{U_{t,p}, t \in \mathbb{Z}\}$, since in this case we have that

$$\sqrt{\frac{n}{M_n}}\widehat{\eta}_{4,\varepsilon} \xrightarrow{D} N\left(0,\ 8\int_{-1}^{1} w^2(x)dx \left(\sum_{h=-\infty}^{\infty} \rho_U^2(h)\right)^2\right),$$

see Theorem 2.3.1. Furthermore, for p large enough, we expect that the filtered time series $U_{t,p}$, $t = p+1, p+2, \ldots, n$, will behave as a time series of approximately uncorrelated random variables. That is, the limiting variance will further simplify to

$$8\int_{-1}^{1} w^2(x)dx.$$

These considerations suggest the following simple bootstrap procedure to perform a test of hypotheses (2.13).

Step 1: Calculate $\widehat{\gamma}_U(0)$ based on the filtered time series $\widehat{U}_{t,p}$, $t = p+1, p+2, \ldots, n$.

Step 2: Generate independent random variables $U^*_{p+1}, U^*_{p+2}, \ldots, U^*_n$ having a Gaussian distribution with mean zero and variance $\widehat{\gamma}_U(0)$.

Step 3: Using $U^*_{p+1}, U^*_{p+2}, \ldots, U^*_n$ calculate the estimator given in (2.7). Denote this estimator by $\widehat{\eta}^*_{4,\varepsilon}$.

Step 4: Repeat Step 2 and Step 3 a large number of times, say B times, and reject H_0 if

$$\widehat{\eta}_{4,\varepsilon} \leq q^*_{\alpha/2} \quad \text{or} \quad \widehat{\eta}_{4,\varepsilon} \geq q^*_{1-\alpha/2}.$$

Here and for $\beta \in (0,1)$, q^*_β denotes the β-percentage point of the distribution of $\widehat{\eta}^*_{4,\varepsilon}$, i.e., $P(\widehat{\eta}^*_{4,\varepsilon} \leq q^*_\beta) = \beta$. The percentage point q^*_β can be consistently estimated using the B bootstrap replications of the estimator $\widehat{\eta}^*_{4,\varepsilon}$.

Notice that in Step 2, the Gaussian distribution has been chosen for the sake of convenience only. The U_t^*'s could be also generated as i.i.d. random variables having a distribution with mean zero, variance $\widehat{\gamma}_U(0)$ and zero fourth order cumulant. Asymptotically, this will not affect the results obtained, because the limiting distribution of $\sqrt{n/M_n}\widehat{\eta}_{4,\varepsilon}$ depends, under the null hypothesis, only on the second order structure of the filtered process $\{U_{t,p}, t \in \mathbb{Z}\}$; see Theorem 2.3.1(iii). Now, it is easily seen that under the assumptions of this theorem,

$$t_n^* = \sqrt{\frac{n}{M_n}}\,\widehat{\eta}_{4,\varepsilon}^* \xrightarrow{D} N\left(0, 8\int_{-1}^{1} w^2(x)dx\right),$$

and this justifies asymptotically the use of the critical values $q_{\alpha/2}^*$ and $q_{1-\alpha/2}^*$ in Step 4 of the above bootstrap algorithm to perform the test.

2.6 Practical Considerations and Numerical Examples

2.6.1 Choice of Parameters

Implementation of the estimator $\widehat{\eta}_{4,\varepsilon}$ requires the choice of two parameters: The order p of the autoregressive filter used and the truncation lag M_n applied for the calculation of (2.7). In this section we give some guidelines of how to choose these parameters in practice.

Concerning the choice of the autoregressive order p, we recommend the use of Akaike's information criterion (AIC). That is, p should be selected as the minimizer of

$$AIC(p) = \text{argmin}_p \left\{\log \widehat{\gamma}_U(0) + \frac{2p}{n}\right\} \qquad (2.14)$$

over a range of values of p where $\widehat{\gamma}_U(0)$ is the estimated variance of the filtered process $\{U_{t,p}, t \in \mathbb{Z}\}$.

The difficult problem to solve concerns, certainly, the choice of the truncation lag M_n, which is common in both nonparametric estimation procedures for $\eta_{4,\varepsilon}$ considered in this chapter. Towards a suggestion for how to choose this parameter in practice, we give first the following alternative expression for the bias of $\widehat{\eta}_{4,\varepsilon}$,

$$\begin{aligned} E(\widehat{\eta}_{4,\varepsilon}) - \eta_{4,\varepsilon} &= \frac{1}{M_n^2}\int_{-1}^{1} x^2 w(x)dx \left\{2\sum_{-\infty}^{\infty} h^2\rho_U^2(h) - \frac{\gamma_{2,U}(0)}{\gamma_U^2(0)}\sum_{h=-\infty}^{\infty} h^2\rho_{2,U}(h)\right\} + o\left(\frac{1}{M_n^2}\right) \\ &= \frac{1}{M_n^2}\int_{-1}^{1} x^2 w(x)dx \frac{1}{c_U(0)}\sum_{h=-\infty}^{\infty} h^2\left(2c_U(0)\rho_U^2(h) - \rho_{2,U}(h)\right) + o\left(\frac{1}{M_n^2}\right) \end{aligned}$$

where,

$$\rho_U(h) = \gamma_U(h)/\gamma_U(0) \quad \rho_{2,U}(h) = \gamma_{2,U}(h)/\gamma_{2,U}(0), \quad \text{and} \quad c_U(0) = \gamma_U^2(0)/\gamma_{2,U}(0).$$

Using assertion (ii) of Theorem 2.3.1 and equation (2.3) an alternative expression for the variance of $\widehat{\eta}_{4,\varepsilon}$ is given by

$$Var(\widehat{\eta}_{4,\varepsilon}) = 2\int_{-1}^{1} w^2(x)dx \frac{1}{c_U^2(0)} \frac{M_n}{n} \left(\sum_{h=-\infty}^{\infty} \rho_{2,U}(h) \right)^2 + o\left(\frac{M_n}{n}\right).$$

Thus, it follows by straightforward calculations that the value of M_n which minimizes the (asymptotic) mean square error

$$E(\widehat{\eta}_{4,\varepsilon} - \eta_{4,\varepsilon})^2 = Var(\widehat{\eta}_{4,\varepsilon}) + (E(\widehat{\eta}_{4,\varepsilon}) - \eta_{4,\varepsilon})^2,$$

is given by

$$M_n^{(opt)} = \left\{ \frac{2W_1^2 \left(\sum_{h=-\infty}^{\infty} h^2(2c_U(0)\rho_U^2(h) - \rho_{2,U}(h)) \right)^2}{W_2 \left(\sum_{h=-\infty}^{\infty} \rho_{2,U}(h) \right)^2} \right\}^{1/5} n^{1/5}. \tag{2.15}$$

Here,

$$W_1 = \int_{-1}^{1} x^2 w(x)dx \quad \text{and} \quad W_2 = \int_{-1}^{1} w^2(x)dx.$$

From (2.15), a crude estimator $\widehat{M_n^{(opt)}}$ of $M_n^{(opt)}$ can be obtained by replacing $c_U(0)$, $\rho_{2,U}(h)$ and $\rho_U(h)$ by sample estimators and truncating the infinite sums to some finite, small value K. As a simple practical rule, we use $K = 1$ in all calculations presented in this section. This choice can be also justified by the fact that, since $\mathbf{U} = \{U_{t,p}, t \in \mathbb{Z}\}$ is a filtered process, we expect that for p large enough, many of its autocorrelations $\rho_U(h)$ will be close to zero.

2.6.2 Numerical Simulations

We first investigated numerically the finite sample performance of the two nonparametric estimators $\breve{\eta}_{4,\varepsilon}$ and $\widehat{\eta}_{4,\varepsilon}$. For this, we generated time series of length $n = 100$ and $n = 500$ of the ARMA(1,1) model

$$X_t = \phi X_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t, \tag{2.16}$$

with different choices of the parameters ϕ and θ and different distributions for the i.i.d. innovations ε_t. In particular and concerning the choice of the pair of parameters (ϕ, θ), the following seven models have been considered:

Model I: $(0, 0.8)$

Model II: $(0, -0.8)$

Model III: $(0, 1.25)$

Model IV: $(0, -1.25)$

Model V: $(0.8, 0)$

Model VI: $(-0.8, 0)$

Model VII: $(0.8, 0.8)$

Notice that Model III and Model IV are non-invertible moving average processes with parameters that are the reciprocal values of the parameters of the invertible counterparts given in Model I and Model II respectively. Furthermore and concerning the innovations, the following four distributions with mean zero and unit variance have been selected:

(I) Uniform on $\left[-\sqrt{3}, \sqrt{3}\right]$, $(\eta_{4,\varepsilon} = -1.2)$

(II) Standard Gaussian, $(\eta_{4,\varepsilon} = 0)$

(III) Logistic, $(\eta_{4,\varepsilon} = 1.2)$

(IV) Double exponential, $(\eta_{4,\varepsilon} = 3.0)$

In all calculations the Bartlett-Priestly lag window has been used.

For the calculation of the estimator $\breve{\eta}_{4,\varepsilon}$ we have used in the simulations the theoretically derived optimal value $M_n^{(opt)}$ of the truncation lag given in formulae (2.15), where the unknown quantities appearing in this formulae have been evaluated using the true parameters of the underlying model. In contrast to this, for the new estimator $\widehat{\eta}_{4,\varepsilon}$, we selected the parameters p and M_n as suggested in Subsection 2.6.1, where the estimator $\widehat{M}_n^{(opt)}$ proposed there has been used. In other words, we compare the performance of the new estimator $\widehat{\eta}_{4,\varepsilon}$ based on data driven choices of p and M_n with the performance of the estimator $\breve{\eta}_{4,\varepsilon}$ based on the theoretically optimal choice of M_n.

Figure 2.1 and Figure 2.2 present boxplots of both estimators obtained over $R = 100$ replications, of each one of the seven models and of each one of the four distributions of the innovations considered. The corresponding mean square errors of both estimators are presented in Figure 2.3 and Figure 2.4 respectively. Furthermore, Table 2.1

and Table 2.2 present the Mean (Mean) and the standard deviation (Std) of the two estimators considered.

As it is seen, the new estimator $\widehat{\eta}_{4,\varepsilon}$ performs extremely well and leads to impressive improvements especially in the case of the more correlated time series (models V,VI and VII). This can be verified by examining the behavior of the boxplots over the different models and the different distributions of the innovations considered as well as the behavior of the corresponding mean square errors. In fact, it seems that pre-whitening the time series stabilizes the mean square error of estimation over the different autocorrelation structures considered towards the case of a time series consisting of non-correlated observations.

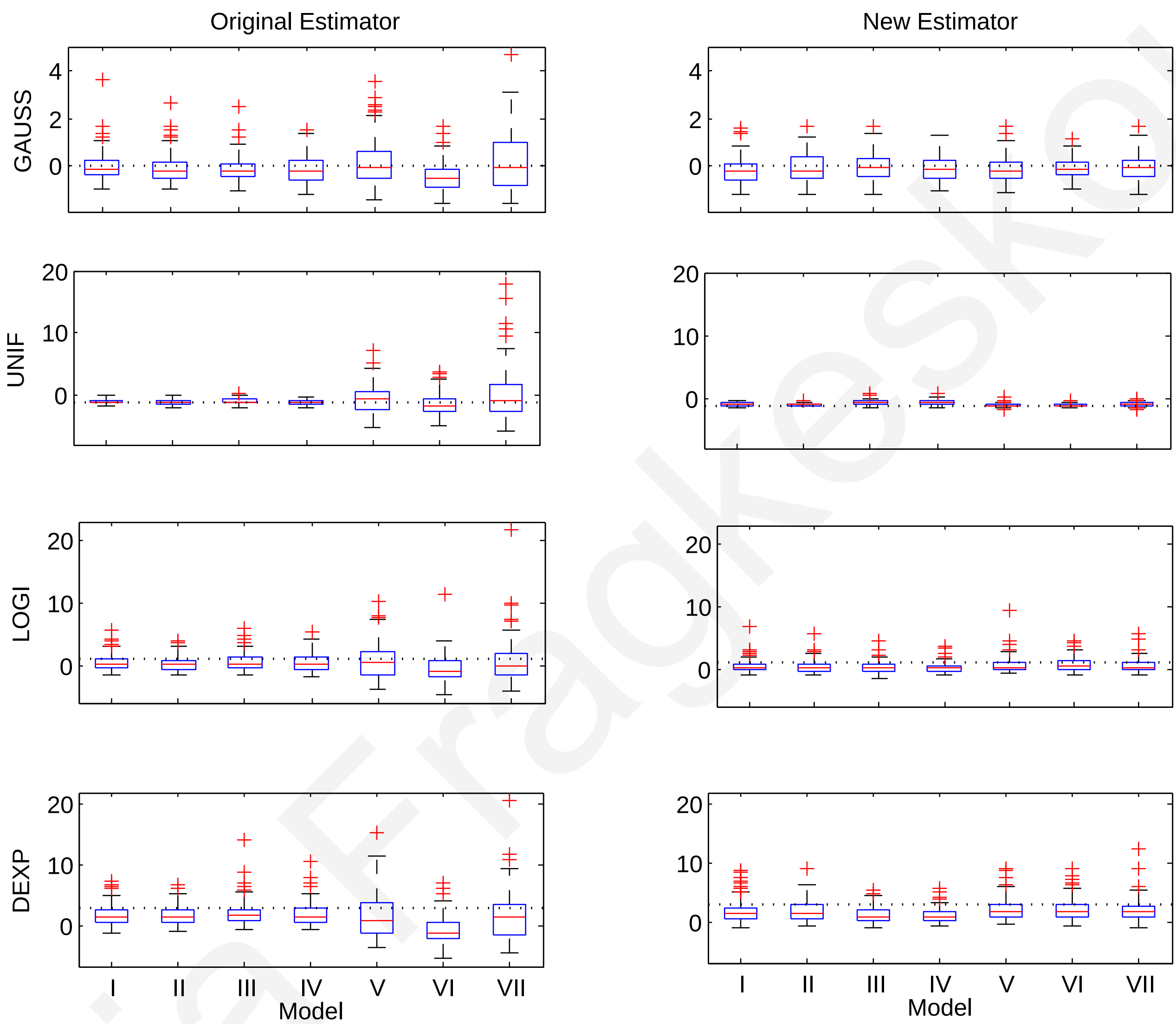

Figure 2.1: Boxplots of the distributions of the estimators $\breve{\eta}_{4,\varepsilon}$ (left panel) and $\widehat{\eta}_{4,\varepsilon}$ (right panel) for the different models, the different innovation distributions considered and the sample size of $n = 100$ observations.

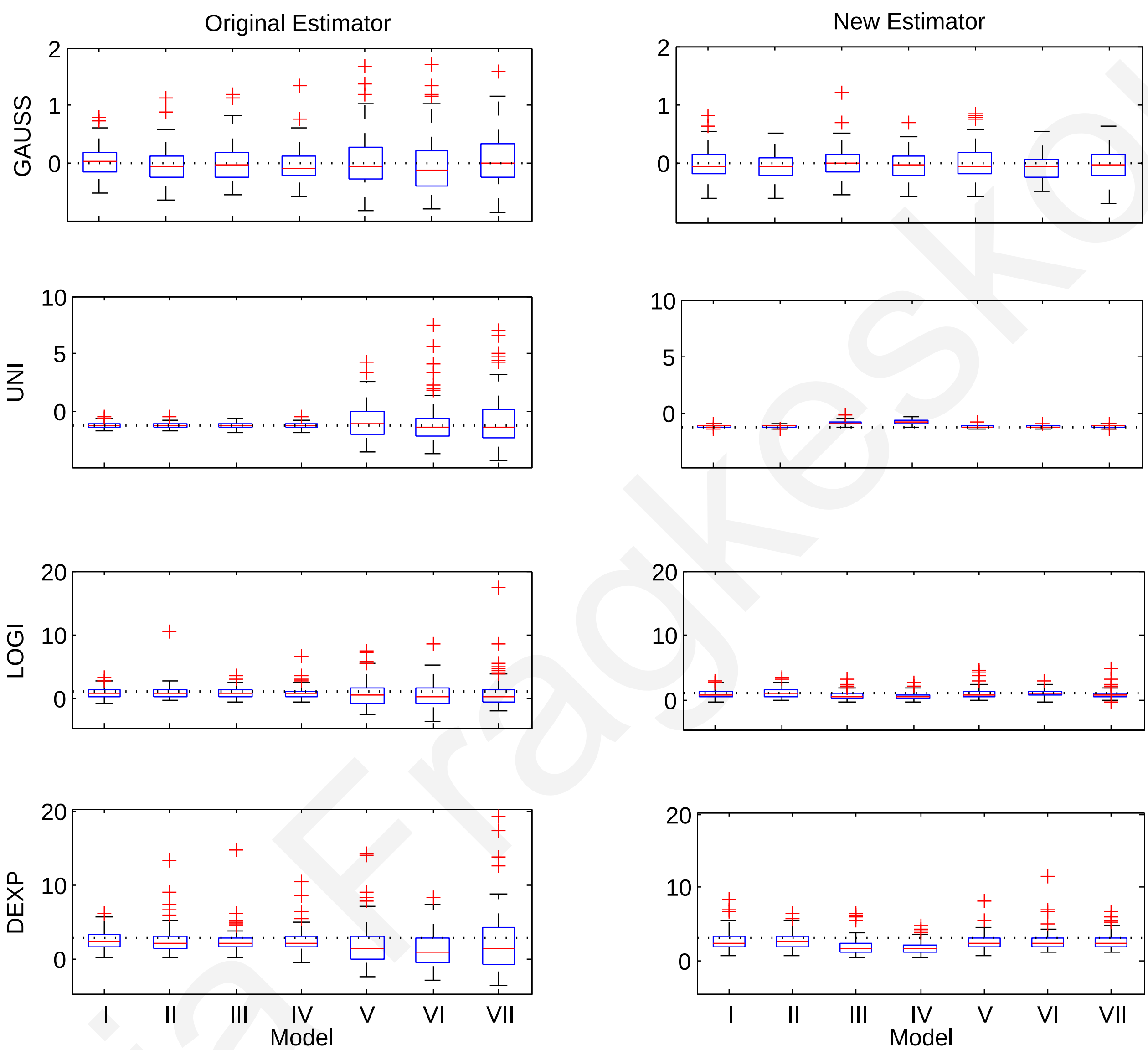

Figure 2.2: Boxplots of the distributions of the estimators $\breve{\eta}_{4,\varepsilon}$ (left panel) and $\widehat{\eta}_{4,\varepsilon}$ (right panel) for the different models, the different innovation distributions considered and the sample size of $n = 500$ observations.

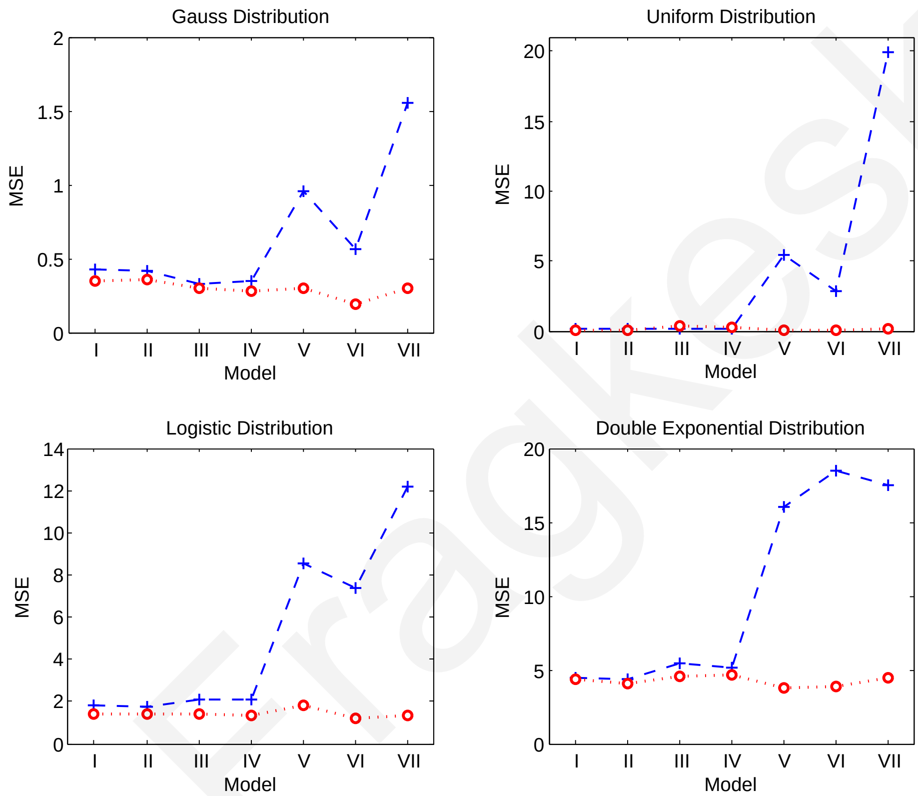

Figure 2.3: Mean square error (MSE) of the two estimators for the different models, the different innovation distributions and the sample size of n = 100 observations. $+ - - - +$ refers to the MSE of the original estimator $\breve{\eta}_{4,\varepsilon}$, while $\circ - - - \circ$ to that of the new estimator $\widehat{\eta}_{4,\varepsilon}$.

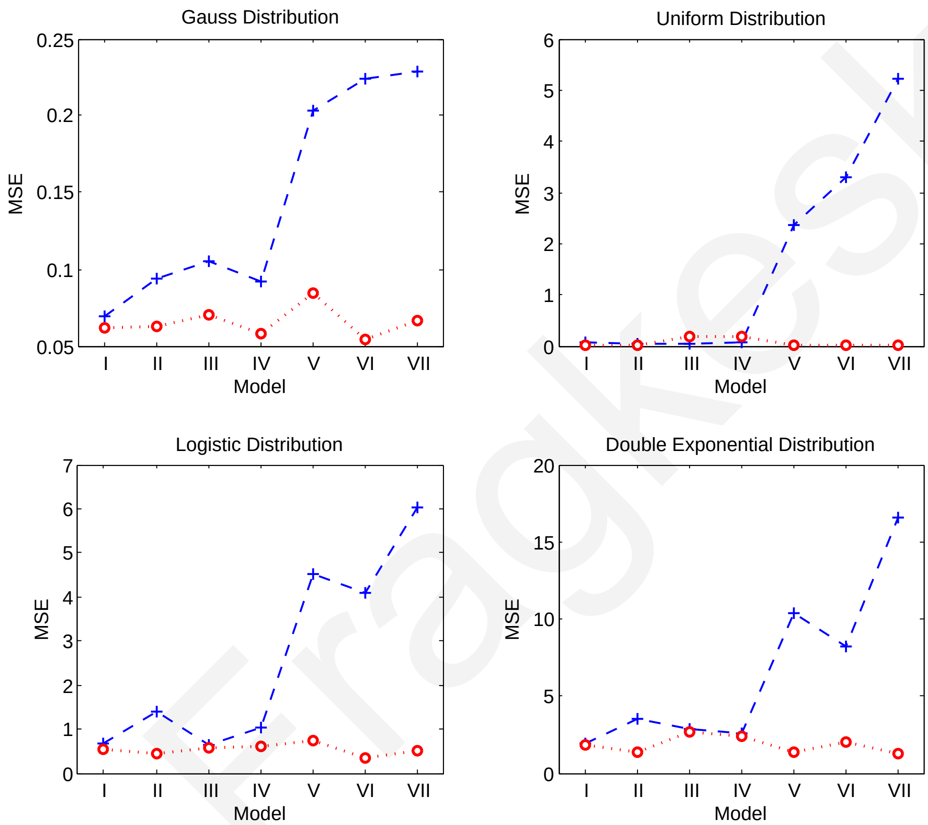

Figure 2.4: Mean square error (MSE) of the two estimators for the different models, the different innovation distributions and the sample size of n = 500 observations. $+ - - - +$ refers to the MSE of the original estimator $\breve{\eta}_{4,\varepsilon}$, while $\circ - - - \circ$ to that of the new estimator $\widehat{\eta}_{4,\varepsilon}$.

θ	ϕ	Est.	UNIF		NORM		LOGI		DEXP	
			Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
0.8	0.0	$\eta_{4,\varepsilon}$	-1.200		0.000		1.200		3.000	
		$\breve{\eta}_{4,\varepsilon}$	-1.053	0.384	-0.002	0.658	0.742	1.281	1.736	1.706
		$\widehat{\eta}_{4,\varepsilon}$	-0.994	0.239	-0.151	0.570	0.616	1.043	2.054	1.877
-0.8	0.0	$\breve{\eta}_{4,\varepsilon}$	-1.113	0.386	-0.060	0.650	0.570	1.163	1.597	1.553
		$\widehat{\eta}_{4,\varepsilon}$	-0.996	0.191	-0.068	0.598	0.614	1.034	1.965	1.735
1.25	0.0	$\breve{\eta}_{4,\varepsilon}$	-1.036	0.416	-0.124	0.559	0.769	1.399	2.080	2.156
		$\widehat{\eta}_{4,\varepsilon}$	-0.756	0.345	-0.013	0.551	0.449	0.937	1.325	1.316
-1.25	0.0	$\breve{\eta}_{4,\varepsilon}$	-1.152	0.387	-0.134	0.575	0.689	1.371	1.906	1.988
		$\widehat{\eta}_{4,\varepsilon}$	-0.764	0.342	-0.124	0.520	0.477	0.908	1.200	1.206
0.0	0.8	$\breve{\eta}_{4,\varepsilon}$	-0.641	2.268	0.175	0.964	1.119	2.942	1.596	3.763
		$\widehat{\eta}_{4,\varepsilon}$	-1.167	0.238	-0.097	0.541	0.887	1.314	2.358	1.847
0.0	-0.8	$\breve{\eta}_{4,\varepsilon}$	-1.475	1.684	-0.391	0.647	-0.283	2.284	-0.573	2.407
		$\widehat{\eta}_{4,\varepsilon}$	-1.183	0.196	-0.088	0.435	1.009	1.093	2.390	1.883
0.8	0.8	$\breve{\eta}_{4,\varepsilon}$	0.108	4.292	0.155	1.243	0.896	3.495	1.644	3.979
		$\widehat{\eta}_{4,\varepsilon}$	-1.006	0.318	-0.030	0.549	0.693	1.053	2.098	1.907

Table 2.1: Mean (Mean) and standard deviation (Std) of the two estimators over $R = 100$ replicates for different models and $n = 100$ observations. NORM, UNIF, LOGI and DEXP refer to the normal, the uniform, the logistic and the double exponential distribution of the innovations ε_t respectively with mean zero and unit variance.

$\boldsymbol{\theta}$	$\boldsymbol{\phi}$	**Est.**	**UNIF**		**NORM**		**LOGI**		**DEXP**	
			Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
0.8	0.0	$\eta_{4,\varepsilon}$	-1.200		0.000		1.200		3.000	
		$\breve{\eta}_{4,\varepsilon}$	-1.128	0.239	0.031	0.264	0.920	0.771	2.649	1.355
		$\widehat{\eta}_{4,\varepsilon}$	-1.130	0.079	-0.005	0.250	1.083	0.731	2.705	1.336
-0.8	0.0	$\breve{\eta}_{4,\varepsilon}$	-1.144	0.224	-0.028	0.307	0.963	1.166	2.584	1.840
		$\widehat{\eta}_{4,\varepsilon}$	-1.144	0.101	-0.064	0.243	1.204	0.671	2.701	1.144
1.25	0.0	$\breve{\eta}_{4,\varepsilon}$	-1.156	0.233	0.001	0.326	0.992	0.779	2.531	1.630
		$\widehat{\eta}_{4,\varepsilon}$	-0.833	0.189	0.012	0.267	0.734	0.611	1.862	1.178
-1.25	0.0	$\breve{\eta}_{4,\varepsilon}$	-1.130	0.2292	-0.036	0.302	0.943	0.993	2.526	1.543
		$\widehat{\eta}_{4,\varepsilon}$	-0.793	0.179	-0.042	0.238	0.668	0.573	1.760	0.924
0.0	0.8	$\breve{\eta}_{4,\varepsilon}$	-0.801	1.497	0.023	0.452	0.716	2.079	2.115	3.113
		$\widehat{\eta}_{4,\varepsilon}$	-1.197	0.074	0.017	0.292	1.139	0.860	2.546	1.073
0.0	-0.8	$\breve{\eta}_{4,\varepsilon}$	-0.988	1.809	-0.054	0.473	0.544	1.920	1.399	2.392
		$\widehat{\eta}_{4,\varepsilon}$	-1.195	0.072	-0.063	0.226	1.112	0.587	2.655	1.376
0.8	0.8	$\breve{\eta}_{4,\varepsilon}$	-0.715	2.247	0.075	0.475	0.900	2.447	2.236	4.023
		$\widehat{\eta}_{4,\varepsilon}$	-1.125	0.090	-0.006	0.259	0.973	0.687	2.544	1.032

Table 2.2: Mean (Mean) and standard deviation (Std) of the two estimators over $R = 100$ replicates for different models and $n = 500$ observations. NORM, UNIF, LOGI and DEXP refer to the normal, the uniform, the logistic and the double exponential distribution of the innovations ε_t respectively with mean zero and unit variance.

We next consider the problem of testing the hypothesis

$$H_0 : \eta_{4,\varepsilon} = 0 \quad \text{vs.} \quad H_1 : \eta_{4,\varepsilon} \neq 0.$$

For this, the size and power behavior of the bootstrap-based testing procedure proposed in Subsection 2.5.2 is investigated. Time series from model (2.16) have been considered with different choices of the parameters (ϕ, θ) and different i.i.d. innovations ε_t. The innovations have been generated as

$$\varepsilon_t = (1-\gamma)z_t + \gamma w_t, \qquad \gamma \in [0,1], \tag{2.17}$$

where $\{z_t, t \in \mathbb{Z}\}$ and $\{w_t, t \in \mathbb{Z}\}$ are i.i.d. random sequences, independent from each other and such that z_t has a standard Gaussian distribution and w_t has one of the following three distributions: uniform, logistic and double exponential. The parameters of the distribution of w_t have been chosen so that ε_t has mean zero and unit variance. Notice that specification (2.17) of the innovations allows for the examination of the size and of the power properties of the bootstrap-based test.

In particular, the case $\gamma = 0$ corresponds to the null hypothesis while the case $\gamma > 0$ to the alternative. Table 2.3 and Table 2.4 present the empirical rejection probabilities calculated over $R = 1000$ replications for sample sizes of $n = 100$ and $n = 500$ observations respectively. The test statistic described in Subsection 2.5.2 has been used, with the smoothing parameters p and M_n selected according to the recommendations made in Subsection 2.6.1. The critical values of the test have been estimated using $B = 1000$ bootstrap replications.

As Table 2.3 and Table 2.4 show, the test retains the correct size over the different model structures considered. It also has reasonable power for deviations from the null even for the sample size of $n = 100$ observations. Furthermore, the power of the test increases as the sample size increases and/or as the deviation from the null becomes larger, that is as the parameter γ becomes larger.

θ	ϕ	α	GAUSS	UNIF		LOGI		DEXP	
			$\gamma = 0.0$	$\gamma = 0.5$	$\gamma = 1.0$	$\gamma = 0.5$	$\gamma = 1.0$	$\gamma = 0.5$	$\gamma = 1.0$
0.0	0.0	5%	0.046	0.233	0.881	0.169	0.283	0.402	0.670
		10%	0.087	0.338	0.924	0.262	0.373	0.510	0.774
0.8	0.0	5%	0.055	0.193	0.666	0.156	0.264	0.352	0.589
		10%	0.103	0.277	0.778	0.238	0.371	0.439	0.694
-0.8	0.0	5%	0.057	0.195	0.712	0.161	0.254	0.307	0.569
		10%	0.116	0.290	0.820	0.249	0.354	0.399	0.677
1.25	0.0	5%	0.052	0.116	0.288	0.103	0.174	0.233	0.438
		10%	0.096	0.184	0.388	0.172	0.268	0.325	0.535
-1.25	0.0	5%	0.050	0.117	0.306	0.102	0.158	0.228	0.368
		10%	0.092	0.187	0.408	0.175	0.237	0.305	0.485
0.0	0.8	5%	0.054	0.226	0.836	0.179	0.274	0.416	0.706
		10%	0.110	0.326	0.900	0.270	0.388	0.535	0.800
0.0	-0.8	5%	0.049	0.222	0.877	0.164	0.281	0.395	0.657
		10%	0.105	0.315	0.935	0.266	0.387	0.505	0.764
0.8	0.8	5%	0.078	0.181	0.556	0.193	0.289	0.355	0.590
		10%	0.134	0.260	0.701	0.262	0.393	0.465	0.692

Table 2.3: Empirical rejection probabilities over $R = 1000$ replications of the bootstrap-based testing procedure for different models and innovation structures and for sample size of n = 100 observations. GAUSS, UNIF, LOGI and DEXP refer to the Gaussian, the uniform, the logistic and the double exponential distribution used in the equation (2.17) to specify the innovations.

θ	ϕ	α	GAUSS	UNIF		LOGI		DEXP	
			$\gamma = 0.0$	$\gamma = 0.5$	$\gamma = 1.0$	$\gamma = 0.5$	$\gamma = 1.0$	$\gamma = 0.5$	$\gamma = 1.0$
0.0	0.0	5%	0.050	0.920	1.000	0.528	0.808	0.952	1.000
		10%	0.102	0.960	1.000	0.626	0.860	0.974	1.000
0.8	0.0	5%	0.050	0.880	1.000	0.444	0.780	0.904	1.000
		10%	0.100	0.930	1.000	0.552	0.850	0.948	1.000
-0.8	0.0	5%	0.040	0.892	1.000	0.434	0.782	0.916	1.000
		10%	0.104	0.946	1.000	0.554	0.866	0.958	1.000
1.25	0.0	5%	0.046	0.458	0.972	0.254	0.480	0.700	0.940
		10%	0.100	0.588	0.984	0.346	0.590	0.798	0.978
-1.25	0.0	5%	0.050	0.420	0.974	0.252	0.526	0.716	0.962
		10%	0.092	0.570	0.982	0.356	0.614	0.808	0.970
0.8	0.0	5%	0.052	0.934	1.000	0.514	0.802	0.926	0.998
		10%	0.110	0.970	1.000	0.628	0.888	0.956	0.998
-0.8	0.0	5%	0.050	0.932	1.000	0.482	0.796	0.930	0.998
		10%	0.104	0.968	1.000	0.612	0.888	0.960	0.998
0.8	0.8	5%	0.068	0.886	1.000	0.450	0.760	0.924	0.992
		10%	0.118	0.948	1.000	0.580	0.842	0.962	0.996

Table 2.4: Empirical rejection probabilities over $R = 1000$ replications of the bootstrap-based testing procedure for different models and innovation structures and for sample size of n = 500 observations. GAUSS, UNIF, LOGI and DEXP refer to the Gaussian, the uniform, the logistic and the double exponential distribution used in the equation (2.17) to specify the innovations.

2.6.3 Real-life Data Applications

We applied the testing procedure for the hypothesis

$$H_0 : \eta_{4,\varepsilon} = 0$$

proposed in this chapter to eleven different datasets commonly used in the time series literature and to which linear models have been fitted after transforming them to stationarity by taking first differences and/or logarithms. For every time series considered, the test has been performed with smoothing parameters p and M_n selected according to the suggestions made in Subsection 2.6.1. Table 2.5 presents the different time series considered together with the source from which they have been obtained, the sample size n, the estimated value $\widehat{\eta}_{4,\varepsilon}$ and the p-value of the test based on $B = 1000$ bootstrap replications.

The results presented in this table are quite interesting. In particular and except from only two out of the eleven time series considered, the null hypothesis $\eta_{4,\varepsilon} = 0$ has been rejected at the commonly used levels. These results, do not only imply that for many time series in practice, the hypothesis of Gaussianity can not be justified, but also that statistical inference based on the simplifying assumption that $\eta_{4,\varepsilon} = 0$, may lead to erroneous conclusions.

	Time Series	n	$\widehat{\eta}_{4,\varepsilon}$	Bootstrap p-value
1	Series A, (first difference) Box and Jenkins (1970), p.525	196	1.2818	0.0075
2	Series C, (first difference) Box and Jenkins (1970), p.528	225	13.128	0.0000
3	Series D, (first difference) Box and Jenkins (1970), p.529	309	2.7310	0.0000
4	Series J, (first difference) Box and Jenkins (1970), p.532-533	295	10.886	0.0000
5	German Egg Prices, (first difference) Fan and Yao (2003), p.113	299	1.2714	0.0024
6	GNP Data, (first difference) Shumway and Stoffer (2006), p.144	222	2.2919	0.0002
7	Recruitment Series, (first difference) Shumway and Stoffer (2006), p.109	452	2.8355	0.0000
8	Southern Oscillation Index, (first difference) Shumway and Stoffer (2006), p.222	452	0.04956	0.8836
9	Federal Reserve Board Production, (first difference) Shumway and Stoffer (2006), p.160	371	5.1467	0.0000
10	Global Temperature, (first difference) Shumway and Stoffer (2006), p.58	97	-0.6983	0.1993
11	Paleoclimatic Glacial Varves, (first difference of log-transform) Shumway and Stoffer (2006), p.62	633	0.5131	0.0303

Table 2.5: Bootstrap p-values of the test of the hypothesis $\eta_{4,\varepsilon} = 0$ for different real-life data sets transformed to stationarity by applying first differences (first diff.) and/or taking logarithms (log-trans.)

2.7 Auxiliary Lemmas and Proofs

In this section we give the proofs of the main theorems presented in Chapter 2 suppressing at some places cumbersome but straightforward calculations.

Throughout the proofs we use the function $n\phi_n(r;h_1,h_2)$ which is defined by

$$n\phi_n(r;h_1,h_2)=\begin{cases} n-\frac{1}{2}\{|h_2|+|h_1|+(h_1-h_2)\}+r, \\ \quad \text{if } r=-\Big\{n-\frac{1}{2}(|h_2|+|h_1|+(h_1-h_2))\Big\},...,\frac{1}{2}(h_1-h_2)-\frac{1}{2}||h_1|-|h_2||, \\ n-\max\{|h_1|,|h_2|\}, \\ \quad \text{if } \ r=\frac{1}{2}(h_1-h_2)-\frac{1}{2}||h_1|-|h_2||,...,\frac{1}{2}(h_1-h_2)+\frac{1}{2}||h_1|-|h_2||, \\ n-\frac{1}{2}\{|h_2|+|h_1|-(h_1-h_2)\}-r, \\ \quad \text{if } \ r=\frac{1}{2}(h_1-h_2)+\frac{1}{2}||h_1|-|h_2||,...,n-\frac{1}{2}(|h_2|+|h_1|-(h_1-h_2)), \\ 0, \\ \quad \text{elsewhere.} \end{cases}$$

Note that for every r, h_1, h_2,

$$0\le\phi_n(r;h_1,h_2)\le 1,\quad \text{and} \quad \lim_{n\to\infty}\phi_n(r;h_1,h_2)=1;$$

see also Anderson (1971), Problem 19, p. 555. Furthermore, denote by

$$Cum\left(X_t,X_{t+h_1},X_{t+h_2},...,X_{t+h_r}\right)\equiv cum_X\left(h_1,h_2,...,h_r\right),$$

the $(r+1)^{th}$ order joint cumulant of $(X_t,X_{t+h_1},X_{t+h_2},...,X_{t+h_r})$, and by

$$Cum\left(X_t^2,X_{t+h_1}^2,X_{t+h_2}^2X_{t+h_3}^2\right)\equiv cum_{X^2}\left(h_1,h_2,h_3\right),$$

the 4^{th} order joint cumulant of $\left(X_t^2,X_{t+h_1}^2,X_{t+h_2}^2,X_{t+h_3}^2\right)$.

The first lemma deals with the absolute summability of certain cumulant functions for linear processes.

Lemma 2.7.1. *Suppose that Assumption 2.1 is satisfied. Then,*

(i) $\sum_{h_1,h_2,...,h_7=-\infty}^{\infty}\left|Cum\left(X_t,X_{t+h_1},...,X_{t+h_7}\right)\right|<\infty,$

(ii) $\sum_{h_1,h_2,h_3=-\infty}^{\infty}\left|Cum\left(X_t^2,X_{t+h_1}^2,X_{t+h_2}^2,X_{t+h_3}^2\right)\right|<\infty.$

Proof:

(i) Following Rosenblatt (1985) the cumulant $Cum\left(X_t, X_{t+h_1}, ..., X_{t+h_7}\right)$ is given by

$$Cum\left(X_t, X_{t+h_1}, ..., X_{t+h_7}\right) = \gamma_8 \sum_{r=-\infty}^{\infty} \psi_r \psi_{r+h_1} ... \psi_{r+h_7}$$

where γ_8 is the 8^{th} order cumulant of ε_t. Then,

$$\begin{aligned}\sum_{h_1,h_2,...,h_7=-\infty}^{\infty} \left|Cum\left(X_t, X_{t+h_1}, ..., X_{t+h_7}\right)\right| &= |\gamma_8| \sum_{h_1,h_2,...,h_7=-\infty}^{\infty} \left|\sum_{r=-\infty}^{\infty} \psi_r \psi_{r+h_1} ... \psi_{r+h_7}\right| \\ &\leq |\gamma_8| \left(\sum_{r=-\infty}^{\infty} |\psi_r|\right)^8 < \infty.\end{aligned}$$

(ii) It is known that for random variables Z_1, Z_2, Z_3, Z_4 the cumulant $Cum\left(Z_1, Z_2, Z_3, Z_4\right)$ is given by

$$\begin{aligned}Cum\left(Z_1, Z_2, Z_3, Z_4\right) =& E\left(Z_1 Z_2 Z_3 Z_4\right) - \Big\{E\left(Z_1\right) E\left(Z_2 Z_3 Z_4\right) + ... + E\left(Z_4\right) E\left(Z_1 Z_2 Z_3\right)\Big\} \\ & - \Big\{E\left(Z_1 Z_2\right) E\left(Z_3 Z_4\right) + E\left(Z_1 Z_3\right) E\left(Z_2 Z_4\right) + E\left(Z_1 Z_4\right) E\left(Z_3 Z_2\right)\Big\} \\ & + 2\Big\{E\left(Z_1\right) E\left(Z_2\right) E\left(Z_3 Z_4\right) + ... + E\left(Z_3\right) E\left(Z_4\right) E\left(Z_1 Z_2\right)\Big\} \\ & - 6E\left(Z_1\right) E\left(Z_2\right) E\left(Z_3\right) E\left(Z_4\right). \qquad (2.18)\end{aligned}$$

Thus, $Cum\left(X_t^2, X_{t+h_1}^2, X_{t+h_2}^2, X_{t+h_3}^2\right)$ equals

$$\sum_{j_1,j_2,...,j_8=-\infty}^{\infty} \psi_{j_1} \psi_{j_2} ... \psi_{j_8} Cum\left(\varepsilon_{t-j_1}\ \varepsilon_{t-j_2}, \varepsilon_{t+h_1-j_3} \varepsilon_{t+h_1-j_4}, \varepsilon_{t+h_2-j_5} \varepsilon_{t+h_2-j_6}, \varepsilon_{t+h_3-j_7} \varepsilon_{t+h_3-j_8}\right).$$

Evaluating the above cumulant of the ε_t's, we get after straightforward calculations that

$$\begin{aligned}&\sum_{h_1,h_2,h_3=-\infty}^{\infty} \left|Cum\left(X_t^2, X_{t+h_1}^2, X_{t+h_2}^2, X_{t+h_3}^2\right)\right| \\ &\leq \Big\{\left|E\left(\varepsilon_1^8\right)\right| + 28E\left(\varepsilon_1^2\right)\left|E\left(\varepsilon_1^6\right)\right| + 35E^2\left(\varepsilon_1^4\right) + 642E^4\left(\varepsilon_1^2\right) + 420\left|E\left(\varepsilon_1^4\right)\right| E^2\left(\varepsilon_1^2\right)\Big\} \\ &\quad \times \left(\sum_{j=-\infty}^{\infty} \psi_j^2\right)^4 + \Big\{456E^4\left(\varepsilon_1^2\right) + 96\left|E\left(\varepsilon_1^4\right)\right| E^2\left(\varepsilon_1^2\right)\Big\}\left(\sum_{j=-\infty}^{\infty} |\psi_j|\right)^6 \left(\sum_{j=-\infty}^{\infty} \psi_j^2\right) \\ &\quad + 24\Big\{E\left(\varepsilon_1^2\right)\left|E\left(\varepsilon_1^6\right)\right| + 24E^2\left(\varepsilon_1^4\right) + 1092E^4\left(\varepsilon_1^2\right) + 504\left|E\left(\varepsilon_1^4\right)\right| E^2\left(\varepsilon_1^2\right)\Big\} \\ &\quad \times \left(\sum_{j=-\infty}^{\infty} \psi_j^2\right)^2 \left(\sum_{j=-\infty}^{\infty} |\psi_j|\right)^4 + \Big\{56E^2\left(\varepsilon_1^4\right) + 96\left|E\left(\varepsilon_1^4\right)\right| E^2\left(\varepsilon_1^2\right)\Big\}\left(\sum_{j=-\infty}^{\infty} |\psi_j|\right)^8 \\ &\leq C\left(\sum_{r=-\infty}^{\infty} |\psi_r|\right)^8,\end{aligned}$$

which yields the expected result. □

Let $\widetilde{\eta}_{4,\varepsilon}$ be the same estimator as $\breve{\eta}_{4,\varepsilon}$ with $\overline{X}_n$ replaced by $\mu = E(X_t)$ and $\overline{X}_{2,n}$ by $\mu_2 = E\left(X_t^2\right)$. In general, and in the following, "$\sim$" refers to estimators using the true means μ and μ_2, while, "$\frown$" to estimators using the sample means $\overline{X}_n$ and $\overline{X}_{2,n}$. The following lemma shows that the asymptotic properties of the estimator $\breve{\eta}_{4,\varepsilon}$ are the same as those of the estimator $\widetilde{\eta}_{4,\varepsilon}$.

Lemma 2.7.2. *Suppose that Assumption 2.1 and Assumption 2.2 are satisfied. Then, as* $n \to \infty$

(i) $E\left(\breve{\eta}_{4,\varepsilon}\right) = E\left(\widetilde{\eta}_{4,\varepsilon}\right) + O\left(\frac{1}{\sqrt{n}} + \frac{M_n}{n}\right)$,

(ii) $\frac{n}{M_n} Var\left(\breve{\eta}_{4,\varepsilon}\right) = \frac{n}{M_n} Var\left(\widetilde{\eta}_{4,\varepsilon}\right) + O\left(\frac{1}{\sqrt{M_n}} + \frac{M_n}{n}\right)$,

(iii) $\sqrt{\frac{n}{M_n}}\left(\breve{\eta}_{4,\varepsilon} - E\left(\breve{\eta}_{4,\varepsilon}\right)\right) = \sqrt{\frac{n}{M_n}}\left(\widetilde{\eta}_{4,\varepsilon} - E\left(\widetilde{\eta}_{4,\varepsilon}\right)\right) + O_P\left(\frac{1}{\sqrt{M_n}} + \sqrt{\frac{M_n}{n}}\right)$.

Proof:

Let,

$$\breve{\eta}_{4,\varepsilon} = \frac{\widehat{N}_n}{\widehat{\gamma}_X^2(0)} \quad \text{and} \quad \widetilde{\eta}_{4,\varepsilon} = \frac{\widetilde{N}_n}{\widetilde{\gamma}_X^2(0)},$$

where

$$\widehat{N}_n = \sum_{h=-(n-1)}^{(n-1)} w\left(\frac{h}{M_n}\right)\left(\widehat{\gamma}_{2,X}(h) - 2\widehat{\gamma}_X^2(h)\right) \text{ and } \widetilde{N}_n = \sum_{h=-(n-1)}^{(n-1)} w\left(\frac{h}{M_n}\right)\left(\widetilde{\gamma}_{2,X}(h) - 2\widetilde{\gamma}_X^2(h)\right).$$

To prove assertion (i) of the lemma observe that for $|h| \leq n-1$,

$$\widehat{\gamma}_{2,X}(h) = \widetilde{\gamma}_{2,X}(h) - \frac{1}{n^2}\sum_{t=1}^{n-|h|}\sum_{s=1}^{n}\left(X_t^2 - \mu_2\right)\left(X_s^2 - \mu_2\right) - \frac{1}{n^2}\sum_{t=1}^{n-|h|}\sum_{s=1}^{n}\left(X_{t+|h|}^2 - \mu_2\right)\left(X_s^2 - \mu_2\right)$$
$$+ \frac{n-|h|}{n^3}\sum_{t,s=1}^{n}\left(X_t^2 - \mu_2\right)\left(X_s^2 - \mu_2\right).$$

Taking expectation on both sides yields

$$E\left(\widehat{\gamma}_{2,X}(h)\right) = E\left(\widetilde{\gamma}_{2,X}(h)\right) - \frac{1}{n^2}\sum_{t=1}^{n-|h|}\sum_{s=1}^{n}\gamma_{2,X}(s-t) - \frac{1}{n^2}\sum_{t=1}^{n-|h|}\sum_{s=1}^{n}\gamma_{2,X}(s-t-|h|)$$
$$+ \frac{n-|h|}{n^3}\sum_{t,s=1}^{n}\gamma_{2,X}(s-t). \tag{2.19}$$

The second term on the right-hand side of (2.19) equals

$$\frac{1}{n^2}\sum_{b=-(n-1)}^{(n-1)} n\phi_n(b;h,0)\,\gamma_{2,X}(b),$$

which is in turn, bounded in absolute value by

$$\frac{1}{n}\sum_{b=-\infty}^{\infty}\left|\gamma_{2,X}\left(b\right)\right| = O\left(\frac{1}{n}\right).$$

Similarly, the third and the fourth term on the right-hand side of (2.19) are $O\left(1/n\right)$. Thus, we conclude that

$$E\left(\widehat{\gamma}_{2,X}\left(h\right)\right) = E\left(\widetilde{\gamma}_{2,X}\left(h\right)\right) + O\left(\frac{1}{n}\right). \tag{2.20}$$

For $\widehat{\gamma}_X^2\left(h\right)$ we have

$$\begin{aligned}
\widehat{\gamma}_X^2\left(h\right) = \widetilde{\gamma}_X^2\left(h\right) &+ \frac{1}{n^4}\sum_{t,r=1}^{n-|h|}\sum_{s,q=1}^{n}\left(X_t-\mu\right)\left(X_r-\mu\right)\left(X_s-\mu\right)\left(X_q-\mu\right)\\
&+ \frac{1}{n^4}\sum_{t,r=1}^{n-|h|}\sum_{s,q=1}^{n}\left(X_{t+|h|}-\mu\right)\left(X_{r+|h|}-\mu\right)\left(X_s-\mu\right)\left(X_q-\mu\right)\\
&+ \frac{\left(n-|h|\right)^2}{n^6}\sum_{t,r,s,q=1}^{n}\left(X_t-\mu\right)\left(X_r-\mu\right)\left(X_s-\mu\right)\left(X_q-\mu\right)\\
&- \frac{2}{n^3}\sum_{t,r=1}^{n-|h|}\sum_{s=1}^{n}\left(X_r-\mu\right)\left(X_{r+|h|}-\mu\right)\left(X_s-\mu\right)\left(X_t-\mu\right)\\
&- \frac{2}{n^3}\sum_{t,r=1}^{n-|h|}\sum_{s=1}^{n}\left(X_r-\mu\right)\left(X_{r+|h|}-\mu\right)\left(X_s-\mu\right)\left(X_{t+|h|}-\mu\right)\\
&+ \frac{2\left(n-|h|\right)}{n^4}\sum_{r=1}^{n-|h|}\sum_{s,t=1}^{n}\left(X_r-\mu\right)\left(X_{r+|h|}-\mu\right)\left(X_s-\mu\right)\left(X_t-\mu\right)\\
&+ \frac{2}{n^4}\sum_{t,r=1}^{n-|h|}\sum_{s,q=1}^{n}\left(X_q-\mu\right)\left(X_{r+|h|}-\mu\right)\left(X_s-\mu\right)\left(X_t-\mu\right)\\
&- \frac{2\left(n-|h|\right)}{n^5}\sum_{t=1}^{n-|h|}\sum_{s,q,r=1}^{n}\left(X_q-\mu\right)\left(X_r-\mu\right)\left(X_s-\mu\right)\left(X_t-\mu\right)\\
&- \frac{2\left(n-|h|\right)}{n^5}\sum_{t=1}^{n-|h|}\sum_{s,q,r=1}^{n}\left(X_q-\mu\right)\left(X_r-\mu\right)\left(X_s-\mu\right)\left(X_{t+|h|}-\mu\right).
\end{aligned} \tag{2.21}$$

Taking expectation on both sides of the equation (2.21), we can show that

$$E\left(\widehat{\gamma}_X^2\left(h\right)\right) = E\left(\widetilde{\gamma}_X^2\left(h\right)\right) + O\left(\frac{1}{n}\right). \tag{2.22}$$

To see (2.22) consider for instance the 6^{th} term of (2.21). Using the relation (2.18) we get

$$\begin{aligned}
\frac{2}{n^3}\sum_{t,r=1}^{n-|h|}\sum_{s=1}^{n}\Big\{&\gamma_X\left(h\right)\gamma_X\left(s-t-|h|\right)+\gamma_X\left(t-r+|h|\right)\gamma_X\left(s-r-|h|\right)+\gamma_X\left(s-r\right)\gamma_X\left(t-r\right)\\
&+ cum_X\left(|h|, t-r+|h|, s-r\right)\Big\}.
\end{aligned} \tag{2.23}$$

The first term of (2.23) equals

$$\frac{2(n-|h|)}{n^3}\gamma_X(h)\sum_{r=-(n-1)}^{(n-1)} n\phi_n(r;h,0)\,\gamma_X(r-|h|),$$

and is bounded in absolute value by

$$\frac{2(n+|h|)}{n^2}\left|\gamma_X(h)\right|\sum_{r=-\infty}^{\infty}\left|\gamma_X(r-|h|)\right| = O\left(\frac{1}{n}\right).$$

Moreover, the second term of (2.23) is bounded in absolute value by

$$\frac{2}{n^3}\sum_{r=1}^{n-|h|}\sum_{z,b=-(n-1)}^{(n-1)}\left|\gamma_X(z+|h|)\right|\left|\gamma_X(b-|h|)\right| \le \frac{2(n-|h|)}{n^3}\sum_{z,b=-\infty}^{\infty}\left|\gamma_X(z+|h|)\right|\left|\gamma_X(b-|h|)\right|.$$

Thus, the second term of (2.23) and similarly, the third term are $O\left(1/n^2\right)$. Finally, for the last term we have

$$\begin{aligned}\frac{2}{n^3}\sum_{t,r=1}^{n-|h|}\sum_{s=1}^{n} cum_X(|h|,t-r+|h|,s-r) &\le \frac{2}{n^3}\sum_{r=1}^{n-|h|}\sum_{z,b=-(n-1)}^{n-1}\left|cum_X(|h|,z+|h|,b)\right| \\ &\le \frac{2(n-|h|)}{n^3}\sum_{z,b=-\infty}^{\infty}\left|cum_X(|h|,z+|h|,b)\right| \\ &= O\left(\frac{1}{n^2}\right).\end{aligned}$$

From the above derivations, we conclude that (2.23) is $O(1/n)$. Similar arguments show that the expectation of all the other terms on the right-hand side of (2.21) which follow the term $E\left(\widetilde{\gamma}_X^2(h)\right)$, have order at least equal to $O(1/n)$. Thus, from (2.20) we get

$$E(\widehat{N}_n) = E(\widetilde{N}_n) + O\left(\frac{M_n}{n}\right). \tag{2.24}$$

Using a Taylor series argument observe that

$$\widetilde{\eta}_{4,\varepsilon} = \frac{\widetilde{N}_n}{\widetilde{\gamma}_X^2(0)} = \frac{\widetilde{N}_n}{\gamma_X^2(0)} + \left(\widetilde{\gamma}_X^2(0)-\gamma_X^2(0)\right)\left(-\frac{\widetilde{N}_n}{\gamma_X^4(0)}\right) + \frac{\widetilde{N}_n}{2c^3}\left(\widetilde{\gamma}_X^2(0)-\gamma_X^2(0)\right)^2, \tag{2.25}$$

where $\min\left\{\gamma_X^2(0),\widetilde{\gamma}_X^2(0)\right\} < c < \max\left\{\gamma_X^2(0),\widetilde{\gamma}_X^2(0)\right\}$. Cauchy-Schwartz's inequality yields then

$$E\left|\left(\frac{\widetilde{N}_n}{c^3}\left(\widetilde{\gamma}_X^2(0)-\gamma_X^2(0)\right)^2\right)\right| \le \sqrt{E\left(\frac{\widetilde{N}_n}{c^3}\right)^2}\sqrt{E\left(\widetilde{\gamma}_X^2(0)-\gamma_X^2(0)\right)^4} = O\left(\frac{1}{n}\right),$$

and

$$E\left|\left(N_n\left(\widetilde{\gamma}_X^2(0)-\gamma_X^2(0)\right)\right)\right| \le \sqrt{E(\widetilde{N}_n)^2}\sqrt{E\left(\widetilde{\gamma}_X^2(0)-\gamma_X^2(0)\right)^2} = O\left(\frac{1}{\sqrt{n}}\right).$$

Therefore,

$$E\left(\widetilde{\eta}_{4,\varepsilon}\right) = \frac{1}{\gamma_X^2(0)}E(\widetilde{N}_n) + O\left(\frac{1}{\sqrt{n}}\right), \tag{2.26}$$

and, similarly,

$$E\left(\breve{\eta}_{4,\varepsilon}\right) = \frac{1}{\gamma_X^2\left(0\right)} E(\widehat{N}_n) + O\left(\frac{1}{\sqrt{n}}\right). \tag{2.27}$$

Thus, the required result is obtained from (2.24), (2.26) and (2.27).

To prove assertion (ii) of the lemma notice first that by tedious algebra we have that

$$Cov\left(\widehat{\gamma}_{2,X}\left(h_1\right), \widehat{\gamma}_{2,X}\left(h_2\right)\right) = Cov\left(\widetilde{\gamma}_{2,X}\left(h_1\right), \widetilde{\gamma}_{2,X}\left(h_2\right)\right) + O\left(\frac{1}{n^2}\right),$$

$$Cov\left(\widehat{\gamma}_{2,X}\left(h_1\right), \widehat{\gamma}_X^2\left(h_2\right)\right) = Cov\left(\widetilde{\gamma}_{2,X}\left(h_1\right), \widetilde{\gamma}_X^2\left(h_2\right)\right) + O\left(\frac{1}{n^2}\right),$$

and

$$Cov\left(\widehat{\gamma}_X^2\left(h_1\right), \widehat{\gamma}_X^2\left(h_2\right)\right) = Cov\left(\widetilde{\gamma}_X^2\left(h_1\right), \widetilde{\gamma}_X^2\left(h_2\right)\right) + O\left(\frac{1}{n^2}\right).$$

From the above relations we get

$$Var(\widehat{N}_n) = Var(\widetilde{N}_n) + O\left(\frac{M_n^2}{n^2}\right). \tag{2.28}$$

Calculating the variance of equation (2.25)

$$\begin{aligned}
Var\left(\widetilde{\eta}_{4,\varepsilon}\right) = {} & \frac{1}{\gamma_X^4\left(0\right)} Var(\widetilde{N}_n) + Var\Big(\left(\widetilde{\gamma}_X^2\left(0\right) - \gamma_X^2\left(0\right)\right) \frac{\widetilde{N}_n}{\gamma_X^4\left(0\right)}\Big) \\
& + \frac{1}{4} Var\Big(\frac{\widetilde{N}_n}{c^3}\left(\widetilde{\gamma}_X^2\left(0\right) - \gamma_X^2\left(0\right)\right)^2\Big) \\
& - \frac{2}{\gamma_X^6\left(0\right)} Cov\left(\widetilde{N}_n, \left(\widetilde{\gamma}_X^2\left(0\right) - \gamma_X^2\left(0\right)\right)\widetilde{N}_n\right) \\
& + \frac{1}{\gamma_X^2\left(0\right)} Cov\Big(\widetilde{N}_n, \frac{\widetilde{N}_n}{c^3}\left(\widetilde{\gamma}_X^2\left(0\right) - \gamma_X^2\left(0\right)\right)^2\Big) \\
& - \frac{1}{\gamma_X^4\left(0\right)} Cov\Big(\widetilde{N}_n\left(\widetilde{\gamma}_X^2\left(0\right) - \gamma_X^2\left(0\right)\right), \frac{\widetilde{N}_n}{c^3}\left(\widetilde{\gamma}_X^2\left(0\right) - \gamma_X^2\left(0\right)\right)^2\Big),
\end{aligned}$$

where

$$Var\Big(\left(\widetilde{\gamma}_X^2\left(0\right) - \gamma_X^2\left(0\right)\right)\frac{\widetilde{N}_n}{\gamma_X^4\left(0\right)}\Big) = O\left(\frac{1}{n}\right), \quad Var\Big(\frac{\widetilde{N}_n}{c^3}\left(\widetilde{\gamma}_X^2\left(0\right) - \gamma_X^2\left(0\right)\right)^2\Big) = O\left(\frac{1}{n^2}\right),$$

and

$$Cov\left(\widetilde{N}_n, \left(\widetilde{\gamma}_X^2\left(0\right) - \gamma_X^2\left(0\right)\right)\widetilde{N}_n\right) \leq \sqrt{Var(\widetilde{N}_n)}\sqrt{Var\left(\left(\widetilde{\gamma}_X^2\left(0\right) - \gamma_X^2\left(0\right)\right)\widetilde{N}_n\right)} = O\left(\frac{\sqrt{M_n}}{n}\right).$$

Similarly,

$$Cov\Big(\widetilde{N}_n\left(\widetilde{\gamma}_X^2\left(0\right) - \gamma_X^2\left(0\right)\right), \frac{\widetilde{N}_n}{c^3}\left(\widetilde{\gamma}_X^2\left(0\right) - \gamma_X^2\left(0\right)\right)^2\Big) = O\left(\frac{1}{n^{3/2}}\right),$$

and

$$Cov\Big(\widetilde{N}_n, \frac{\widetilde{N}_n}{c^3}\left(\widetilde{\gamma}_X^2\left(0\right) - \gamma_X^2\left(0\right)\right)^2\Big) = O\left(\frac{\sqrt{M_n}}{n^{3/2}}\right).$$

Thus,

$$\frac{n}{M_n} Var\left(\widetilde{\eta}_{4,\varepsilon}\right) = \frac{1}{\gamma_X^4(0)} \frac{n}{M_n} Var(\widetilde{N}_n) + O\left(\frac{1}{\sqrt{M_n}}\right). \tag{2.29}$$

Along the same lines and with $\widetilde{\eta}_{4,\varepsilon}$, $\widetilde{N}_n$ and $\widetilde{\gamma}_X^2(0)$ replaced by $\breve{\eta}_{4,\varepsilon}$, $\widehat{N}_n$ and $\widehat{\gamma}_X^2(0)$ respectively we can show that

$$\frac{n}{M_n} Var\left(\breve{\eta}_{4,\varepsilon}\right) = \frac{1}{\gamma_X^4(0)} \frac{n}{M_n} Var(\widehat{N}_n) + O\left(\frac{1}{\sqrt{M_n}}\right). \tag{2.30}$$

Finally, equations (2.28), (2.29) and (2.30) yield assertion (ii).

Consider next assertion (iii) and notice that

$$\begin{aligned}\widehat{\gamma}_X(h) = \widetilde{\gamma}_X(h) &+ \frac{1}{n}\left(\mu - \overline{X}_n\right) \sum_{t=1}^{n-|h|} (X_t - \mu) + \frac{1}{n}\left(\mu - \overline{X}_n\right) \sum_{t=1}^{n-|h|} \left(X_{t+|h|} - \mu\right) \\ &+ \frac{n-|h|}{n}\left(\mu - \overline{X}_n\right)^2,\end{aligned}$$

and

$$\begin{aligned}\widehat{\gamma}_{2,X}(h) = \widetilde{\gamma}_{2,X}(h) &+ \frac{1}{n}\left(\mu_2 - \overline{X}_{2,n}\right) \sum_{t=1}^{n-|h|} \left(X_t^2 - \mu_2\right) + \frac{1}{n}\left(\mu_2 - \overline{X}_{2,n}\right) \sum_{t=1}^{n-|h|} \left(X_{t+|h|}^2 - \mu_2\right) \\ &+ \frac{n-|h|}{n}\left(\mu_2 - \overline{X}_{2,n}\right)^2.\end{aligned}$$

Using

(I) $\sum_{t=1}^{n-|h|} (X_t - \mu) = O_P(\sqrt{n})$,

(II) $\sum_{t=1}^{n-|h|} (X_t^2 - \mu_2) = O_P(\sqrt{n})$,

(III) $\mu - \overline{X}_n = O_P(1/\sqrt{n})$,

(IV) $\mu_2 - \overline{X}_{2,n} = O_P(1/\sqrt{n})$,

we get that

$$\widehat{\gamma}_{2,X}(h) = \widetilde{\gamma}_{2,X}(h) + O_P\left(\frac{1}{n}\right) \quad \text{and} \quad \widehat{\gamma}_X^2(h) = \widetilde{\gamma}_X^2(h) + O_P\left(\frac{1}{n}\right),$$

and, therefore,

$$\breve{\eta}_{4,\varepsilon} = \widetilde{\eta}_{4,\varepsilon} + O_P\left(\frac{M_n}{n}\right). \tag{2.31}$$

From equation (2.31) and assertion (i) we get that

$$\begin{aligned}\sqrt{\frac{n}{M_n}}\left(\breve{\eta}_{4,\varepsilon} - E\left(\breve{\eta}_{4,\varepsilon}\right)\right) &= \sqrt{\frac{n}{M_n}}\left(\breve{\eta}_{4,\varepsilon} - \widetilde{\eta}_{4,\varepsilon}\right) + \sqrt{\frac{n}{M_n}}\left(\widetilde{\eta}_{4,\varepsilon} - E\left(\breve{\eta}_{4,\varepsilon}\right)\right) \\ &= \sqrt{\frac{n}{M_n}}\left(\widetilde{\eta}_{4,\varepsilon} - E\left(\widetilde{\eta}_{4,\varepsilon}\right)\right) + O_P\left(\frac{1}{\sqrt{M_n}} + \sqrt{\frac{M_n}{n}}\right),\end{aligned}$$

which concludes the proof. □

Proof of Theorem 2.2.1:

Let

$$\widetilde{\eta}_{4,\varepsilon} = \frac{\widetilde{N}_n}{\widetilde{\gamma}^2_X(0)}, \quad \text{where} \quad \widetilde{N}_n = \sum_{h=-(n-1)}^{(n-1)} w\left(\frac{h}{M_n}\right)\left(\widetilde{\gamma}_{2,X}(h) - 2\widetilde{\gamma}^2_X(h)\right).$$

(i) By a Taylor series argument and as in the proof of Lemma 2.7.2 (i), equation (2.25), we get

$$E(\widetilde{\eta}_{4,\varepsilon}) = \frac{1}{\gamma^2_X(0)} E(\widetilde{N}_n) + O\left(\frac{1}{\sqrt{n}}\right). \tag{2.32}$$

Now,

$$E(\widetilde{N}_n) = \sum_{h=-(n-1)}^{n-1} w\left(\frac{h}{M_n}\right)\Big(E\big(\widetilde{\gamma}_{2,X}(h)\big) - 2E\left(\widetilde{\gamma}^2_X(h)\right)\Big),$$

where $E\left(\widetilde{\gamma}^2_X(h)\right)$ equals

$$\begin{aligned}
&\frac{1}{n^2}\sum_{t,s=1}^{n-|h|} E\left(X_t-\mu\right)\left(X_{t+|h|}-\mu\right)\left(X_s-\mu\right)\left(X_{s+|h|}-\mu\right)\\
&= \frac{(n-|h|)^2}{n^2}\gamma^2_X(h) + \frac{1}{n^2}\sum_{t,s=1}^{n-|h|}\gamma^2_X(t-s) + \frac{1}{n^2}\sum_{t,s=1}^{n-|h|}\gamma_X(t-s-|h|)\,\gamma_X(t+|h|-s)\\
&\quad + \frac{1}{n^2}\sum_{t,s=1}^{n-|h|} cum_X(|h|, s-t, s-t+|h|)\\
&= \frac{(n-|h|)^2}{n^2}\gamma^2_X(h) + \frac{1}{n}\sum_{r=-(n-1)}^{n-1}\phi_n(r;h,h)\gamma^2_X(r)\\
&\quad + \frac{1}{n}\sum_{r=-(n-1)}^{n-1}\phi_n(r;h,h)\gamma_X(r+|h|)\,\gamma_X(r-|h|)\\
&\quad + \frac{1}{n}\sum_{r=-(n-1)}^{n-1}\phi_n(r;h,h) cum_X(|h|, -r, -r+|h|).
\end{aligned}$$

Furthermore, it is easy to see that

$$E\left(\widetilde{\gamma}_{2,X}(h)\right) = \frac{n-|h|}{n}\gamma_{2,X}(h).$$

Thus,

$$\begin{aligned}
E(\widetilde{N}_n) = &\sum_{h=-(n-1)}^{n-1} w\left(\frac{h}{M_n}\right)\gamma_{2,X}(h) - \sum_{h=-(n-1)}^{n-1}\frac{|h|}{n} w\left(\frac{h}{M_n}\right)\gamma_{2,X}(h)\\
&- 2\sum_{h=-(n-1)}^{n-1} w\left(\frac{h}{M_n}\right)\gamma^2_X(h)
\end{aligned}$$

$$
\begin{aligned}
&- 2 \sum_{h=-(n-1)}^{n-1} \frac{h^2}{n^2} w\left(\frac{h}{M_n}\right)\gamma_X^2(h) + 4 \sum_{h=-(n-1)}^{n-1} \frac{|h|}{n} w\left(\frac{h}{M_n}\right)\gamma_X^2(h) \\
&- \frac{2}{n} \sum_{h=-M_n}^{M_n} \sum_{r=-(n-1)}^{n-1} \phi_n(r; h, h) w\left(\frac{h}{M_n}\right)\gamma_X^2(r) \\
&- \frac{2}{n} \sum_{h=-M_n}^{M_n} \sum_{r=-(n-1)}^{n-1} \phi_n(r; h, h) w\left(\frac{h}{M_n}\right)\gamma_X(r + |h|)\, \gamma_X(r - |h|) \\
&- \frac{2}{n} \sum_{h=-M_n}^{M_n} \sum_{r=-(n-1)}^{n-1} \phi_n(r; h, h) w\left(\frac{h}{M_n}\right) cum_X(|h|, -r, -r + |h|).
\end{aligned}
$$

By Lemma 2.7.1 and the summability conditions

$$
\sum_{h=-\infty}^{\infty} |h|\,|\gamma_{2,X}(h)| < \infty \quad \text{and} \quad \sum_{h=-\infty}^{\infty} h^2 \gamma_X^2(h) < \infty,
$$

we get that

$$
E(\widetilde{N}_n) = \sum_{h=-(n-1)}^{n-1} w\left(\frac{h}{M_n}\right)\left(\gamma_{2,X}(h) - 2{\gamma_X}^2(h)\right) + O\left(\frac{M_n}{n} + \frac{1}{n}\right).
$$

Furthermore,

$$
\begin{aligned}
M_n^2\big[E(\widetilde{N}_n) - \kappa_{4,\varepsilon}\big] = {} & M_n^2 \sum_{h=-M_n}^{M_n} \left[w\left(\frac{h}{M_n}\right) - 1\right]\left[\gamma_{2,X}(h) - 2\gamma_X^2(h)\right] \\
& - 2M_n^2 \sum_{h=M_n+1}^{\infty} \left[\gamma_{2,X}(h) - 2\gamma_X^2(h)\right] + O\left(\frac{M_n^3}{n}\right).
\end{aligned} \tag{2.33}
$$

The first term on the right-hand side of (2.33) can be written as

$$
\sum_{h=-M_n^*}^{M_n^*} \frac{w\left(\frac{h}{M_n}\right) - 1}{\left(\frac{h}{M_n}\right)^2} h^2 \left[\gamma_{2,X}(h) - 2\gamma_X^2(h)\right] + 2 \sum_{h=M_n^*+1}^{M_n} \frac{w\left(\frac{h}{M_n}\right) - 1}{\left(\frac{h}{M_n}\right)^2} h^2 \left[\gamma_{2,X}(h) - 2\gamma_X^2(h)\right], \tag{2.34}
$$

for any integer $M_n^* < M_n$. Consider the first term of (2.34), For any $\varepsilon > 0$ we can choose $\delta = \delta(\varepsilon)$ so that for $|h/M_n| < \delta$,

$$
\left| \frac{w\left(\frac{h}{M_n}\right) - 1}{\left(\frac{h}{M_n}\right)^2} + C_w \right| < \varepsilon, \quad \text{where} \quad C_w := \int_{-1}^{1} x^2 w(x)\, dx.
$$

Thus for $M_n^* = [\delta M_n]$, the first term in (2.34) is within

$$
\varepsilon' = \varepsilon \sum\nolimits_{h=-\infty}^{\infty} h^2 \left|\gamma_{2,X}(h) - 2\gamma_X^2(h)\right| \quad \text{of} \quad - C_w \sum\nolimits_{h=-M_n^*}^{M_n^*} h^2 \left[\gamma_{2,X}(h) - 2\gamma_X^2(h)\right],
$$

which implies that this term converges to

$$
-C_w \sum_{h=-\infty}^{\infty} h^2 \left[\gamma_{2,X}(h) - 2\gamma_X^2(h)\right],
$$

as $n \to \infty$. Since $|w(x)| \leq 1$ we have

$$\frac{|1 - w(x)|}{x^2} \leq \frac{2}{\delta^2}, \quad |x| \geq \delta.$$

The second term in (2.34) is in absolute value no greater than

$$\frac{4}{\delta^2} \sum_{h=[\delta M_n]+1}^{\infty} h^2 \left|\gamma_{2,X}(h) - 2\gamma_X^2(h)\right|,$$

which converges to zero as $n \to \infty$.

The second term on the right-hand side of (2.33) is bounded in absolute value by

$$2M_n^2 \sum_{h=M_n+1}^{\infty} \left|\gamma_{2,X}(h) - 2\gamma_X^2(h)\right| \leq 2 \sum_{h=M_n+1}^{\infty} h^2 \left|\gamma_{2,X}(h) - 2\gamma_X^2(h)\right|,$$

which also converges to zero, as $n \to \infty$. Thus,

$$M_n^2 \left[E(\widetilde{N}_n) - \kappa_{4,\varepsilon}\right] \to -C_w \sum_{h=-\infty}^{\infty} h^2 \left[\gamma_{2,X}(h) - 2\gamma_X^2(h)\right], \tag{2.35}$$

as $n \to \infty$. Using equation (2.32) and Lemma 2.7.2 (i), we conclude that

$$\begin{aligned}
\lim_{n\to\infty} M_n^2 \left(E\left(\breve{\eta}_{4,\varepsilon}\right) - \eta_{4,\varepsilon}\right) &= -\frac{C_w}{\gamma_X^2(0)} \sum_{h=-\infty}^{\infty} h^2 \left[\gamma_{2,X}(h) - 2\gamma_X^2(h)\right] \\
&= C_w \frac{\kappa_{4,\varepsilon}}{\sigma_\varepsilon^2\, \gamma_X^2(0)} 2\pi f_{\widetilde{X}}''(0),
\end{aligned}$$

where $f_{\widetilde{X}}''$ denotes the second derivative of the spectral density of the process $\widetilde{\mathbf{X}} = \{\widetilde{X}_t : t \in \mathbb{Z}\}$, where $\widetilde{X}_t = \sum_{j=-\infty}^{\infty} \psi_j^2 \varepsilon_{t-j}$.

(ii) Notice that

$$\begin{aligned}
\operatorname{Var}(\widetilde{N}_n) = &\sum_{h_1,h_2=-(n-1)}^{n-1} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_2}{M_n}\right) Cov\big(\widetilde{\gamma}_{2,X}(h_1), \widetilde{\gamma}_{2,X}(h_2)\big) \\
&- 4 \sum_{h_1,h_2=-(n-1)}^{n-1} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_2}{M_n}\right) Cov\left(\widetilde{\gamma}_{2,X}(h_1), \widetilde{\gamma}_X^2(h_2)\right) \\
&+ 4 \sum_{h_1,h_2=-(n-1)}^{n-1} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_2}{M_n}\right) Cov\left(\widetilde{\gamma}_X^2(h_1), \widetilde{\gamma}_X^2(h_2)\right) \\
= &\, V_{1,n} - 4V_{2,n} + 4V_{3,n} \text{ (say).}
\end{aligned}$$

We show that

$$\frac{n}{M_n} V_{1,n} = 2\Big(\sum_{u=-\infty}^{\infty} \gamma_{2,X}(u)\Big)^2 \int_{-1}^{1} w^2(x)\, dx + O\left(\frac{1}{M_n}\right), \tag{2.36}$$

$$\frac{n}{M_n} V_{2,n} = O\left(\frac{1}{M_n}\right), \tag{2.37}$$

and

$$\frac{n}{M_n} V_{3,n} = O\left(\frac{1}{M_n}\right). \tag{2.38}$$

To see why relation (2.36) is true notice that

$$\begin{aligned} &nCov\left(\widetilde{\gamma}_{2,X}\left(h_1\right), \widetilde{\gamma}_{2,X}\left(h_2\right)\right) \\ &= \sum_{r=-(n-1)}^{n-1} \phi_n\left(r; h_1, h_2\right) \Big\{ \gamma_{2,X}\left(r\right) \gamma_{2,X}\left(r + \left|h_1\right| - \left|h_2\right|\right) + \gamma_{2,X}\left(r + \left|h_1\right|\right) \gamma_{2,X}\left(r - \left|h_2\right|\right) \\ &\quad + cum_{X^2}\left(\left|h_1\right|, -r, \left|h_2\right| - r\right) \Big\}. \end{aligned}$$

Thus, $(n/M_n)V_{1,n}$ equals

$$\begin{aligned} &\frac{1}{M_n} \sum_{h_1,h_2=-M_n}^{M_n} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_2}{M_n}\right) \sum_{r=-(n-1)}^{n-1} \phi_n\left(r; h_1, h_2\right) \Big\{ \gamma_{2,X}\left(r\right) \gamma_{2,X}\left(r + \left|h_1\right| - \left|h_2\right|\right) \\ &\quad + \gamma_{2,X}\left(r + \left|h_1\right|\right) \gamma_{2,X}\left(r - \left|h_2\right|\right) + cum_{X^2}\left(\left|h_1\right|, -r, \left|h_2\right| - r\right) \Big\}. \end{aligned} \tag{2.39}$$

We deal with each of the three terms of (2.39) separately. Consider first the term

$$\begin{aligned} &\frac{1}{M_n} \sum_{h_1,h_2=-M_n}^{M_n} \sum_{r=-(n-1)}^{n-1} \phi_n\left(r; h_1, h_2\right) w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_2}{M_n}\right) \gamma_{2,X}\left(r + \left|h_1\right|\right) \gamma_{2,X}\left(r - \left|h_2\right|\right) \\ &= \frac{1}{M_n} \sum_{u,v=-(M_n+n-1)}^{M_n+n-1} \sum_{r=\max[u-M_n, v-M_n, -(n-1)]}^{\min[u-M_n, v-M_n, (n-1)]} \phi_n\left(r; r-u, v-r\right) \\ &\quad \times w\left(\frac{r-u}{M_n}\right) w\left(\frac{v-r}{M_n}\right) \gamma_{2,X}\left(u\right) \gamma_{2,X}\left(v\right), \end{aligned} \tag{2.40}$$

of (2.39), where the sum over r is equal to zero if the lower limit is greater than the upper limit. The difference between (2.40) and

$$\begin{aligned} &\frac{1}{M_n} \sum_{u,v=-m}^{m} \sum_{r=\max[u-M_n, v-M_n, -(n-1)]}^{\min[u-M_n, v-M_n, (n-1)]} \phi_n\left(r; r-u, v-r\right) w\left(\frac{r-u}{M_n}\right) w\left(\frac{v-r}{M_n}\right) \\ &\quad \times \gamma_{2,X}\left(u\right) \gamma_{2,X}\left(v\right), \end{aligned} \tag{2.41}$$

is, in absolute value, bounded by

$$4\left(2 + \frac{1}{M_n}\right) \sum_{u=-\infty}^{\infty} \sum_{v=m+1}^{\infty} \left|\gamma_{2,X}\left(u\right)\right| \left|\gamma_{2,X}\left(v\right)\right|,$$

which can be made arbitrarily small if $m\,(\leq M_n)$ is sufficiently large taking into account that $|w\,(x)|$ is bounded and $\sum_{r=-\infty}^{\infty}|\gamma_X\,(r)| < \infty$. Since $w\,(x)$ is continuous, we have for $|u| \leq m, |v| \leq m$ and M_n sufficiently large that

$$\left|w\left(\frac{r-u}{M_n}\right)w\left(\frac{v-r}{M_n}\right) - w^2\left(\frac{r}{M_n}\right)\right| < \varepsilon, \tag{2.42}$$

for r such that $-M_n \leq r-u \leq M_n$, $-M_n \leq r-v \leq M_n$, and $-M_n \leq r \leq M_n$. For $|u| \leq m, |v| \leq m, |r| \leq m + M_n$

$$\phi_n\,(r; r-u, v-r) \geq 1 - \frac{5m+3M_n}{n}.$$

Thus, the difference between (2.41) and

$$\frac{1}{M_n}\sum_{r=-M_n}^{M_n} w^2\left(\frac{r}{M_n}\right)\Big(\sum_{v=-m}^{m}\gamma_{2,X}\,(v)\Big)^2 \tag{2.43}$$

can be made arbitrarily small for n sufficiently large. Now, since

$$\frac{1}{M_n}\sum_{r=-M_n}^{M_n} w^2\left(\frac{r}{M_n}\right) \to \int_{-1}^{1} w^2\,(x)\,dx, \quad \text{as } M_n \to \infty,$$

(2.43) converges to

$$\Big(\sum_{u=-\infty}^{\infty}\gamma_{2,X}\,(u)\Big)^2\int_{-1}^{1} w^2\,(x)\,dx.$$

Consider next the term

$$\begin{aligned}
&\frac{1}{M_n}\sum_{h_1,h_2=-M_n}^{M_n}\sum_{r=-(n-1)}^{n-1}\phi_n\,(r;h_1,h_2)w\left(\frac{h_1}{M_n}\right)w\left(\frac{h_2}{M_n}\right)\gamma_{2,X}\,(r)\,\gamma_{2,X}\,(r+|h_1|-|h_2|)\\
&=\frac{1}{M_n}\sum_{u=-(n-1)}^{n-1}\sum_{v=u-2M_n}^{u+2M_n}\sum_{s=\max\{u,v\}-M_n}^{\min\{u,v\}+M_n}\phi_n\,(u;u-s,v-s)w\left(\frac{u-s}{M_n}\right)w\left(\frac{v-s}{M_n}\right)\\
&\quad\times\gamma_{2,X}\,(u)\,\gamma_{2,X}\,(v)\,,
\end{aligned} \tag{2.44}$$

of (2.39). Using the same arguments as for (2.40), the right hand side of equation (2.44) equals

$$\frac{1}{M_n}\sum_{u,v=-m}^{m}\sum_{s=\max\{u,v\}-M_n}^{\min\{u,v\}+M_n}\phi_n\,(u;u-s,v-s)w\left(\frac{u-s}{M_n}\right)w\left(\frac{v-s}{M_n}\right)\gamma_{2,X}\,(u)\,\gamma_{2,X}\,(v)+o(1). \tag{2.45}$$

Taking into account (2.42), the difference between (2.45) and

$$\sum_{u,v=-m}^{m} \sum_{s=-M_n}^{M_n} \frac{1}{M_n} w^2 \left(\frac{s}{M_n} \right) \gamma_{2,X}(u) \gamma_{2,X}(v), \tag{2.46}$$

can be made arbitrarily small for n sufficiently large. Thus, for m sufficiently large the limit of (2.46) is

$$\Big(\sum_{u=-\infty}^{\infty} \gamma_{2,X}(u) \Big)^2 \int_{-1}^{1} w^2(x)\, dx.$$

Finally, consider the term

$$\frac{1}{M_n} \sum_{h_1,h_2=-M_n}^{M_n} \sum_{r=-(n-1)}^{n-1} \phi_n(r; h_1, h_2) w\left(\frac{h_1}{M_n} \right) w\left(\frac{h_2}{M_n} \right) cum_{X^2}(h_1, -r, h_2 - r)$$

of (2.39).This term is in absolute value no greater than

$$\frac{1}{M_n} \sum_{r,s,t=-\infty}^{\infty} |cum_{X^2}(r,s,t)| = O\left(\frac{1}{M_n} \right),$$

by Lemma 2.7.1. Thus,

$$\begin{aligned} \frac{n}{M_n} V_{1,n} &= 2 \Big(\sum_{u=-\infty}^{\infty} \gamma_{2,X}(u) \Big)^2 \int_{-1}^{1} w^2(x)\, dx + O\left(\frac{1}{M_n} \right) \\ &= 8\pi^2 f_{X^2}^2(0) \int_{-1}^{1} w^2(x)\, dx + O\left(\frac{1}{M_n} \right). \end{aligned}$$

To show (2.37), notice that

$$\begin{aligned} \frac{n}{M_n} V_{2,n} = \frac{1}{M_n n^2} \sum_{h_1,h_2=-(n-1)}^{n-1} w\left(\frac{h_1}{M_n} \right) w\left(\frac{h_2}{M_n} \right) \sum_{t=1}^{n-|h_1|} \sum_{z,s=1}^{n-|h_2|} Cov\Big(\left(X_t^2 - \mu_2\right) \left(X_{t+|h_1|}^2 - \mu_2\right), \\ \left(X_z - \mu\right) \left(X_{z+|h_2|} - \mu\right) \left(X_s - \mu\right) \left(X_{s+|h_2|} - \mu\right) \Big). \end{aligned} \tag{2.47}$$

Using the notation

$$Cum\left(X_t^2, X_r^2, X_z, X_g\right) \equiv cum_{X,X^2}(r-t, z-t, g-t),$$

we get from Lemma 2.7.1 that

$$\sum_{t,r,z,g=-\infty}^{\infty} |cum_{X,X^2}(r-t, z-t, g-t)| < \infty.$$

To bound (2.47) we evaluate the expectation $E\left(\prod_{i=1}^{6} Z_i \right)$ of the random variables Z_i with $E(Z_i) = 0$, $i = 1, 2, ..., 6$, using all possible decompositions in products of triples

$E\left(Z_iZ_jZ_k\right)E\left(Z_lZ_mZ_n\right)$, of quadriples and pairs $E\left(Z_iZ_jZ_kZ_l\right)\,E\left(Z_mZ_n\right)$, and of pairs $E\left(Z_iZ_j\right)E\left(Z_kZ_l\right)E\left(Z_mZ_n\right)$ for indices $i,j,k,l,m,n\in\{1,2,...,6\}$ and the cumulant term $Cum\left(Z_1,Z_2,Z_3,Z_4,Z_5,Z_6\right)$. Evaluating the expansion term in (2.47) in that way, it follows by similar arguments that all terms have at least order $O(1/M_n)$. For instance, the term

$$\frac{1}{M_nn^2}\sum_{h_1,h_2=-(n-1)}^{n-1}w\left(\frac{h_1}{M_n}\right)w\left(\frac{h_2}{M_n}\right)\sum_{t=1}^{n-|h_1|}\sum_{z,s=1}^{n-|h_2|}\gamma_X\left(h_2\right)Cum\left(X_t^2,X_{t+|h_1|}^2,X_s,X_{s+|h_1|}\right),$$

appearing in the evaluation of the expectation term in (2.47) is bounded in absolute value by

$$\frac{1}{M_n}\sup_{-1\leq x\leq 1}w^2(x)\sum_{g=-\infty}^{\infty}\left|\gamma_X\left(g\right)\right|\sum_{r,b,z,h=-\infty}^{\infty}\left|cum_{X,X^2}\left(r,b,z,h\right)\right|=O\left(\frac{1}{M_n}\right).$$

Finally, to see (2.38), notice that

$$\begin{aligned}
&Cov\left(\widetilde{\gamma}_X^2\left(h_1\right),\widetilde{\gamma}_X^2\left(h_2\right)\right)\\
&=Cov\Big(\Big\{\frac{1}{n}\sum_{t=1}^{n-|h_1|}\left(X_t-\mu\right)\left(X_{t+|h_1|}-\mu\right)\Big\}^2,\Big\{\frac{1}{n}\sum_{z=1}^{n-|h_2|}\left(X_z-\mu\right)\left(X_{z+|h_2|}-\mu\right)\Big\}^2\Big)\\
&=\frac{1}{n^4}\sum_{t,s=1}^{n-|h_1|}\sum_{z,q=1}^{n-|h_2|}Cov\Big(\left(X_t-\mu\right)\left(X_{t+|h_1|}-\mu\right)\left(X_s-\mu\right)\left(X_{s+|h_1|}-\mu\right),\\
&\quad\left(X_z-\mu\right)\left(X_{z+|h_2|}-\mu\right)\left(X_q-\mu\right)\left(X_{q+|h_2|}-\mu\right)\Big).
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{n}{M_n}V_{3,n}=\frac{1}{M_nn^3}&\sum_{h_1,h_2=-(n-1)}^{n-1}w\left(\frac{h_1}{M_n}\right)w\left(\frac{h_2}{M_n}\right)\sum_{t,s=1}^{n-|h_1|}\sum_{z,q=1}^{n-|h_2|}Cov\Big(\left(X_t-\mu\right)\left(X_{t+|h_1|}-\mu\right)\\
&\left(X_s-\mu\right)\left(X_{s+|h_1|}-\mu\right),\left(X_z-\mu\right)\left(X_{z+|h_2|}-\mu\right)\left(X_q-\mu\right)\left(X_{q+|h_2|}-\mu\right)\Big).
\end{aligned}\tag{2.48}$$

To bound (2.48) we evaluate the expectation $E\left(\prod_{i=1}^{8}Z_i\right)$ of the random variables Z_i with $E\left(Z_i\right)=0$, $i=1,2,...,8$, using all possible decompositions in products of

$$E\left(Z_iZ_jZ_kZ_h\right)E\left(Z_lZ_mZ_nZ_r\right),$$

$$E\left(Z_iZ_jZ_kZ_hZ_lZ_m\right)E\left(Z_nZ_r\right),$$

$$E\left(Z_lZ_mZ_nZ_r\right)E\left(Z_iZ_j\right)E\left(Z_kZ_h\right),$$

$$E\left(Z_iZ_jZ_kZ_hZ_l\right)E\left(Z_mZ_nZ_r\right),$$

$$E\left(Z_iZ_jZ_k\right)E\left(Z_mZ_nZ_r\right)E\left(Z_hZ_l\right),$$

$$E\left(Z_iZ_j\right)E\left(Z_kZ_h\right)E\left(Z_lZ_m\right)E\left(Z_nZ_r\right),$$

for indices $i, j, k, l, m, n, r, h \in \{1, 2, ..., 8\}$ and the cumulant term $Cum\left(Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7, Z_8\right)$. Evaluating the expansion term in (2.48) in that way, it follows by similar arguments that all terms have at least order $O(1/M_n)$. For instance, consider the case where product of four autocovariances appears

$$\frac{1}{M_n n^3}\sum_{h_1,h_2=-(n-1)}^{n-1} w\left(\frac{h_1}{M_n}\right)w\left(\frac{h_2}{M_n}\right)(n-|h_2|)\sum_{t,s=1}^{n-|h_1|}\sum_{z=1}^{n-|h_2|}\gamma_X(h_2)\gamma_X(t-z+|h_1|)$$
$$\times\gamma_X(t-s-|h_1|)\,\gamma_X(s-z-|h_2|).$$

The above expression is bounded in absolute value by

$$\frac{1}{M_n}\sup_{-1\le x\le 1} w^2(x)\Big(\sum_{r=-\infty}^{\infty}|\gamma_X(r)|\Big)^4 = O\left(\frac{1}{M_n}\right).$$

Thus, from (2.36), (2.37) and (2.38) we get that

$$\frac{n}{M_n}Var(\widetilde{N}_n) = 2\Big(\sum_{u=-\infty}^{\infty}\gamma_{2,X}(u)\Big)^2\int_{-1}^{1} w^2(x)\,dx + O\left(\frac{1}{M_n}\right).$$

Lemma 2.7.2 (ii) and equation (2.29) imply then that

$$\frac{n}{M_n}Var\left(\breve{\eta}_{4,\varepsilon}\right) = \frac{8\pi^2 f^2_{X^2}(0)}{\gamma^4_X(0)}\int_{-1}^{1} w^2(x)\,dx + O\left(\frac{1}{\sqrt{M_n}} + \frac{M_n}{n}\right),$$

from which assertion (ii) of the theorem follows since

$$\sum_{h=-\infty}^{\infty}\gamma_{2,X}(h) = \eta_{4,\varepsilon}\gamma^2_X(0) + 2\sum_{h=-\infty}^{\infty}\gamma^2_X(h);\ \text{see (2.3)}.$$

(iii) Notice that because

$$\sqrt{\frac{n}{M_n}}(\widetilde{N}_n - E(\widetilde{N}_n)) = \sqrt{\frac{n}{M_n}}\sum_{h=-(n-1)}^{(n-1)} w\left(\frac{h}{M_n}\right)\left(\widetilde{\gamma}_{2,X}(h) - E\left(\widetilde{\gamma}_{2,X}(h)\right)\right) + O_P\left(\frac{1}{\sqrt{M_n}}\right), \tag{2.49}$$

it suffices to consider the limiting distribution of the first term on the right hand-side of (2.49) only. We have

$$\sqrt{\frac{n}{M_n}}\sum_{h=-(n-1)}^{(n-1)} w\left(\frac{h}{M_n}\right)\left(\widetilde{\gamma}_{2,X}(h) - E(\widetilde{\gamma}_{2,X}(h))\right)$$
$$= 2\sqrt{\frac{n}{M_n}}\sum_{h=1}^{M_n} w\left(\frac{h}{M_n}\right)\frac{1}{n}\sum_{t=1}^{n}\left\{U_tU_{t+h} - E\left(U_tU_{t+h}\right)\right\} + o_P(1)$$
$$= W_n + o_P(1)\ \text{(say)},$$

where

$$U_t = X_t^2 - \mu_2 = \sum_{j_1, j_2 = -\infty}^{\infty} \psi_{j_1} \psi_{j_2} \varepsilon_{t-j_1} \varepsilon_{t-j_2} - \sigma_\varepsilon^2 \sum_{j=-\infty}^{\infty} \psi_j^2.$$

Write

$$U_t = U_{t,L} + V_{t,L},$$

where

$$U_{t,L} = \sum_{j_1, j_2 = -L}^{L} \psi_{j_1} \psi_{j_2} \varepsilon_{t-j_1} \varepsilon_{t-j_2} - \sigma_\varepsilon^2 \sum_{j=-L}^{L} \psi_j^2,$$

and

$$V_{t,L} = \sum_{|j_1|, |j_2| > L} \psi_{j_1} \psi_{j_2} \varepsilon_{t-j_1} \varepsilon_{t-j_2} - \sigma_\varepsilon^2 \sum_{|j|>L} \psi_j^2 + 2 \sum_{-L \le j_1 \le L, |j_2| > L} \psi_{j_1} \psi_{j_2} \varepsilon_{t-j_1} \varepsilon_{t-j_2}.$$

Notice that $U_{t,L}$ depends only on a finite number of the i.i.d. innovations ε_t's. Let

$$W_{n,L} = 2\sqrt{\frac{n}{M_n}} \sum_{h=1}^{M_n} w\left(\frac{h}{M_n}\right) \frac{1}{n} \sum_{t=1}^{n} \{U_{t,L} U_{t+h,L} - E\left(U_{t,L} U_{t+h,L}\right)\}.$$

In order to show that

$$W_n \xrightarrow{D} N\Big(0,\ 8\pi^2 f_{X^2}^2(0) \int_{-1}^{1} w^2(x)\, dx\Big) \quad \text{as} \quad n \to \infty,$$

it suffices by Proposition 6.3.9 of Brockwell and Davis (1991), to show that

(a) $W_{n,L} \underset{n\to\infty}{\overset{D}{\rightarrow}} Y_L$ for all L, where

$$Y_L \sim N\Big(0, 8\pi^2 f_{X_L^2}^2(0) \int_{-1}^{1} w^2(x) dx\Big),$$

$f_{X_L^2}$ is the spectral density of the process $\mathbf{X}_L^2 = \{X_{t,L}^2, t \in \mathbb{Z}\}$ and $X_{t,L}^2 = (\sum_{j=-L}^{L} \psi_j \varepsilon_{t-j})^2$.

(b) $Y_L \xrightarrow{D} Y$ as $L \to \infty$ where

$$Y \sim N\Big(0, 8\pi^2 f_{X^2}^2(0) \int_{-1}^{1} w^2(x)\, dx\Big),$$

(c) $\lim_{L\to\infty} \limsup_{n\to\infty} P\left(|W_n - W_{n,L}| > \varepsilon\right) = 0$ for every $\varepsilon > 0$.

To establish (a) notice that $E\left(W_{n,L}\right) = 0$, while

$$Var\left(W_{n,L}\right) = 8\pi^2 f_{X_L^2}^2(0) \int_{-1}^{1} w^2(x) + O\left(\frac{1}{M_n}\right).$$

Write,

$$W_{n,L} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_{t,L},$$

where

$$Z_{t,L} = 2\frac{1}{\sqrt{M_n}} \sum_{h=1}^{M_n} w\left(\frac{h}{M_n}\right) \{U_{t,L}U_{t+h,L} - E\left(U_{t,L}U_{t+h,L}\right)\}.$$

Let $\{Q_n\}$ be a sequence of integers such that $\left(M_n + 2L\right)/Q_n \to 0$ and $Q_n/n \to 0$ as $n \to \infty$. Furthermore, let

$$Q_n = \left[\sqrt{n\left(M_n + 2L\right)}\right]$$

and

$$Y_{j,n} = \frac{1}{\sqrt{Q_n}} \left\{Z_{(j-1)Q_n+1,L} + ... + Z_{jQ_n - M_n - 2L,L}\right\}, \quad j = 1, 2, ..., [n/Q_n].$$

Observe that $Y_{1,n}, Y_{2,n}, ..., Y_{[n/Q_n],n}$ are i.i.d. and that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_{t,L} - \frac{1}{\sqrt{[n/Q_n]}} \sum_{j=1}^{[n/Q_n]} Y_{j,n} = o_P\left(1\right).$$

Thus, it suffices to consider the asymptotic distribution of

$$\frac{1}{\sqrt{[n/Q_n]}} \sum_{j=1}^{[n/Q_n]} Y_{j,n}.$$

Moreover,

$$E\left({Y_{j,n}}^4\right) = \frac{1}{Q_n^2} \sum_{t,s,r,q=1}^{Q_n - M_n} E\left(Z_t Z_s Z_r Z_q\right), \tag{2.50}$$

where

$$E\left(Z_t Z_s Z_r Z_q\right) = \begin{cases} E\left(Z_t^4\right), & \text{if } t = s = r = q, \\ E\left(Z_t^2 Z_s^2\right), & \text{if } t = r \neq s = q \text{ or } t = s \neq r = q \text{ or } t = q \neq s = r, \\ E\left(Z_t^2 Z_s Z_r\right), & \text{if two indices are same, different with the other two indices} \\ & \quad \text{which are different with each other} \\ E\left(Z_t^3 Z_s\right), & \text{if } t = r = q \neq s \text{ or } s = r = q \neq t \text{ or } t = s = q \neq r \text{ or } t = r = s \neq q, \\ E\left(Z_t Z_s Z_r Z_q\right), & \text{if all indices are different.} \end{cases}$$

We proceed by evaluating the expectation term $E\left(Z_t Z_s Z_r Z_q\right)$. Now, consider for instance, the case where $t = s = r = q$. Then,

$$E\left(Y_{j,n}{}^4\right) = \frac{1}{Q_n^2}\sum_{t=1}^{Q_n-M_n} E\left(Z_t^4\right)$$

$$= \frac{2^4}{M_n^2 Q_n^2}\sum_{h_1,h_2,h_3,h_4=1}^{M_n} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_2}{M_n}\right) w\left(\frac{h_3}{M_n}\right) w\left(\frac{h_4}{M_n}\right)\sum_{t=1}^{Q_n-M_n} E\Bigg\{$$

$$\Big\{U_{t,L}U_{t+|h_1|,L} - E\left(U_{t,L}U_{t+|h_1|,L}\right)\Big\}\Big\{U_{t,L}U_{t+|h_2|,L} - E\left(U_{t,L}U_{t+|h_2|,L}\right)\Big\}$$

$$\times\Big\{U_{t,L}U_{t+|h_3|,L} - E\left(U_{t,L}U_{t+|h_3|,L}\right)\Big\}\Big\{U_{t,L}U_{t+|h_4|,L} - E\left(U_{t,L}U_{t+|h_4|,L}\right)\Big\}\Bigg\}.$$

The last expression above equals

$$\frac{2^4}{M_n^2 Q_n^2}\sum_{h_1,h_2,h_3,h_4=1}^{M_n} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h}{M_n}\right) w\left(\frac{h}{M_n}\right) w\left(\frac{h}{M_n}\right)\sum_{t=1}^{Q_n-M_n} E(W_{t,h_1}W_{t,h_2}W_{t,h_3}W_{t,h_4})$$

$$= \frac{2^4}{M_n^2 Q_n^2}\sum_{h_1,h_2,h_3,h_4=1}^{M_n} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_2}{M_n}\right) w\left(\frac{h_3}{M_n}\right) w\left(\frac{h_4}{M_n}\right)\sum_{t=1}^{Q_n-M_n}$$

$$\big\{E\left(W_{t,h_1}W_{t,h_2}\right)E\left(W_{t,h_3}W_{t,h_4}\right) + E\left(W_{t,h_1}W_{t,h_3}\right)E\left(W_{t,h_2}W_{t,h_4}\right)$$

$$+ E\left(W_{t,h_1}W_{t,h_4}\right)E\left(W_{t,h_2}W_{t,h_3}\right) + Cum\left(W_{t,h_1}, W_{t,h_2}, W_{t,h_3}, W_{t,h_4}\right)\big\}, \tag{2.51}$$

where $W_{t,h} = U_{t,L}U_{t+|h|,L} - E\left(U_{t,L}U_{t+|h|,L}\right)$. Denote by

$$cum_{U,L}\left(h_1, h_2, h_3\right) \equiv Cum\left(U_{t,L}, U_{t+h_1,L}, U_{t+h_2,L}, U_{t+h_3,L}\right)$$

the fourth order joint cumulant of $U_{t,L}, U_{t+h_1,L}, U_{t+h_2,L}$ and $U_{t+h_3,L}$ and let $\gamma_{U,L}\left(h_1\right) = Cov\left(U_{t,L}, U_{t+h_1,L}\right)$, be the autocovariance function of $\{U_{t,L}, t \in \mathbb{Z}\}$. From Lemma 2.7.1 we get that

$$\sum_{h_1=-\infty}^{\infty}\sum_{h_2=-\infty}^{\infty}\sum_{h_3=-\infty}^{\infty} cum_{U,L}\left(h_1, h_2, h_3\right) < \infty. \tag{2.52}$$

The first term of (2.51) which refers to $E\left(W_{t,h_1}W_{t,h_2}\right)E\left(W_{t,h_3}W_{t,h_4}\right)$ equals

$$\frac{2^4(Q_n - M_n)}{M_n^2 Q_n^2}\sum_{h_1,h_2,h_3,h_4=1}^{M_n} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_2}{M_n}\right) w\left(\frac{h_3}{M_n}\right) w\left(\frac{h_4}{M_n}\right)\Bigg\{$$

$$\Big\{\gamma_{U,L}\left(0\right)\gamma_{U,L}\left(|h_2| - |h_1|\right) + \gamma_{U,L}\left(h_1\right)\gamma_{U,L}\left(h_2\right) + cum_{U,L}\left(|h_1|, 0, |h_2|\right)\Big\}$$

$$\times\Big\{\gamma_{U,L}\left(0\right)\gamma_{U,L}\left(|h_4| - |h_3|\right) + \gamma_{U,L}\left(h_3\right)\gamma_U, L\left(h_4\right) + cum_{U,L}\left(|h_3|, 0, |h_4|\right)\Big\}\Bigg\}. \tag{2.53}$$

To show that (2.53) converges to zero, consider for instance the term

$$\frac{2^4\gamma_{U,L}^2\left(0\right)\left(Q_n - M_n\right)}{M_n^2 Q_n^2}\sum_{h_1,h_2=1}^{M_n} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_2}{M_n}\right)\gamma_{U,L}\left(|h_2| - |h_1|\right)$$

$$\times\sum_{h_3,h_4=1}^{M_n} w\left(\frac{h_3}{M_n}\right) w\left(\frac{h_4}{M_n}\right)\gamma_{U,L}\left(|h_4| - |h_3|\right).$$

The above expression is bounded in absolute value by

$$\frac{2^4\gamma_{U,L}^2(0)(Q_n+M_n)}{M_n^2Q_n^2}\sum_{h_1,h_2=1}^{M_n}|\gamma_{U,L}(|h_2|-|h_1|)|\sum_{h_3,h_4=1}^{M_n}|\gamma_{U,L}(|h_4|-|h_3|)|$$

$$=\frac{2^4\gamma_{U,L}^2(0)(Q_n+M_n)}{Q_n^2}\sum_{r=-(M_n-1)}^{(M_n-1)}|\gamma_{U,L}(r)|\phi_{M_n}(r)\sum_{z=-(M_n-1)}^{(M_n-1)}|\gamma_{U,L}(z)|\phi_{M_n}(z), \quad (2.54)$$

where

$$M_n\phi_{M_n}(r)=\begin{cases} M_n+r, & r=-M_n,...,0\\ M_n, & r=0,\\ M_n-r, & r=0,...,M_n,\\ 0, & \text{elsewhere.}\end{cases}$$

Note that for every r,

$$0\le\phi_{M_n}(r)\le 1, \quad \text{and} \quad \lim_{n\to\infty}\phi_{M_n}(r)=1.$$

Thus, equation (2.54), is bounded by

$$\frac{2^4\gamma_{U,L}^2(0)(Q_n+M_n)}{Q_n^2}\Big(\sum_{r=-\infty}^{\infty}|\gamma_{U,L}(r)|\Big)^2\to 0 \quad \text{as} \quad n\to\infty.$$

Consider next from (2.53), the term

$$\frac{2^4\gamma_{U,L}(0)(Q_n-M_n)}{M_n^2Q_n^2}\sum_{h_1,h_2=1}^{M_n}w\left(\frac{h_1}{M_n}\right)w\left(\frac{h_2}{M_n}\right)\gamma_{U,L}(|h_2|-|h_1|)$$
$$\times\sum_{h_3,h_4=1}^{M_n}w\left(\frac{h_3}{M_n}\right)w\left(\frac{h_4}{M_n}\right)cum_{U,L}(|h_3|,0,|h_4|).$$

The above expression is bounded in absolute value by

$$\frac{2^4|\gamma_{U,L}(0)|(Q_n+M_n)}{M_n^2Q_n^2}\sum_{h_1,h_2=1}^{M_n}|\gamma_{U,L}(|h_2|-|h_1|)|\sum_{h_3,h_4=1}^{M_n}|cum_{U,L}(|h_3|,0,|h_4|)|$$
$$=\frac{2^4|\gamma_{U,L}(0)|(Q_n+M_n)}{M_nQ_n^2}\sum_{r=-(M_n-1)}^{(M_n-1)}|\gamma_{U,L}(r)|\varphi_{M_n}(r)\sum_{h_3,h_4=1}^{M_n}|cum_{U,L}(|h_3|,0,|h_4|)|$$
$$\le\frac{2^4|\gamma_{U,L}(0)|(Q_n+M_n)}{M_nQ_n^2}\sum_{r=-\infty}^{\infty}|\gamma_{U,L}(r)|\sum_{h_3,h_4=-\infty}^{\infty}|cum_{U,L}(|h_3|,0,|h_4|)|$$
$$\to 0 \quad \text{as} \quad n\to\infty.$$

By similar arguments it follows that all the other terms of (2.53) as well as the second and the third term of (2.51) converge to zero as $n\to\infty$. Finally, the fourth term of

(2.51) is bounded by

$$\frac{2^4(Q_n+M_n)}{M_n^2Q_n^2}\sum_{h_1,h_2,h_3,h_4=1}^{M_n}\left|Cum(U_{t,L}U_{t+|h_1|,L},U_{t,L}U_{t+|h_2|,L},U_{t,L}U_{t+|h_3|,L},U_{t,L}U_{t+|h_4|,L}\right|,$$

which converges to zero as $n\to\infty$ because of Lemma 2.7.1. Next, consider $E\left(Y_{j,n}{}^4\right)$ in (2.50), when all indices t,s,q and r are different. Using the same arguments as in dealing with the case $t=s=q=r$ we get

$$\begin{aligned}E\left(Y_{j,n}{}^4\right)=\frac{2^4}{M_n^2Q_n^2}&\sum_{h_1,h_2,h_3,h_4=1}^{M_n}w\left(\frac{h_1}{M_n}\right)w\left(\frac{h_2}{M_n}\right)w\left(\frac{h_3}{M_n}\right)w\left(\frac{h_4}{M_n}\right)\sum_{t,r,s,q=1}^{Q_n-M_n}\\&\big\{E\left(W_{t,h_1}W_{r,h_2}\right)E\left(W_{s,h_3}W_{q,h_4}\right)+E\left(W_{t,h_1}W_{q,h_3}\right)E\left(W_{s,h_2}W_{r,h_4}\right)\\&+E\left(W_{t,h_1}W_{s,h_4}\right)E\left(W_{q,h_2}W_{r,h_3}\right)+Cum\left(W_{t,h_1},W_{s,h_2},W_{q,h_3},W_{r,h_4}\right)\big\}.\end{aligned}\tag{2.55}$$

The first term of (2.55) which refers to $E\left(W_{t,h_1}W_{r,h_2}\right)E\left(W_{s,h_3}W_{q,h_4}\right)$ equals

$$\begin{aligned}&\frac{2^4}{M_n^2Q_n^2}\sum_{h_1,h_2,h_3,h_4=1}^{M_n}w\left(\frac{h_1}{M_n}\right)w\left(\frac{h_2}{M_n}\right)w\left(\frac{h_3}{M_n}\right)w\left(\frac{h_4}{M_n}\right)\Bigg\{\\&\Big\{\gamma_{U,L}\left(r-t\right)\gamma_{U,L}\left(r-t+|h_2|-|h_1|\right)+\gamma_{U,L}\left(r-t+h_1\right)\gamma_{U,L}\left(r-t+h_2\right)\\&\quad+cum_{U,L}\left(|h_1|,r-t,r-t+|h_2|\right)\Big\}\\&\times\Big\{\gamma_{U,L}\left(s-q\right)\gamma_{U,L}\left(s-q+|h_4|-|h_3|\right)+\gamma_{U,L}\left(s-q+h_3\right)\gamma_{U,L}\left(s-q+h_4\right)\\&\quad+cum_{U,L}\left(|h_3|,s-q,s-q+|h_4|\right)\Big\}\Bigg\},\end{aligned}\tag{2.56}$$

which is bounded. To see this, consider for instance the term

$$\begin{aligned}&\frac{2^4}{M_n^2Q_n^2}\sum_{t,s,r,q=1}^{Q_n-M_n}\gamma_{U,L}\left(r-t\right)\gamma_{U,L}\left(s-q\right)\sum_{h_1,h_2=1}^{M_n}w\left(\frac{h_1}{M_n}\right)w\left(\frac{h_2}{M_n}\right)\gamma_{U,L}\left(r-t+|h_2|-|h_1|\right)\\&\times\sum_{h_3,h_4=1}^{M_n}w\left(\frac{h_3}{M_n}\right)w\left(\frac{h_4}{M_n}\right)\gamma_{U,L}\left(s-q+|h_4|-|h_3|\right).\end{aligned}$$

The above expression is bounded in absolute value by

$$\begin{aligned}&\frac{2^4}{M_n^2Q_n^2}\sum_{t,s,r,q=1}^{Q_n-M_n}\left|\gamma_{U,L}\left(r-t\right)\right|\left|\gamma_{U,L}\left(s-q\right)\right|\sum_{h_1,h_2=1}^{M_n}\left|\gamma_{U,L}\left(r-t+|h_2|-|h_1|\right)\right|\\&\times\sum_{h_3,h_4=1}^{M_n}\left|\gamma_{U,L}\left(s-q+|h_4|-|h_3|\right)\right|,\end{aligned}$$

which is equals to

$$\frac{2^4(Q_n-M_n)^2}{Q_n^2}\sum_{l,m=-(Q_n-M_n-1)}^{(Q_n-M_n-1)}\phi_{Q_n-M_n}(l)\,\phi_{Q_n-M_n}(m)\,|\gamma_{U,L}(l)\,||\gamma_{U,L}(m)\,|$$
$$\times\sum_{h=-(M_n-1)}^{(M_n-1)}|\gamma_{U,L}(l+h)\,|\sum_{z=-(M_n-1)}^{(M_n-1)}|\gamma_{U,L}(m+z)\,|$$
$$\leq\frac{2^4(Q_n-M_n)^2}{Q_n^2}\Big(\sum_{h=-\infty}^{\infty}|\gamma_{U,L}(h)\,|\Big)^4$$
$$<\infty.$$

By similar arguments and using relation (2.52) it follows that all other terms of (2.55) as well as all terms of (2.56) are negligible. Thus, $E\left(Y_{1,n}^4\right)$ is uniformly bounded in n, which, by a verification of Lyapunov's condition, implies that

$$\frac{1}{\sqrt{[n/Q_n]}}\sum_{j=1}^{[n/Q_n]}Y_{j,n}$$

has a limiting Gaussian distribution, that is, as $n\to\infty$,

$$W_{n,L}=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}Z_{t,L}\xrightarrow{D}Y_L,\quad\text{where}\quad Y_L\sim N\Big(0,8\pi^2f_{X_L^2}^2(0)\int_{-1}^{1}w^2(x)\,dx\Big).$$

To see assertion (b) notice first that

$$f_{X^2}^2(0)-f_{X_L^2}^2(0)=\Big(\frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\gamma_{2,X}(h)\Big)^2-\Big(\frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\gamma_{2,X_L}(h)\Big)^2,$$

where γ_{2,X_L} is the autocovariance function of the process $\mathbf{X}_L^2=\{X_{t,L}^2,t\in\mathbb{Z}\}$. Using $a^2-b^2=(a-b)(a+b)$ and since

$$\sum_{h=-\infty}^{\infty}\gamma_{2,X}(h)=\big(E\left(\varepsilon_1^4\right)-3\sigma_\varepsilon^4\big)\Big(\sum_{j=-\infty}^{\infty}\psi_j^2\Big)^2+2\sigma_\varepsilon^4\sum_{h=-\infty}^{\infty}\Big(\sum_{j=-\infty}^{\infty}\psi_j\psi_{j+h}\Big)^2,$$

and

$$\sum_{h=-\infty}^{\infty}\gamma_{2,X_L}(h)=\big(E\left(\varepsilon_1^4\right)-3\sigma_\varepsilon^4\big)\Big(\sum_{j=-L}^{L}\psi_j^2\Big)^2+2\sigma_\varepsilon^4\sum_{h=-\infty}^{\infty}\Big(\sum_{j=-L}^{L}\psi_j\psi_{j+h}\Big)^2,$$

we easily get

$$\Big|f_{X^2}^2(0)-f_{X_L^2}^2(0)\Big|\to 0\quad\text{as}\quad L\to\infty.$$

Consider assertion (c). Write

$$W_n-W_{n,L}=S_1+S_2+S_3$$

where

$$S_1 = \frac{2}{\sqrt{nM_n}} \sum_{h=1}^{M_n} w\left(\frac{h}{M_n}\right) \sum_{t=1}^{n} \Big\{U_{t,L}V_{t+h,L} - E\left(U_{t,L}V_{t+h,L}\right)\Big\},$$

$$S_2 = \frac{2}{\sqrt{nM_n}} \sum_{h=1}^{M_n} w\left(\frac{h}{M_n}\right) \sum_{t=1}^{n} \Big\{V_{t,L}U_{t+h,L} - E\left(V_{t,L}U_{t+h,L}\right)\Big\},$$

and

$$S_3 = \frac{2}{\sqrt{nM_n}} \sum_{h=1}^{M_n} w\left(\frac{h}{M_n}\right) \sum_{t=1}^{n} \Big\{V_{t,L}V_{t+h,L} - E\left(V_{t,L}V_{t+h,L}\right)\Big\}.$$

Straightforward calculations yield that the variances of the terms S_i, $i = 1, 2, 3$, are of order $O\left(\sum_{|j|>L} |\psi_j|\right)$ uniformly in n. Since

$$E(W_n - W_{n,L})^2 \leq 3\left(Var\left(S_1\right) + Var\left(S_2\right) + Var\left(S_3\right)\right),$$

assertion (c) follows by Markov's inequality.

To conclude the proof notice that

$$\sqrt{\frac{n}{M_n}}(\widetilde{\eta}_{4,\varepsilon} - E(\widetilde{\eta}_{4,\varepsilon})) = \sqrt{\frac{n}{M_n}}\frac{1}{\gamma_X^2(0)}(\widetilde{N}_n - E(\widetilde{N}_n)) + o_P\left(1\right),$$

and, therefore, using Lemma 2.7.2 (iii) and the relation

$$\frac{8\pi^2 f_{X^2}^2(0)}{\gamma_X^4(0)} = 2\Big(\eta_{4,\varepsilon} + 2\sum_{h=-\infty}^{\infty} \rho_X^2(h)\Big)^2,$$

we get

$$\sqrt{\frac{n}{M_n}}(\breve{\eta}_{4,\varepsilon} - E(\breve{\eta}_{4,\varepsilon})) \xrightarrow{D} N\left(0, \tau_X^2\right), \text{ as } n \to \infty$$

where

$$\tau_X^2 = 2\Big(\eta_{4,\varepsilon} + 2\sum_{h=-\infty}^{\infty} \rho_X^2(h)\Big)^2 \int_{-1}^{1} w^2(x)dx.$$

□

Let $\mathring{\eta}_{4,\varepsilon}$ be the same estimator as $\widehat{\eta}_{4,\varepsilon}$ with the estimated filtered observations $\widehat{U}_{t,p} = X_t - \sum_{j=1}^{p} \widehat{a}_{j,p} X_{t-j}$, replaced by $U_{t,p} = X_t - \sum_{j=1}^{p} a_{j,p} X_{t-j}$, $t = p+1, p+2, \ldots, n$. The following lemma allows to approximate bias, variance and distribution of $\widehat{\eta}_{4,\varepsilon}$ using the corresponding quantities of $\mathring{\eta}_{4,\varepsilon}$.

Lemma 2.7.3. *Suppose that Assumption 2.1 and Assumption 2.2 are satisfied and let $p \in \mathbb{N}$ be fixed. Then, as $n \to \infty$*

(i) $E\left(\widehat{\eta}_{4,\varepsilon}\right) = E\left(\mathring{\eta}_{4,\varepsilon}\right) + O\left(\frac{M_n}{n} + \frac{1}{\sqrt{n}}\right)$,

(ii) $\frac{n}{M_n} Var\left(\widehat{\eta}_{4,\varepsilon}\right) = \frac{n}{M_n} Var\left(\mathring{\eta}_{4,\varepsilon}\right) + O\left(\frac{1}{\sqrt{M_n}} + \frac{M_n}{n}\right)$,

(iii) $\sqrt{\frac{n}{M_n}}\left(\widehat{\eta}_{4,\varepsilon} - E\left(\widehat{\eta}_{4,\varepsilon}\right)\right) = \sqrt{\frac{n}{M_n}}\left(\mathring{\eta}_{4,\varepsilon} - E\left(\mathring{\eta}_{4,\varepsilon}\right)\right) + O_P\left(\sqrt{\frac{M_n}{n}} + \frac{1}{\sqrt{M_n}}\right)$.

Proof:

Let "$\sim$" refer to estimators based on the time series $\widehat{U}_{t,p}$, $t = p+1, p+2, \ldots, n$ using the means $E(\widehat{U}_{t,p})$ and $E(\widehat{U}_{t,p}^2)$, instead of the sample means $\overline{U}_n$ and $\overline{U}_{2,n}$. Let also "$\mathring{\sim}$" refer to estimators based on the time series $U_{t,p}$, $t = p+1, p+2, \ldots, n$ using the means $E(U_{t,p})$ and $E(U_{t,p}^2)$, while, "$\circ$" to estimators based on the same time series but using the sample means $\overline{\overline{U}}_n = (1/N)\sum_{t=1}^{N} U_{t,p}$ and $\overline{\overline{U}}_{2,n} = (1/N)\sum_{t=1}^{N} U_{t,p}^2$. Notice that

$$\widehat{U}_{t,p} = U_{t,p} + \sum_{j=1}^{p} \left(a_{j,p} - \widehat{a}_{j,p}\right) X_{t-j}. \tag{2.57}$$

Since for fixed p, $E(a_{j,p} - \widehat{a}_{j,p})^2 = O_P(1/n)$, Cauchy-Schwartz's inequality yields, uniformly in h,

$$\sum_{j=1}^{p} \left(a_{j,p} - \widehat{a}_{j,p}\right) X_{t+|h|-j} = O_P\left(\frac{1}{\sqrt{n}}\right). \tag{2.58}$$

To show assertion (i), observe first that

$$\sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right) E(\widetilde{\gamma}_{2,U}(h) - 2\widetilde{\gamma}_U^2(h)) = \sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right) E(\mathring{\widetilde{\gamma}}_{2,U}(h) - 2\mathring{\widetilde{\gamma}}_U^2(h)) + O\left(\frac{M_n}{n}\right).$$

The above equation follows after straightforward calculations using relations (2.57), (2.58) and applying Cauchy-Schwartz's inequality. Using the same arguments as in the proof of Lemma 2.7.2 (i), it follows that,

$$\sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right) E(\widehat{\gamma}_{2,U}(h) - 2\widehat{\gamma}_U^2(h)) = \sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right) E(\widetilde{\gamma}_{2,U}(h) - 2\widetilde{\gamma}_U^2(h)) + O\left(\frac{M_n}{N}\right) \tag{2.59}$$

and

$$\sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right) E(\mathring{\widetilde{\gamma}}_{2,U}(h) - 2\mathring{\widetilde{\gamma}}_U^2(h)) = \sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right) E(\mathring{\gamma}_{2,U}(h) - 2\mathring{\gamma}_U^2(h)) + O\left(\frac{M_n}{N}\right).$$

The above relations and a Taylor series argument; see equation (2.25) yield assertion (i).

To prove assertion (ii) we first show that

$$\begin{aligned} &\frac{n}{M_n} Var\Big(\sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right)(\widetilde{\gamma}_{2,U}(h) - 2\widetilde{\gamma}_U^2(h))\Big) \\ &= \frac{n}{M_n} Var\Big(\sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right)(\mathring{\widetilde{\gamma}}_{2,U}(h) - 2\mathring{\widetilde{\gamma}}_U^2(h))\Big) + O\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \tag{2.60}$$

Equation (2.60) follows because, by straightforward calculations using Cauchy-Schwartz's inequality and relations (2.57) and (2.58) we get

$$\begin{aligned} &\frac{n}{M_n} \sum_{h_1,h_2=-M_n}^{M_n} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_1}{M_n}\right) Cov(\widetilde{\gamma}_U^2(h_1), \widetilde{\gamma}_U^2(h_2)) \\ &= \frac{n}{M_n} \sum_{h_1,h_2=-M_n}^{M_n} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_2}{M_n}\right) Cov(\mathring{\widetilde{\gamma}}_U^2(h_1), \mathring{\widetilde{\gamma}}_U^2(h_2)) + O\left(\frac{1}{n^{3/2} M_n}\right), \end{aligned}$$

$$\begin{aligned} &\frac{n}{M_n} \sum_{h_1,h_2=-M_n}^{M_n} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_1}{M_n}\right) Cov(\widetilde{\gamma}_{2,U}(h_1), \widetilde{\gamma}_{2,U}(h_2)) \\ &= \frac{n}{M_n} \sum_{h_1,h_2=-M_n}^{M_n} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_2}{M_n}\right) Cov(\mathring{\widetilde{\gamma}}_{2,U}(h_1), \mathring{\widetilde{\gamma}}_{2,U}(h_2)) + O\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

and

$$\begin{aligned} &\frac{n}{M_n} \sum_{h_1,h_2=-M_n}^{M_n} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_1}{M_n}\right) Cov(\widetilde{\gamma}_{2,U}(h_1), \widetilde{\gamma}_U^2(h_2)) \\ &= \frac{n}{M_n} \sum_{h_1,h_2=-M_n}^{M_n} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_2}{M_n}\right) Cov(\mathring{\widetilde{\gamma}}_{2,U}(h_1), \mathring{\widetilde{\gamma}}_U^2(h_2)) + O\left(\frac{1}{n}\right). \end{aligned}$$

Furthermore, using the same arguments as in the proof of Lemma 2.7.2 (ii) it follows that

$$
\begin{aligned}
&Var\Big(\sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right)\big(\widehat{\gamma}_{2,U}\,(h) - 2\widehat{\gamma}_U^2\,(h)\big)\Big) \\
&= Var\Big(\sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right)\big(\widetilde{\gamma}_{2,U}\,(h) - 2\widetilde{\gamma}_U^2\,(h)\big)\Big) + O\left(\frac{M_n^2}{N^2}\right)
\end{aligned}
\tag{2.61}
$$

and

$$
\begin{aligned}
&Var\Big(\sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right)\big(\mathring{\widetilde{\gamma}}_{2,U}\,(h) - 2\mathring{\widetilde{\gamma}}_U^2\,(h)\big)\Big) \\
&= Var\Big(\sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right)\big(\mathring{\gamma}_{2,U}\,(h) - 2\mathring{\gamma}_U^2\,(h)\big)\Big) + O\left(\frac{M_n^2}{N^2}\right).
\end{aligned}
$$

The above two relations, equation (2.60) and a Taylor series argument yield assertion (ii).

Assertion (iii) follows using assertion (i) and

$$
\sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right)\left(\widehat{\gamma}_{2,U}\,(h) - 2\widehat{\gamma}_U^2\,(h)\right) = \sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right)\left(\mathring{\gamma}_{2,U}\,(h) - 2\mathring{\gamma}_U^2\,(h)\right) + O_P\left(\frac{M_n}{n}\right).
$$

□

Proof of Theorem 2.3.1:

Follows from Lemma 2.7.2 and Lemma 2.7.3 applied to $\mathring{\eta}_{4,\varepsilon}$ and exactly along the same lines as the proof of Theorem 2.2.1. □

Proof of Theorem 2.4.1:

(i) It suffices to show that $p_0 \in \mathbb{N}$ exists such that for all $p \geq p_0$,

$$\sum_{h=1}^{\infty} \rho_U^2(h) \leq \sum_{h=1}^{\infty} \rho_X^2(h),$$

or equivalently, that for every $\varepsilon > 0$, $p_0 = p(\varepsilon, \mathbf{X})$ exists, such that for all $p \geq p_0$,

$$\sum_{h=1}^{\infty} \rho_U^2(h) < \varepsilon.$$

Using the AR-representation (2.8), we get

$$U_{t,p} = \sum_{j=1}^{p} (a_j - a_{j,p}) X_{t-j} + \sum_{j=p+1}^{\infty} a_j X_{t-j} + v_t. \tag{2.62}$$

Thus,

$$\begin{aligned}
\sum_{h=1}^{\infty} \gamma_U^2(h) = \sum_{h=1}^{\infty} \Bigg\{ & \sum_{j,r=1}^{p} (a_j - a_{j,p})(a_r - a_{r,p}) \gamma_X(h + j - r) \\
& + \sum_{j=1}^{p} \sum_{r=p+1}^{\infty} (a_j - a_{j,p}) a_r \gamma_X(h + j - r) \\
& + \sum_{j=p+1}^{\infty} \sum_{r=1}^{p} a_j (a_r - a_{r,p}) \gamma_X(h + j - r) + \sum_{j,r=p+1}^{\infty} a_j a_r \gamma_X(h + j - r) \\
& + \sum_{r=1}^{p} (a_r - a_{r,p}) \mathrm{Cov}(v_t, X_{t-r+h}) + \sum_{r=p+1}^{\infty} a_r \mathrm{Cov}(v_t, X_{t-r+h}) \\
& + \sum_{j=1}^{p} (a_j - a_{j,p}) \mathrm{Cov}(v_{t+h}, X_{t-j}) + \sum_{j=p+1}^{\infty} a_j \mathrm{Cov}(v_{t+h}, X_{t-j}) \Bigg\}^2,
\end{aligned} \tag{2.63}$$

where

$$\mathrm{Cov}(v_{t+h}, X_{t-j}) = \gamma_X(h + j) - \sum_{j_1=1}^{\infty} a_{j_1} \gamma_X(h + j - j_1)$$

and

$$\mathrm{Cov}(v_t, X_{t-r+h}) = \gamma_X(h - r) - \sum_{j_1=1}^{\infty} a_{j_1} \gamma_X(h + j_1 - r).$$

Bounding the differences $|a_{j,p} - a_j|$ by Baxter's inequality

$$\sum_{j=0}^{p} |a_{j,p} - a_j| \leq C \sum_{j=p+1}^{\infty} |a_j|, \tag{2.64}$$

where $C > 0$ is a constant independent of p, see Lemma 2.2 of Kreiss et al. (2011), and using the summability of the autocovariances and of the coefficients $|a_j|$, it follows that all eight terms on the right hand side of (2.63) can be made arbitrary small. Consider for instance, the term

$$\sum_{h=1}^{\infty}\Big\{\sum_{j,r=1}^{p}(a_j-a_{j,p})(a_r-a_{r,p})\gamma_X\left(h+j-r\right)\Big\}^2,$$

which is bounded by

$$\Big(\sum_{j,r=1}^{p}|a_j-a_{j,p}||a_r-a_{r,p}|\Big)^2\sum_{h=-\infty}^{\infty}\gamma_X^2\left(h\right). \tag{2.65}$$

Using (2.64) there exists $p^{(1)}\in\mathbb{N}$ and $C_1>0$ (independent of $p^{(1)}$) such that for all $p\geq p^{(1)}$, (2.65) is less or equal to

$$C_1\Big(\sum_{r=p+1}^{\infty}|a_r|\Big)^4\sum_{h=-\infty}^{\infty}\gamma_X^2\left(h\right)<\frac{\varepsilon\sigma_\varepsilon^4}{8}.$$

Similar arguments can be applied to the other seven terms on the right-hand side of (2.63) showing that $p^{(i)}$, $i=2,3,\ldots,8$ exist such that each one of the corresponding terms can be made arbitrary small, i.e., less than $\varepsilon\sigma_\varepsilon^4/8$. Choosing $p_0=\max\left\{p^{(1)},p^{(2)},...,p^{(8)}\right\}$, we get that

$$\sum_{h=1}^{\infty}\gamma_U^2\left(h\right)<\varepsilon\sigma_\varepsilon^4.$$

The assertion follows then since

$$\sum_{h=1}^{\infty}\rho_U^2\left(h\right)=\frac{1}{\gamma_U^2\left(0\right)}\sum_{h=1}^{\infty}\gamma_U^2\left(h\right)<\frac{\varepsilon\sigma_\varepsilon^4}{\gamma_U^2\left(0\right)}\leq\frac{\varepsilon\sigma_\varepsilon^4}{\sigma_\varepsilon^4}=\varepsilon.$$

(ii) From Theorem 2.1 of Hannan and Kavalieris (1986), we obtain under the assumptions made, that

$$\max_{1\leq j\leq p}|a_{j,p}-\widehat{a}_{j,p}|=O_P\Big(\sqrt{\frac{\log(n)}{n}}\Big),$$

and, therefore,

$$\sum_{j=1}^{p}\left(a_{j,p}-\widehat{a}_{j,p}\right)X_{t+|h|-j}=O_P\Big(p\sqrt{\frac{\log(n)}{n}}\Big). \tag{2.66}$$

We first prove that

$$\begin{aligned}&\frac{n}{M_n}Var\Big(\sum_{h=-M_n}^{M_n}w\left(\frac{h}{M_n}\right)\left(\widetilde{\gamma}_{2,U}\left(h\right)-2\widetilde{\gamma}_U^2\left(h\right)\right)\Big)\\&=\frac{n}{M_n}Var\Big(\sum_{h=-M_n}^{M_n}w\left(\frac{h}{M_n}\right)\left(\mathring{\widetilde{\gamma}}_{2,U}\left(h\right)-2\mathring{\widetilde{\gamma}}_U^2\left(h\right)\right)\Big)+O\Big(\frac{p\sqrt{n\log(n)}}{N}\Big).\end{aligned} \tag{2.67}$$

Equation (2.67) follows using the relations

$$\frac{n}{M_n}\sum_{h_1,h_2=-M_n}^{M_n} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_2}{M_n}\right) Cov(\widetilde{\gamma}_U^2(h_1), \widetilde{\gamma}_U^2(h_2))$$
$$= \frac{n}{M_n}\sum_{h_1,h_2=-M_n}^{M_n} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_2}{M_n}\right) Cov(\mathring{\widetilde{\gamma}}_U^2(h_1), \mathring{\widetilde{\gamma}}_U^2(h_2)) + O\Big(\frac{p\sqrt{n\log(n)}}{M_n N^2}\Big),$$

$$\frac{n}{M_n}\sum_{h_1,h_2=-M_n}^{M_n} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_2}{M_n}\right) Cov(\widetilde{\gamma}_{2,U}(h_1), \widetilde{\gamma}_{2,U}(h_2))$$
$$= \frac{n}{M_n}\sum_{h_1,h_2=-M_n}^{M_n} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_2}{M_n}\right) Cov(\mathring{\widetilde{\gamma}}_{2,U}(h_1), \mathring{\widetilde{\gamma}}_{2,U}(h_2)) + O\Big(\frac{p\sqrt{n\log(n)}}{N}\Big)$$

and

$$\frac{n}{M_n}\sum_{h_1,h_2=-M_n}^{M_n} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_2}{M_n}\right) Cov(\widetilde{\gamma}_{2,U}(h_1), \widetilde{\gamma}_U^2(h_2))$$
$$= \frac{n}{M_n}\sum_{h_1,h_2=-M_n}^{M_n} w\left(\frac{h_1}{M_n}\right) w\left(\frac{h_2}{M_n}\right) Cov(\mathring{\widetilde{\gamma}}_{2,U}(h_1), \mathring{\widetilde{\gamma}}_U^2(h_2)) + O\Big(\frac{p\sqrt{n\log(n)}}{N^{3/2}}\Big),$$

where the above assertions can be verified using Cauchy-Scwartz's inequality and equations (2.57) and (2.66). Using equations (2.61) and (2.67) and a Taylor series argument; see equation (2.25), we get

$$\frac{n}{M_n}Var(\widehat{\eta}_{4,\varepsilon}) = \frac{n}{M_n}\frac{1}{\gamma_U^4(0)}Var\Big(\sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right)(\mathring{\widetilde{\gamma}}_{2,U}(h) - 2\mathring{\widetilde{\gamma}}_U^2(h))\Big)$$
$$+ O\Big(\frac{p\sqrt{n\log(n)}}{N}\Big).$$

Next, by the same arguments as in the proof of Theorem 2.2.1(ii), we obtain

$$\frac{n}{M_n}Var(\widehat{\eta}_{4,\varepsilon}) = 2\Big(\eta_{4,\varepsilon} + 2\sum_{h=-m}^{m}\rho_U^2(h)\Big)^2\frac{1}{M_n}\sum_{r=-M_n}^{M_n} w^2\left(\frac{r}{M_n}\right) + O\Big(\frac{p\sqrt{n\log(n)}}{N}\Big),$$

where $m \leq M_n$ is an integer. Furthermore, since $\sum_{h=1}^{\infty}\gamma_U^2(h) \to 0$ as $p \to \infty$, we get

$$\sum_{h=-m}^{m}\rho_U^2(h) \leq 2\sum_{h=1}^{\infty}\rho_U^2(h) + 1 \leq \frac{2}{\sigma_\varepsilon^4}\sum_{h=1}^{\infty}\gamma_U^2(h) + 1 \to 1, \text{ as } p \to \infty,$$

and, therefore,

$$\frac{n}{M_n}Var(\widehat{\eta}_{4,\varepsilon}) \to 2(\eta_{4,\varepsilon}+2)^2\int_{-1}^{1} w^2(x)\,dx.$$

□

Proof of Theorem 2.4.2:

(i) It suffices to show that $p_0 \in \mathbb{N}$ exists such that for all $p \geq p_0$,

$$\frac{\left|2\pi f''_{\widetilde{U}}(0)\right|}{\gamma^2_U(0)} \leq \frac{\left|2\pi f''_{\widetilde{X}}(0)\right|}{\gamma^2_X(0)},$$

or equivalently, that for every $\varepsilon > 0$, $p_0 = p(\varepsilon, \mathbf{X})$ exists such that for all $p \geq p_0$,

$$\frac{\left|2\pi f''_{\widetilde{U}}(0)\right|}{\gamma^2_U(0)} < \varepsilon.$$

Notice that

$$\begin{aligned}
2\pi\left|f''_{\widetilde{U}}(0)\right| &= \sigma^2_\varepsilon \sum_{h=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h^2 c^2_{l,p} c^2_{l+h,p} \\
&\leq \sigma^2_\varepsilon \Big(\sum_{l=-\infty}^{\infty} |c_{l,p}| \sum_{h=-\infty}^{\infty} |h||c_{l+h,p}| \Big)^2 \\
&\leq \sigma^2_\varepsilon \Big(\sum_{h=-\infty}^{\infty} |h||c_{h,p}| \Big)^4.
\end{aligned}$$

Thus it suffices to show that for every $\delta > 0$, $p_0(\delta, \mathbf{X})$ exists such that for all $p \geq p_0$,

$$\sum_{h=-\infty}^{\infty} |h||c_{h,p}| < \delta.$$

Recall that the coefficients $c_{h,p}$ are those appearing in the power series

$$C_p(z) = \sum_{h=-\infty}^{\infty} c_{h,p} z^h, \ z \in \mathbb{C},$$

where

$$C_p(z) = A_p(z)\Psi(z), \ A_p(z) = 1 - \sum_{h=1}^{p} a_{h,p} z^h \quad \text{and} \quad \Psi(z) = \sum_{h=0}^{\infty} \psi_h z^h,$$

that is,

$$c_{h,p} = \begin{cases} \psi_h - \sum_{k=1}^{\min\{h,p\}} a_{k,p}\psi_{h-k}\,, & h \geq 0, \\ 0, & h < 0, \end{cases}$$

where $\psi_0 = 1$. Furthermore, for

$$A(z) = 1 - \sum_{h=1}^{\infty} a_h z^h$$

we have by Assumption 2.4, that

$$A(z) = \Psi^{-1}(z), \quad \text{and} \quad \psi_h - \sum\nolimits_{k=1}^{\infty} a_k \psi_{h-k} = 0, \quad \text{for} \quad h \neq 0.$$

From the above we get

$$\sum_{h=1}^{\infty}|h|\,|c_{h,p}| = \sum_{h=1}^{\infty}|h|\left|\sum\nolimits_{k=1}^{\min\{h,p\}}(a_k - a_{k,p})\psi_{h-k} + \sum\nolimits_{k=\min\{h,p\}+1}^{\infty} a_k\psi_{h-k}\right|. \qquad (2.68)$$

By Baxter's inequality, see (2.64), there exists $p_0 \in \mathbb{N}$ and $C_1 > 0$ (independent of p_0) such that for all $p \geq p_0$, (2.68) is less or equal to

$$C_1\sum_{h=1}^{\infty}|h|\sum\nolimits_{k=\min\{h,p\}+1}^{\infty}|a_k|\,|\psi_{h-k}| < \delta.$$

Thus, for $\delta = \sigma_\varepsilon^{1/2}\varepsilon^{1/4}$ we have that

$$2\pi\left|f''_{\widetilde{U}}(0)\right| < \sigma_\varepsilon^4\varepsilon,$$

and, therefore,

$$\frac{2\pi|f''_{\widetilde{U}}(0)|}{\gamma_U^2(0)} < \frac{\varepsilon\sigma_\varepsilon^4}{\gamma_U^2(0)} \leq \frac{\varepsilon\sigma_\varepsilon^4}{\sigma_\varepsilon^4} = \varepsilon.$$

(ii) First, we show

$$\sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right)E(\widehat{\gamma}_{2,U}(h) - 2\widehat{\gamma}_U^2(h)) = \sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right)E(\mathring{\widetilde{\gamma}}_{2,U}(h) - 2\mathring{\widetilde{\gamma}}_U^2(h)) + O\Big(\frac{p\sqrt{\log(n)}}{\sqrt{nN}}\Big). \qquad (2.69)$$

Equation (2.69) follows from equation (2.59) and because

$$\sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right)E(\widetilde{\gamma}_{2,U}(h) - 2\widetilde{\gamma}_U^2(h)) = \sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right)E(\mathring{\widetilde{\gamma}}_{2,U}(h) - 2\mathring{\widetilde{\gamma}}_U^2(h)) + O\Big(\frac{p\sqrt{\log(n)}}{\sqrt{nN}}\Big).$$

Let,

$$\widehat{\eta}_{4,\varepsilon} = \frac{\widehat{\kappa}_{4,\varepsilon}}{\widehat{\gamma}_U^2(0)} \quad \text{where} \quad \widehat{\kappa}_{4,\varepsilon} = \sum_{h=-M_n}^{M_n} w\left(\frac{h}{M_n}\right)(\widehat{\gamma}_{2,U}(h) - 2\widehat{\gamma}_U^2(h)).$$

Using the same arguments as in the proof of Theorem 2.2.1(i) we get

$$M_n^2[E(\widehat{\kappa}_{4,\varepsilon}) - \kappa_{4,\varepsilon}] = -C_w\sum_{h=-\infty}^{\infty}h^2[\gamma_{2,U}(h) - 2\gamma_U^2(h)] + O\left(\frac{M_n^2}{\sqrt{N}}p\sqrt{\frac{\log(n)}{n}}\right).$$

Recall that under Assumption 2.4, relation (2.62) is true with the v_t's replaced by the i.i.d. innovations ε_t. Straightforward calculations yield then

$$\sum_{h=-\infty}^{\infty}h^2[\gamma_{2,U}(h) - 2\gamma_U^2(h)] = o(1),$$

and, consequently,

$$M_n^2[E(\widehat{\kappa}_{4,\varepsilon}) - \kappa_{4,\varepsilon}] \to 0.$$

Finally, using

$$\gamma_U^2(0) - \gamma_\varepsilon^2(0) = o(1),$$

and a Taylor series argument we get, under the assumptions made, that

$$M_n^2(E(\widehat{\eta}_{4,\varepsilon}) - \eta_{4,\varepsilon}) = \frac{M_n^2}{\gamma_\varepsilon^2(0)}\left\{E(\widehat{\kappa}_{4,\varepsilon}) - \kappa_{4,\varepsilon}\right\} + O\left(\frac{M_n^2}{N} + \frac{M_n^2}{\sqrt{N}}p\sqrt{\frac{\log(n)}{n}}\right) \to 0,$$

which concludes the proof. □

Chapter 3

Extending the Range of Validity of the Autoregressive (Sieve) Bootstrap

3.1 Introduction

The aim of this chapter is twofold. First, for linear processes **X**, we extend the range of validity of the AR-sieve bootstrap to important classes of statistics which include, for instance, sample autocovariances. This is achieved by appropriately modifying the way the pseudo-innovations used in this bootstrap algorithm are generated. Using some recent developments in nonparametric estimation of the fourth order moments of the (unobserved) i.i.d. innovations ε_t driving the linear process (1.8), described in Chapter 2, we propose an AR-sieve bootstrap procedure where the pseudo-innovations are not obtained by i.i.d. resampling from the empirical distribution of the estimated residuals but from some appropriate three point distribution. This three point distribution, delivers i.i.d. pseudo-innovations which imitate asymptotically correct also the rescaled fourth order moment cumulant of the true innovations ε_t, a quantity which is important for some statistics belonging to the class (1.7). We call this procedure the AR-sieve bootstrap with i.i.d. wild innovations and we show that, for the linear process class (1.8), this modification extends the range of validity of the AR-sieve to important statistics belonging to the class (1.7) and for which the classical AR-sieve fails.

However, for general processes and due to the retained i.i.d. structure of the generated pseudo-innovations, the range of validity of the modified AR-sieve bootstrap for the class (1.7) is essentially restricted to statistics that only depend on the second order structure of the process. To overcome this limitation we propose a new version of the AR-bootstrap. This version works by fitting an autoregressive model of order p, not to the observed time series $X_1, X_2, \ldots, X_n$ itself, but to the time series of transformed random variables $Y_1, Y_2, \ldots, Y_{n-m+1}$, where Y_t is given by

$$Y_t = g(X_t, X_{t+1}, \ldots, X_{t+m-1});$$

see expression (1.7). New pseudo-time series $Y_1^*, Y_2^*, \ldots, Y_{n-m+1}^*$ are generated using this autoregressive fit and pseudo-innovations obtained by means of a dependent wild bootstrap procedure, Shao (2010), applied to the estimated residuals

$$\widehat{V}_{t,p} = Y_t - \sum_{j=1}^{p} \widehat{b}_{j,p} Y_{t-j}, \quad t = p+1, p+2, ..., n-m+1.$$

Since the dependent wild bootstrap appropriately mimics the dependence structure of the filtered process

$$V_{t,p} = Y_t - \sum_{j=1}^{p} b_{j,p} Y_{t-j},$$

the autoregressive order p used in the autoregressive fit does not need to increase to infinity with n, in order for this bootstrap procedure to capture the entire autocovariance structure of the transformed process $\mathbf{Y} = \{Y_t, t \in \mathbb{Z}\}$. Notice that for general nonlinear functions $g(\cdot)$, like those in (1.7), it is in general not easy to derive properties of the spectral density of the transformed process $\mathbf{Y}$ based on properties of the process $\mathbf{X}$. In other words, it is not clear under what circumstances an autoregressive representation like (1.6) exists for the transformed process $\mathbf{Y}$. This makes the application of an AR-sieve bootstrap procedure to $\mathbf{Y}$ difficult to justify theoretically. However, fitting a fixed, pth-order linear autoregression, is always possible, provided that

$$p < n - m, \quad Var(Y_t) > 0 \quad \text{and} \quad Cov(Y_t, Y_{t+h}) \to 0, \quad \text{as} \quad h \to \infty;$$

see Brockwell and Davis (1991), Proposition 5.1.1. The later requirement is, however, satisfied by the conditions imposed on T_n in order for this statistic to have a proper limiting distribution; see Subsection 3.2.1. We show that the proposed AR-bootstrap with dependent wild innovations, is asymptotically valid for a wide range of weakly dependent processes and for the entire class of statistics (1.7). This asymptotic validity coincides with a good finite sample behavior, which is demonstrated by means of several numerical simulations. In these simulations, comparisons to some alternative bootstrap methods are also given. Notice that although the last discussed version of the AR-bootstrap applied to the transformed process $\mathbf{Y}$ is also valid for the cases for which the modified AR-sieve with i.i.d. wild innovations works, the later bootstrap procedure retains its attractivity due to its potential efficiency in cases where the underlying process is indeed linear. This justifies the consideration of both bootstrap modifications in this chapter.

There are many applications of the AR, respectively, of the AR-sieve bootstrap, in the econometric time series literature which use wild bootstrap procedures to generate the pseudo-innovations. These applications concerns mostly the case of (stationary or non-stationary) autoregressive processes (of finite or infinite order) with heteroskedastic innovations or innovations having infinite variance; see among others Hansen (2000), Goncalves and Kilian (2004), Concalves and Kilian (2007) and Cavaliere et al. (2013). However, the situation considered in this chapter is different. We do not deal with non-stationary or heteroscedastic processes and our wild bootstrap proposals are concerned with the limitations of the AR-sieve bootstrap caused by the fact that the standard resampling procedures applied to generate the pseudo-innovations do not correctly mimic the rescaled fourth order cumulant of the true innovations (in the linear process case) or the fourth order moment structure of the process (in the general process case). These limitations turn out to be important for many statistics of interest, like for instance those described by the class (1.7). Our proposals resolve the problems caused by these limitations and considerably extend the range of validity of autoregressive bootstrap procedures. Furthermore, for general processes, our bootstrap consistency results are established under quite minimal assumptions on the underlying process requiring essentially summability of second and fourth order cumulants, and therefore, avoiding mixing or any other type of weak dependence conditions.

The remaining of the chapter is organized as follows. In Section 3.2 the basic assumptions needed as well as a precise description of the modified AR-sieve bootstrap procedure are given. Validity of the modified AR-sieve procedure driven by appropriately i.i.d. wild generated innovations is then established. Section 3.3 describes the AR-bootstrap proposal applied to the transformed process **Y** and driven by dependent wild pseudo-innovations. It establishes the asymptotic validity of this bootstrap procedure for the entire class of statistics (1.7) and under quite general weak dependence assumptions on the underlying process **X**. A fully data driven procedure to select the parameters involved in both bootstrap procedures is described in Section 3.4. Extensive simulations are also presented in this section which investigate the finite sample behavior of both methods and compare their performances with that of the classical AR-sieve and of the block bootstrap. All technical proofs are deferred to Section 3.5 .

3.2 Autoregressive Sieve Bootstrap with i.i.d. Wild Innovations

3.2.1 Assumptions and preliminaries

Throughout this section we assume that the underlying process **X** is linear, that is, X_t is generated as in (1.8). Moreover, the following assumption is made.

Assumption 3.1. *The power series*

$$\Psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j, \quad z \in \mathbb{C},$$

satisfies $\Psi(z) \neq 0$ *for* $|z| = 1$*. The coefficients* ψ_j *fulfill the condition*

$$\sum_{j=-\infty}^{\infty} |j||\psi_j| < \infty$$

and the i.i.d. innovations in (1.8) have finite fourth moments, i.e., $E\varepsilon_t^4 < \infty$*.*

Notice that Assumption 3.1 implies that the linear process **X** is strictly stationary with mean zero and autocovariance function

$$\gamma_X(h) = E(X_t X_{t+h}), \quad h \in \mathbb{Z}.$$

Furthermore,

$$\sum_{h=1}^{\infty} h|\gamma_X(h)| < \infty,$$

and a spectral density

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_X(h) e^{-ih\omega}$$

of $\mathbf{X}$ exists, is differentiable and bounded away from zero everywhere in the interval $[0, \pi]$. The linear process considered obeys, therefore, the autoregressive representation (1.6), where the white noise sequence $\{e_t\}$ appearing in this representation is not necessarily identical to the sequence of i.i.d. innovations $\{\varepsilon_t\}$ appearing in (1.8).

To illustrate the last point, consider as an example the simple linear process

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1} \quad \text{with} \quad \theta > 1.$$

With the help of the backshift operator L we can express X_t as

$$\begin{aligned} X_t &= (1 + \theta L)\varepsilon_t \\ &= (1 + \theta L)(1 + \theta^{-1} L)(1 + \theta^{-1} L)^{-1} \varepsilon_t \\ &= (1 + \theta^{-1} L) e_t, \end{aligned}$$

where

$$e_t = (1 + \theta L)(1 + \theta^{-1} L)^{-1} \varepsilon_t = \varepsilon_t + \sum_{j=1}^{\infty} d_j \varepsilon_{t-j}, \quad \text{with} \quad d_j = (1 - \theta^2)(-\theta)^{-j}.$$

Notice that $\{e_t, t \in \mathbb{Z}\}$ is a white noise process, that is, $E\left(e_t e_{t+h}\right) = 0 \; \forall \; h \geq 1$. This is easily seen, since $\forall \; h \geq 1$,

$$\begin{aligned} E\left(e_t e_{t+h}\right) &= E\left(\varepsilon_t \varepsilon_{t+h}\right) + \sum_{j=1}^{\infty} d_j E\left(\varepsilon_t \varepsilon_{t+h-j}\right) + \sum_{j=1}^{\infty} d_j E\left(\varepsilon_{t+h} \varepsilon_{t-j}\right) + \sum_{j,r=1}^{\infty} d_j d_r E\left(\varepsilon_{t-j} \varepsilon_{t+h-r}\right) \\ &= \sigma_\varepsilon^2 d_h + \sigma_\varepsilon^2 \sum_{j=1}^{\infty} d_j d_{j+h} \\ &= \sigma_\varepsilon^2 \left(1 - \vartheta^2\right) \left(-\frac{1}{\theta}\right)^h + \sigma_\varepsilon^2 (1 - \vartheta^2)^2 \left(-\frac{1}{\theta}\right)^h \sum_{j=1}^{\infty} \left(-\frac{1}{\theta}\right)^{2j} \\ &= 0. \end{aligned}$$

Thus, X_t obeys the AR-representation

$$X_t = \sum_{j=1}^{\infty} a_j X_{t-j} + e_t \quad \text{where} \quad a_j = -(-1/\theta)^j, \quad j = 1, 2, \ldots.$$

Furthermore, for $\theta = 2$ it is easily seen that

$$\sigma_e^2 = 4\sigma_\varepsilon^2 \quad \text{and} \quad \eta_{4,e} = \frac{2}{5}\eta_{4,\varepsilon},$$

where $\sigma_e^2 = E(e_t^2)$, $\sigma_\varepsilon^2 = E(\varepsilon_t^2)$, are the variances and $\eta_{4,e} = Ee_t^4/\sigma_e^4 - 3$ and $\eta_{4,\varepsilon} = E\varepsilon_t^4/\sigma_\varepsilon^4 - 3$, are the rescaled fourth order cumulants of the innovations e_t and ε_t, respectively.

Recall from Chapter 2 that fitting an autoregressive model of order p to $\mathbf{X}$ by means of minimizing the mean square error

$$E(X_t - \sum_{j=1}^{p} \beta_j X_{t-j})^2$$

with respect to $\beta_1, \beta_2, ..., \beta_p$, leads to the uniquely determined coefficients $a_p = (a_{1,p}, a_{2,p}, \ldots, a_{p,p})^\top$ given by

$$a_p = \Gamma_p^{-1}\gamma_p,$$

where

$$\Gamma_p = (\gamma_X(i-j))_{i,j=1,2,\ldots,p} \quad \text{and} \quad \gamma_p = (\gamma_X(j), j = 1, 2, \ldots, p)^\top.$$

Notice that under Assumption 3.1, the matrix Γ_p is invertible for every $p \in \mathbb{N}$; cf. Proposition 5.1.1 of Brockwell and Davis (1991). Suppose we have estimators $\widehat{a}_p = (\widehat{a}_{j,p}, j = 1, 2, \ldots, p)^\top$ of a_p and define the estimated residuals

$$\widehat{e}_{t,p} = X_t - \sum_{j=1}^{p} \widehat{a}_{j,p} X_{t-j}, \quad t = p+1, p+2, ..., n. \tag{3.1}$$

$\widehat{a}_p$ could be for instance the Yule-Walker estimator which is obtained by replacing $\gamma_X(h)$ in $a_p = \Gamma_p^{-1}\gamma_p$ by the sample autocovariances $\widehat{\gamma}_X(h)$, $0 \leq h \leq p$, given by

$$\widehat{\gamma}_X(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(X_t - \overline{X}_n)(X_{t+|h|} - \overline{X}_n), \quad \overline{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t. \tag{3.2}$$

Since we do not restrict our considerations to the case of the Yule-Walker estimator, we require that the estimator used satisfy the following condition.

Assumption 3.2. *The sequence of estimators* $\widehat{a}_p = (\widehat{a}_{1,p}, \widehat{a}_{2,p}, \ldots, \widehat{a}_{p,p})^\top$ *satisfies*

$$p^2 \sum_{j=1}^{p} |\widehat{a}_{j,p} - a_{j,p}| = O_P(1), \quad as \quad p \to \infty.$$

Assumption 3.2 is quite general and is fulfilled for instance by the Yule-Walker and the least squares estimator; see also Kreiss et al. (2011). For instance, if $\widehat{a}_{j,p}$ is the Yule-Walker estimator then,

$$\max_{1 \le j \le p} |a_{j,p} - \widehat{a}_{j,p}| = O_P\left(\sqrt{\frac{\log(n)}{n}}\right);$$

see Theorem 2.1 of Hannan and Kavalieris (1986). Thus,

$$p^2 \sum_{j=1}^{p} |\widehat{a}_{j,p} - a_{j,p}| = O_P\left(p^3 \sqrt{\frac{\log(n)}{n}}\right),$$

which is $O_P(1)$ if $p \to \infty$ with $n \to \infty$ such that $p = O(n/\log(n))^{1/6}$.

Consider next the class of statistics (1.7) and assume that the functions f and g satisfy the following smoothness conditions.

Assumption 3.3. *$f(x)$ has continuous partial derivative for all x in a neighborhood of $\theta = E(g(X_t, ..., X_{t+m-1}))$ and the differential*

$$\sum_{i=1}^{m} \partial f(x) / \partial x_i |_{x=\theta}$$

does not vanish. The function

$$g : \mathbb{R}^m \to \mathbb{R}^d, \ d \le m,$$

has continuous partial derivatives of order h $(h \ge 1)$ which satisfy a Lipschitz condition.

Under Assumption 3.1 and Assumption 3.3 it can be shown that, as $n \to \infty$,

$$\sqrt{n}(T_n - f(\theta)) \xrightarrow{D} N(0, H_f(\theta)\Sigma_g(\theta)H_f^{\top}(\theta)), \tag{3.3}$$

where "$\xrightarrow{D}$" denotes convergence in distribution,

$$\Sigma_g(\theta) = \left(\sum_{h=-\infty}^{\infty} Cov(g_i(X_0, ..., X_{m-1}), g_j(X_h, ..., X_{h+m-1}))\right)_{i,j=1,2,...,d}, \tag{3.4}$$

and

$$H_f(\theta) = (\partial f(x)/\partial x_i|_{x=\theta}, i = 1, 2, \ldots, d).$$

In the following and for simplicity, we assume that $d = 1$. The goal, is then to approximate the distribution of

$$\sqrt{n}(T_n - f(\theta)),$$

by means of the following modified AR-sieve bootstrap procedure, which we call the AR-sieve bootstrap with i.i.d. wild innovations.

3.2.2 The bootstrap algorithm

Step 1: Fit an autoregressive model of order $p = p(n) \in \mathbb{N}$, $p < n$, to the time series $X_1, X_2, ..., X_n$, obtain estimates $\widehat{a}_p$ and residuals $\widehat{e}_{t,p}$, $t = p+1, p+2, \ldots, n$, defined as in (3.1).

Step 2: Let $\widehat{\eta}_{4,\varepsilon}$ be the estimator of $\eta_{4,\varepsilon}$ given by

$$\widehat{\eta}_{4,\varepsilon} = \frac{1}{\widehat{\gamma}_e^2(0)} \sum_{h=-(N-1)}^{N-1} k\left(\frac{h}{M_n}\right)\left(\widehat{\gamma}_{2,e}(h) - 2\widehat{\gamma}_e^2(h)\right), \tag{3.5}$$

where $N = n - p$, $0 \leq h \leq N-1$,

$$\widehat{\gamma}_e(h) = \frac{1}{N} \sum_{t=p+1}^{n-|h|} \left(\widehat{e}_{t,p} - \overline{e}_n\right)\left(\widehat{e}_{t+|h|,p} - \overline{e}_n\right), \quad \overline{e}_n = \frac{1}{N} \sum_{t=p+1}^{n} \widehat{e}_{t,p},$$

$$\widehat{\gamma}_{2,e}(h) = \frac{1}{N} \sum_{t=p+1}^{n-|h|} \left(\widehat{e}_{t,p}^2 - \overline{e}_{2,n}\right)\left(\widehat{e}_{t+|h|,p}^2 - \overline{e}_{2,n}\right) \text{ and } \overline{e}_{2,N} = \frac{1}{N} \sum_{t=p+1}^{n} \widehat{e}_{t,p}^2.$$

Here k is a so-called lag-window satisfying $k \geq 0$, $k(x) = 0$ for $|x| \geq 1$ while $M_n < n$ is a truncation parameter to be specified later.

Step 3: Generate $X_1^*, X_2^*, \ldots, X_n^*$ as

$$X_t^* = \sum_{j=1}^{p} \widehat{a}_{j,p} X_{t-j}^* + \sqrt{\widehat{\gamma}_e(0)}\varepsilon_t^*, \ t \in \mathbb{Z}, \tag{3.6}$$

where the innovations ε_t^* are i.i.d. random variables having the following (three point) distribution

$$P\left(\varepsilon_t^* = \sqrt{\widehat{\eta}_{4,\varepsilon} + 3}\right) = P\left(\varepsilon_t^* = -\sqrt{\widehat{\eta}_{4,\varepsilon} + 3}\right) = \frac{1}{2\left(\widehat{\eta}_{4,\varepsilon} + 3\right)}$$

and

$$P\left(\varepsilon_t^* = 0\right) = 1 - \frac{1}{\left(\widehat{\eta}_{4,\varepsilon} + 3\right)}.$$

Step 4: Let T_n^* be the same statistic as T_n defined in (1.7) but with X_t replaced by X_t^*, that is,

$$T_n^* = f\left(\frac{1}{n-m+1} \sum_{t=1}^{n-m+1} g(X_t^*, X_{t+1}^*, \ldots, X_{t+m-1}^*)\right), \tag{3.7}$$

and $\theta^* = E^*(g(X_t^*, X_{t+1}^*, \ldots, X_{t+m-1}^*))$, the analogue of θ associated with the bootstrap process $\mathbf{X}^* = \{X_t^*, t \in \mathbb{Z}\}$, where X_t^* is generated as in (3.6). Use the distribution of $\sqrt{n}(T_n^* - f(\theta^*))$ to approximate the distribution of $\sqrt{n}(T_n - f(\theta))$.

Concerning the estimator $\widehat{\eta}_{4,\varepsilon}$ used in Step 2 of the above AR-sieve bootstrap algorithm the following is mentioned. Recall that in the classical AR-sieve bootstrap, the pseudo-innovations are generated by choosing with replacement from the empirical distribution of the estimated residuals $\widehat{e}_{t,p}$. Furthermore, it yields that (under certain conditions) the empirical fourth order moment of the estimated residuals, that is $(n-p)^{-1}\sum_{t=p+1}^{n}\widehat{e}_{t,p}^{4}$, converges in probability to $E(e_t^4)$, as $p\to\infty$ and $n\to\infty$. However, $E(e_t^4)$ and $E(\varepsilon_t^4)$ and, consequently, $\eta_{4,e}$ and $\eta_{4,\varepsilon}$ may be different; recall the example discussed in Subsection 3.2.1. The statistic $\widehat{\eta}_{4,\varepsilon}$ used in Step 2 is a consistent, nonparametric estimator of $\eta_{4,\varepsilon}$, i.e. of the rescaled, fourth order cumulant of the innovations $\{\varepsilon_t, t\in\mathbb{Z}\}$ appearing in (1.8). Thus, the fourth order cumulant of ε_t is appropriately captured by this modification of the AR-sieve bootstrap. As we have seen in Chapter 2, the estimator $\widehat{\eta}_{4,\varepsilon}$ has certain advantages compared to alternative estimators of the same parameter previously proposed in the literature; see Grenander and Rosenblatt (1957), Janas and Dahlhaus (1994) and Kreiss and Paparoditis (2012). Consistency of this estimator requires that the lag-window k and the truncation parameter M_n satisfy the following assumption.

Assumption 3.4.

(i) $k:[-1,1]\to\mathbb{R}$ *is a symmetric, non-negative and continuous function and satisfies*

$$k(x)=\int_{-\infty}^{\infty}K(u)e^{-iux}du,$$

where K *is a non-negative kernel function. Furthermore,* $k(0)=1$, $|k(u)|\le 1$, *and* $\int_{-\infty}^{\infty}k^2(u)du<\infty$.

(ii) $M_n\to\infty$ *as* $n\to 0$ *such that* $M_n^4/n\to 0$.

The procedure used in Step 3 of the algorithm to generate the i.i.d. pseudo-innovations ε_t^* implies that

$$E^*(\varepsilon_t^*)=0,\quad E^*(\varepsilon_t^{*^2})=1\quad\text{and}\quad E^*(\varepsilon_t^*)^4-3=\widehat{\eta}_{4,\varepsilon},$$

where under certain conditions

$$\widehat{\eta}_{4,\varepsilon}\xrightarrow{P}\eta_{4,\varepsilon},\quad\text{as}\quad n\to\infty.$$

Thus the generation mechanism of the i.i.d. pseudo-innovations ensures that these innovations imitate (asymptotically) correct also the rescaled fourth order cumulant of the true innovations $\{\varepsilon_t\}$. As we will see, this is important for some statistics belonging to the class (1.7). Notice that one could generate the ε_t^*'s using another consistent estimator of $\eta_{4,\varepsilon}$ and/or a different distribution, for instance a distribution from the Pearson family of distributions with mean zero, variance one, zero third moment and kurtosis $\widehat{\eta}_{4,\varepsilon}+3$. Such alternative choices, will not affect the asymptotic results presented in the next section.

3.2.3 Bootstrap validity

The asymptotic validity of the modified AR-sieve bootstrap procedure proposed, is easily established using the concept of the companion process introduced in Kreiss and Paparoditis (2011). To elaborate, consider the process $\widetilde{\mathbf{X}}=\{\widetilde{X}_t, t\in\mathbb{Z}\}$, called the companion process, with $\widetilde{X}_t$ generated as

$$\widetilde{X}_t=\sum_{j=1}^{\infty}a_j\widetilde{X}_{t-j}+\sqrt{\gamma_e(0)}\widetilde{\varepsilon}_t,\ t\in\mathbb{Z}, \tag{3.8}$$

and $\{\widetilde{\varepsilon}_t, t\in\mathbb{Z}\}$ i.i.d. random variables having distribution

$$P\left(\widetilde{\varepsilon}_t=\sqrt{\eta_{4,\varepsilon}+3}\right)=P\left(\widetilde{\varepsilon}_t=-\sqrt{\eta_{4,\varepsilon}+3}\right)=\frac{1}{2\left(\eta_{4,\varepsilon}+3\right)},$$

and

$$P\left(\widetilde{\varepsilon}_t=0\right)=1-\frac{1}{\left(\eta_{4,\varepsilon}+3\right)}.$$

We denote such a sequence of innovations by

$$\widetilde{\varepsilon}_t\sim IID(0,1,m_{4,\varepsilon}),\quad \text{where}\quad m_{4,\varepsilon}:=E\left(\widetilde{\varepsilon}_t^4\right)=\eta_{4,\varepsilon}+3.$$

Let $\widetilde{e}_t=\sqrt{\gamma_e\left(0\right)}\widetilde{\varepsilon}_t$, where $\gamma_e\left(0\right)\equiv\sigma_e^2$. Then, the i.i.d. innovations $\{\widetilde{e}_t, t\in\mathbb{Z}\}$ driving the linear process (3.8) satisfy

$$E\left(\widetilde{e}_t\right)=0,\ E\left(\widetilde{e}_t^2\right)=\sigma_e^2\ \text{ and }\ \eta_{4,\widetilde{e}}=\frac{E\left(\widetilde{e}_t^4\right)-3E^2\left(\widetilde{e}_t^2\right)}{E^2\left(\widetilde{e}_t^2\right)}=\eta_{4,\varepsilon}.$$

The coefficients a_j appearing in (3.8) are those of the AR-representation (1.6) of the underlying process $\mathbf{X}$. Thus, $\mathbf{X}$ and $\widetilde{\mathbf{X}}$ have the same autocovariance structure. Furthermore, notice that $\widetilde{\mathbf{X}}$ is called the companion process because it is the stochastic process the dependence structure of which the modified autoregressive sieve bootstrap proposal asymptotically mimics.

Now, let $\widetilde{T}_n$ be the same statistic as T_n but with $X_1, X_2, \ldots, X_n$ replaced by a fictious time series $\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_n$ stemming from the companion process $\widetilde{\mathbf{X}}$. Let further

$$\widetilde{\theta} = E(g(\widetilde{X}_t, \widetilde{X}_{t+1}, \ldots, \widetilde{X}_{t+m-1})).$$

Thus, and as we will see in Theorem 3.2.1, what the bootstrap sequence $\sqrt{n}(T_n^* - f(\theta^*))$ consistently estimates, is the distribution of $\sqrt{n}(\widetilde{T}_n - f(\widetilde{\theta}))$. Hence the AR-sieve bootstrap procedure with i.i.d. wild generated errors, will be asymptotically valid if and only if the asymptotic distributions of $\sqrt{n}(\widetilde{T}_n - f(\widetilde{\theta}))$ and of $\sqrt{n}(T_n - f(\theta))$ are identical. This simple check criterion for examining the validity of the modified AR-sieve bootstrap procedure proposed, is the consequence of the following theorem.

Theorem 3.2.1. *Suppose that Assumption 3.1 to Assumption 3.4 are satisfied and that $p_n = o(n/\log(n))^{1/4}$ as $n \to \infty$. Then,*

$$d_k\Big(\mathcal{L}\left(\sqrt{n}\left(T_n^* - f\left(\theta^*\right)\right)\right), \mathcal{L}\big(\sqrt{n}(\widetilde{T}_n - f(\widetilde{\theta}))\big)\Big) \to 0, \text{ in probability},$$

where d_k denotes Kolmogorov's distance and $\mathcal{L}(X)$ the distribution of the random variable X.

Notice that the AR-sieve bootstrap with i.i.d. wild innovations works for all statistics for which the classical AR-sieve with i.i.d. innovations obtained from the empirical distribution of the estimated residual $\widehat{e}_{t,p}$ works. This is true since, as we have seen, the autocovariance structure of $\widetilde{\mathbf{X}}$ is identical to that of the underlying process $\mathbf{X}$, that is, $\gamma_X(h) = \gamma_{\widetilde{X}}(h), \forall h \in \mathbb{Z}$. This is easily seen since,

$$\gamma_X(h) = \sigma_e^2 \sum_{j=0}^{\infty} c_j c_{j+h},$$

and

$$\gamma_{\widetilde{X}}(h) = Var(\sqrt{\gamma_e(0)}\widetilde{\varepsilon}_t) \sum_{j=0}^{\infty} c_j c_{j+h} = \sigma_e^2 \sum_{j=0}^{\infty} c_j c_{j+h}.$$

Moreover, because the innovations ε_t^* also imitate asymptotically correct the fourth order moment structure of the true innovations ε_t of the linear process (1.8), the modified

AR-sieve bootstrap is valid for an extended range of statistics for which the classical AR-sieve fails. The following is an example.

Example: Consider the estimator

$$\widehat{\gamma}_X(h) = \frac{1}{n}\sum_{t=1}^{n-h} X_t X_{t+h}$$

of the autocovariance $\gamma_X(h)$, $0 \leq h < n$, which is a special case of (1.7) and recall that for linear processes it yields that

$$\sqrt{n}(\widehat{\gamma}_X(h) - \gamma_X(h)) \xrightarrow{D} N(0, \tau_h^2),$$

where

$$\tau_h^2 = \eta_{4,\varepsilon}\gamma_X^2(h) + \sum_{k=-\infty}^{\infty}(\gamma_X^2(k) + \gamma_X(k+h)\gamma_X(k-h));$$

cf. Brockwell and Davis (1991), Prop. 7.3.1. Since for the companion process we have for the estimator

$$\widetilde{\gamma}_{\widetilde{X}}(h) = \frac{1}{n}\sum_{t=1}^{n-h} \widetilde{X}_t \widetilde{X}_{t+h},$$

that

$$\sqrt{n}(\widetilde{\gamma}_{\widetilde{X}}(h) - \gamma_{\widetilde{X}}(h)) \xrightarrow{D} N(0, \widetilde{\tau}_h^2),$$

where

$$\widetilde{\tau}_h^2 = \eta_{4,\varepsilon}\gamma_{\widetilde{X}}^2(h) + \sum_{k=-\infty}^{\infty}(\gamma_{\widetilde{X}}^2(k) + \gamma_{\widetilde{X}}(k+h)\gamma_{\widetilde{X}}(k-h),$$

i.e., $\widetilde{\tau}_h^2 = \tau_h^2$, we immediately get by Theorem 3.2.1 the validity of AR-sieve bootstrap with i.i.d. wild innovations for this statistic.

3.3 Autoregressive Bootstrap with Dependent Wild Innovations

3.3.1 Motivation

The previous modification extends the range of validity of the classical AR-sieve bootstrap. However, this bootstrap procedure is not valid for general stationary processes and for the entire class of statistics (1.7) due to the i.i.d. structure of the generated pseudo-innovations ε_t^*. Moreover, even for linear processes, this bootstrap procedure does not necessarily imitate correct the entire fourth order moment structure of $\mathbf{X}$. To elaborate, denote by $f_{4,X}$ the fourth order cumulant spectral density of $\mathbf{X}$ and recall that if this process is the linear process (1.8), then $f_{4,X}$ is given by

$$f_{4,X}(\omega_1,\omega_2,\omega_3) = \frac{1}{(2\pi)^3}(E\varepsilon_1^4 - 3\sigma_\varepsilon^4)\Psi(\omega_1)\Psi(\omega_2)\Psi(\omega_3)\Psi(-\omega_1-\omega_2-\omega_3),$$

where

$$\omega_j \in [0,\pi] \quad \text{and} \quad \Psi(\omega) = \sum_{j=-\infty}^{\infty} \psi_j e^{-ij\omega};$$

see Rosenblatt (1985). The fourth order cumulant spectral density $f_{4,\widetilde{X}}$ of the companion process $\widetilde{\mathbf{X}}$, which is the process the dependence structure of which is (asymptotically) imitated by the AR-sieve with i.i.d. wild innovations, see (3.8), is given by

$$f_{4,\widetilde{X}}(\omega_1,\omega_2,\omega_3) = \frac{1}{(2\pi)^3}\frac{\sigma_e^4}{\sigma_\varepsilon^4}(E\varepsilon_1^4 - 3\sigma_\varepsilon^4)\widetilde{\Psi}(\omega_1)\widetilde{\Psi}(\omega_2)\widetilde{\Psi}(\omega_3)\widetilde{\Psi}(-\omega_1-\omega_2-\omega_3),$$

where

$$\widetilde{\Psi}(z) = A^{-1}(z), \quad |z| \leq 1.$$

Since in general

$$f_{4,X} \neq f_{4,\widetilde{X}},$$

it follows that even for linear processes, the AR-sieve bootstrap does not imitate correctly the entire fourth order structure of $\mathbf{X}$. As an example, recall the non-invertible MA(1) process considered in Subsection 3.2.1 and observe that for this process we have

$$\Psi(z) = 1+\theta z, \quad \widetilde{\Psi}(z) = 1+\theta^{-1}z \quad \text{and} \quad \sigma_e^4 = \theta^4\sigma_\varepsilon^4.$$

Thus,

$$f_{4,X}(\omega_1,\omega_2,\omega_3) = \frac{1}{(2\pi)^3}\sigma_\varepsilon^4\eta_{4,\varepsilon}\prod_{j=1}^{3}\left(1+\theta e^{-i\omega_j}\right)\left(1+\theta e^{i\sum_{l=1}^3\omega_l}\right),$$

while

$$f_{4,\widetilde{X}}(\omega_1,\omega_2,\omega_3) = \frac{1}{(2\pi)^3}\sigma_\varepsilon^4\eta_{4,\varepsilon}\prod_{j=1}^{3}\left(\theta+e^{-i\omega_j}\right)\left(\theta+e^{i\sum_{l=1}^3\omega_l}\right).$$

The goal of this section is to develop an AR-bootstrap procedure which is valid for the entire class of statistics (1.7) and for stochastic processes satisfying quite general weak dependence conditions. It is obvious from the previous discussion that such a procedure has to imitate correctly the high order dependence structure of the underlying process which affects the limiting distribution of the statistic T_n. Towards this goal, it is important to observe that it is not necessary for the AR-bootstrap procedure to mimic the entire dependence structure of $\mathbf{X}$. It suffices if it imitates correctly the autocovariance structure of the transformed process

$$\mathbf{Y} = \{Y_t = g(X_t, X_{t+1}, \ldots, X_{t+m-1}), t \in \mathbb{Z}\}.$$

This is true since, as equation (3.4) shows, it is the autovariance structure of $\mathbf{Y}$ that affects the limiting distribution of T_n. Based on this observation, we apply an AR-bootstrap procedure not to the time series $X_1, X_2, \ldots, X_n$ itself but to the transformed time series $Y_1, Y_2, ..., Y_N$, where $N = n - m + 1$.

To elaborate, recall that the coefficients $b_{j,p}$, $j = 1, 2, \ldots, p$, of a linear AR(p)-fit obtained by minimizing the mean square error

$$E(Y_t - \sum_{j=1}^{p} b_{j,p} Y_{t-j})^2,$$

are for every fixed $p \in \mathbb{N}$ uniquely determined provided

$$Var(Y_t) > 0 \quad \text{and} \quad Cov(Y_0, Y_h) \to 0 \quad \text{for} \quad h \to \infty,$$

see Proposition 5.1.1 of Brockwell and Davis (1991). This requirement is fulfilled if Σ_g given in (3.4) is well defined which is true if the following assumption is satisfied.

Assumption 3.5. *The autocovariance function of* $\mathbf{Y}$ *denoted by* $\gamma_Y(h) = Cov(Y_t, Y_{t+h})$, $h \in \mathbb{Z}$, *is absolute summable, i.e.,*

$$\sum_{h=-\infty}^{\infty} |\gamma_Y(h)| < \infty.$$

Furthermore, we assume that

$$\sum_{h_1,h_2,h_3=-\infty}^{\infty} |Cum(Y_0, Y_{h_1}, Y_{h_2}, Y_{h_3})| < \infty.$$

Under the above assumption, the process $\mathbf{Y}$ possesses a continuous spectral density f_Y. However, it is not clear if the transformed process $\mathbf{Y}$ also obeys an autoregressive representation like (1.6), since for this to be true, the spectral density f_Y should also be bounded away from zero from below. For general functions g, like those appearing in the definition of (1.7), such a property is in general difficult to verify. This makes the application of an AR-sieve type bootstrap procedure to the transformed process $\mathbf{Y}$, where p is allowed to increase to infinity with n, difficult to justify theoretically. This problem does not exist in our new proposal since the order p of the autoregression fitted is kept fix. However, since in general, an AR-bootstrap with fixed order p can not capture the entire autocovariance structure of $\mathbf{Y}$, we appropriately modify the i.i.d. resampling scheme applied to the residuals. In particular, we replace the i.i.d. resampling used in the classical AR-bootstrap by a generation of pseudo-innovations using a dependent wild bootstrap procedure; see Shao (2010). Thus, the bootstrap procedure proposed is an AR-bootstrap with dependent wild pseudo-innovations applied to the transformed time series $Y_1, Y_2, \ldots, Y_N$. It is precisely described in the following algorithm.

3.3.2 Bootstrap algorithm and bootstrap validity

Step 1: Fit an autoregressive model of order p to the series $Y_1, Y_2, ..., Y_N$ and obtain estimated residuals

$$\widehat{V}_{t,p} = Y_t - \sum_{j=1}^{p} \widehat{b}_{j,p} Y_{t-j}, \; t = p+1, p+2, ..., N.$$

Step 2: Generate the bootstrap sample $Y_1^*, Y_2^*, ..., Y_N^*$ using

$$Y_j^* = Y_j, \quad j = 1, 2, \ldots, p$$

and

$$Y_t^* = \sum_{j=1}^{p} \widehat{b}_{j,p} Y_{t-j}^* + V_t^*, \quad t = p+1, p+2, \ldots, N,$$

where the pseudo-innovations V_t^* are obtained as

$$V_t^* = (\widehat{V}_{t,p} - \overline{V}_n) W_t^*, \quad t = p+1, p+2, ..., N.$$

Here,

$$\overline{V}_n = \frac{1}{N-p} \sum_{t=p+1}^{N} \widehat{V}_{t,p},$$

and W_t^* $t = p+1, p+2, ..., N$, is a time series stemming from a stationary process $\{W_t^*, t \in \mathbb{Z}\}$ which is independent of $\mathbf{Y}$, with

$$E\left(W_t^*\right) = 0, \quad Var(W_t^*) = 1 \quad \text{and} \quad Cov(W_t^*, W_s^*) = w\left[(t-s)/l_n\right].$$

The function $w\left(\cdot\right)$ is a kernel function which satisfies

$$K_w\left(x\right) = \int_{-\infty}^{\infty} w(z) e^{-izx} dz \geq 0, \; x \in \mathbb{R},$$

and l_n is a bandwidth parameter.

Step 3: Approximate the distribution of

$$\sqrt{n}(T_n - f(\theta))$$

by that of

$$\sqrt{n}(f(\overline{Y}_n^*) - f(E^*(\overline{Y}_n^*))) \quad \text{where} \quad \overline{Y}_n^* = \frac{1}{N}\sum_{t=1}^{N} Y_t^*.$$

The conditions stated about the kernel w ensure the non-negative definiteness of the covariance matrix of W_t^*, while l_n is a resampling parameter used in the dependent wild bootstrap and which will be specified later on.

The following theorem shows that the proposed AR-bootstrap procedure with dependent wild innovations, is valid for the entire class of statistics (1.7) under quite general conditions on the dependence structure of the underlying process $\mathbf{X}$.

Theorem 3.3.1. *Suppose that the statistic T_n given in (1.7) based on a time series of length n from the strictly stationary process $\mathbf{X}$ fulfills (3.3) and that $E|Y_1|^{2+\delta} < \infty$ for some $\delta > 0$. Furthermore, assume that Assumption 3.3 is fulfilled, that the process $\mathbf{Y}$ satisfies Assumption 3.5 and that $\{W_t^*, t \in \mathbb{Z}\}$ is a l_n-dependent process with $E^*|W_t^*|^{2+\delta} < \infty$. If $l_n \to \infty$ as $n \to \infty$ such that $l_n^{2+2/\delta}/n \to 0$, then we have, for every fixed $p \in \mathbb{N}$, that, as $n \to \infty$,*

$$d_k\left(\mathcal{L}\big(\sqrt{n}\big(f(\overline{Y}_n^*) - f(E^*(\overline{Y}_n^*))\big)\big), \mathcal{L}\big(\sqrt{n}(T_n - f(\theta))\big)\right) \to 0,$$

in probability.

We stress here the fact that in order to establish the above theorem and beyond Assumption 3.5, we do not impose any specific weak dependence assumptions on the underlying process $\mathbf{X}$, like mixing or other type of weak dependence conditions. The requirement that the statistic T_n convergence in distribution as stated in (3.3) and (3.4) is very weak and is fulfilled for a wide range of weak dependent processes including for instance, mixing processes, linear processes and processes satisfying other weak dependence assumptions; see for instance Künsch (1989).

3.4 Numerical Results

3.4.1 Choosing the bootstrap parameters

Implementation of the autoregressive bootstrap methods proposed in previous sections, requires the selection of two bootstrap parameters, the order p and the truncation lag M_n respectively the resampling parameter l_n. Concerning the choice of the autoregressive order p, we recommend for both bootstrap procedures to use Akaike's information criterion (AIC); see also (2.14). For the truncation lag M_n or the resampling bandwidth l_n, we provide in the following some heuristic rules which lead to some data-driven procedures to automatically select these two parameters.

For the AR-sieve bootstrap with i.i.d. wild innovations, the nonparametric estimation of $\eta_{4,\varepsilon}$ used, requires the choice of the truncation lag M_n. In Chapter 2, Subsection 2.6.1, a procedure has been proposed for the selection of this parameter which can be also used in the current context. Recall that the idea is to choose M_n in order to minimize an approximation of the (asymptotic) mean square error $E(\widehat{\eta}_{4,\varepsilon} - \eta_{4,\varepsilon})^2$. This approach leads to the formulae

$$M_n^{(opt)} = \left\{ \frac{2K_1^2 \left(\sum_{h=-\infty}^{\infty} h^2 (2c_e(0)\rho_e^2(h) - \rho_{2,e}(h)) \right)^2}{K_2 \left(\sum_{h=-\infty}^{\infty} \rho_{2,e}(h) \right)^2} \right\}^{1/5} n^{1/5}, \tag{3.9}$$

for the optimal value of M_n. Here,

$$c_e(0) = \gamma_e^2(0)/\gamma_{2,e}(0), \quad K_1 = \int_{-1}^{1} x^2 k(x)dx \quad \text{and} \quad K_2 = \int_{-1}^{1} k^2(x)dx.$$

Furthermore,

$$\rho_e(h) = \gamma_e(h)/\gamma_e(0) \quad \text{and} \quad \rho_{2,e}(h) = \gamma_{2,e}(h)/\gamma_{2,e}(0)$$

are the autocorrelation functions at lag h of the filtered processes $\{e_{t,p}, t \in \mathbb{Z}\}$ and $\{e_{t,p}^2, t \in \mathbb{Z}\}$ respectively, while $\gamma_{2,e} = Cov(e_{t,p}^2, e_{t+h,p}^2)$ is the autocovariance at lag h of the filtered squared process. Using this formulae, M_n can be chosen by replacing $c_e(0)$, $\rho_{2,e}(h)$ and $\rho_e(h)$ by the corresponding sample estimators based on the estimated residuals $\widehat{e}_{t,p}$, see (3.1), and by truncating the infinite sums in (3.9) to some finite, small value L, i.e., $L = 1$.

For the AR-bootstrap procedure with dependent wild innovations, we use the following heuristic rule to select l_n. Since the autoregressive fit intends to capture the second order structure of $Y_1, Y_2, \ldots, Y_N$, the choice of l_n should concerned with the imitation of the fourth order structure of the filtered process

$$V_{t,p} = Y_t - \sum_{j=1}^{p} b_{j,p} Y_{t-j}.$$

Making the working assumption that $\mathbf{Y}$ is a linear process implies that the fourth order structure of the filtered process $V_{t,p}$ can estimated using the same strategy as for the filtered process

$$e_{t,p} = X_t - \sum_{j=1}^{p} a_{j,p} X_{t-j}.$$

This suggests the use of formulae (3.9) to select l_n where $c_e(0)$, $\rho_e(h)$ and $\rho_{2,e}(h)$ are now replaced by

$$c_V(0) = \gamma_V^2(0)/\gamma_{2,V}(0), \quad \rho_V(h) = \gamma_V(h)/\gamma_V(0) \quad \text{and} \quad \rho_{2,V}(h) = \gamma_{2,V}(h)/\gamma_{2,V}(0).$$

Here $\gamma_V(h)$ and $\gamma_{2,V}(h)$ are the autocovariances at lag h of the process $\{V_{t,p}, t \in \mathbb{Z}\}$ and of the squared process $\{V_{t,p}^2, t \in \mathbb{Z}\}$ respectively. Replacing these quantities by sample estimates based on the estimated residuals $\widehat{V}_{t,p}$, $t = p+1, p+2, \ldots, N$, and truncating the infinite sums, as in the case of M_n, leads to a pactical rule for selecting the parameter l_n of the dependent wild bootstrap. As our simulations in the next subsection show, the rules proposed in this subsection to select the bootstrap parameters p, M_n and l_n work very good in practise.

3.4.2 Simulations

We investigate the ability of the different bootstrap methods to estimate the standard deviation of the first order sample autocovariance, i.e., of $\sqrt{n}\,\widehat{\gamma}(1)$, for time series of length $n = 100$ and $n = 300$ stemming from five different models and driven by i.i.d. innovations having four different distributions. Furthermore, four different bootstrap methods are compared. The autoregressive sieve bootstrap (ARS), the autoregressive sieve bootstrap with i.i.d. wild innovations (ARSW), the autoregressive bootstrap with dependent wild innovations (ARDW) and the block bootstrap (BB). The following five time series models have been considered in the simulation study:

Model I: $X_t = \phi X_{t-1} + \varepsilon$, with $\phi = 0.8$,

Model II: $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$, with $\theta = 0.8$,

Model III: $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$, with $\theta = 2$,

Model IV: $X_t = 0.6\sin(X_{t-1}) + \varepsilon_t$,

Model V: $X_t = \big\{0.8 - 1.1\exp\{-50X_{t-1}^2\}\big\}X_{t-1} + 0.1\varepsilon_t$.

Models I-III are linear models with Model III being a non-invertible first order moving average process. The nonlinear Model IV has been used in Paparoditis and Politis (2001) and Shao (2010), while the nonlinear Model V by Auestad and Tjøstheim (1990). Concerning the i.i.d. innovations, the following distributions with mean zero and unit variance have been used:

- **(I)** Standard Gaussian, ($\eta_{4,\varepsilon} = 0$),
- **(II)** Logistic, ($\eta_{4,\varepsilon} = 1.2$),
- **(III)** Double Exponential, ($\eta_{4,\varepsilon} = 3.0$),
- **(IV)** A distribution from the Pearson family with $\eta_{4,\varepsilon} = 5$.

The autoregressive order for the ARS bootstrap has been selected using AIC while the block size b in the BB procedure has been selected as follows. We calculated the mean square error of the BB estimates for several values of b between 1 and 20 and selected the value of b which minimizes the (empirical) mean square error (MSE) over $R = 200$ replications. In other words, the BB estimates presented in this section are the best (in the MSE sense) estimates that can be obtained using this bootstrap method.

Figure 3.1 and Figure 3.2 present the ratios of the mean bootstrap estimates of the standard deviation and of the estimated exact standard deviation of $\sqrt{n}\widehat{\gamma}(1)$ over $R = 200$ replications, for each of the different models and of the different distributions of the innovations considered. Table 3.1 and Table 3.2 present the estimated exact standard deviations, the mean bootstrap estimates, the standard deviations and the corresponding mean square errors of the different bootstrap estimates.

As it is seen from the two tables and the two exhibits presented, the ARS and the ARSW behave quite good in the case of the three linear models considered and for these models, they outperform the BB. This is expected since these models are taylor made for linear bootstrap procedures. Furthermore, the ARSW estimates seem to be less biased for the case of the non-invertible Model III compared to the estimates of the classical ARS. However, both linear procedures, that is, the ARS and the ARSW, become quite biased in the case of the nonlinear models considered, i.e., in the case of Model IV and Model V. For both nonlinear models, the BB estimates turn out to be also biased although, their bias is smaller compared to that of the linear bootstrap procedures ARS and ARSW. The ARDW estimates are quite stable and less biased for all models and all different distributions of the innovations considered. Notice that the biases of the ARDW method are, in most of the cases considered, the smallest among the biases of all bootstrap methods compared in this simulation study. Also regarding the mean square error, the ARDW estimates behave quite well with their MSE being in many cases close to the lowest MSE that can be achieved by the BB procedure using the best possible choice of the block length as explained before. This observation is especially true for the case of $n = 300$ observations.

Model	Estimated Exact	AR-Sieve IID			AR-Sieve Wild IID			AR-Dependent Wild			Block Bootstrap		
		Mean	**Std**	**MSE**	**Mean**	**Std**	**MSE**	**Mean**	**Std**	**MSE**	**Mean**	**Std**	**MSE**
Gauss													
Model (1)	7.38180	6.18840	2.76270	9.01900	6.30480	2.95080	9.82400	6.41790	3.39520	12.3990	4.77810	2.20200	11.6040
Model (2)	2.06720	2.04720	0.53370	0.28400	2.06880	0.56650	0.31900	2.08180	0.68336	0.46500	1.73840	0.38390	0.25500
Model (3)	5.93700	5.85200	1.43530	2.05700	5.91170	1.44390	2.07500	5.83840	1.78460	3.17900	4.88460	0.80270	1.74800
Model (4)	1.23420	1.43170	0.31370	0.13700	1.46420	0.31500	0.15200	1.24660	0.33190	0.11000	1.12910	0.17410	0.04100
Model (5)	0.04927	0.02791	0.01699	0.21630	0.02813	0.01705	0.21611	0.04302	0.03209	0.20322	0.03067	0.02046	0.21388
Logis													
Model (1)	7.70240	6.32210	2.82770	9.86100	6.26000	2.71500	9.41500	6.39390	3.32160	12.6900	4.69540	2.23420	14.0090
Model (2)	2.23350	2.15630	0.64990	0.42600	2.16830	0.66810	0.44800	2.19130	0.88670	0.78400	1.81580	0.51040	0.43400
Model (3)	6.23940	5.75780	1.67740	3.03100	5.88170	1.71250	3.04600	5.88700	2.28140	5.30300	4.96200	1.05580	2.74100
Model (4)	1.19300	1.41290	0.34090	0.16400	1.42090	0.34580	0.17100	1.18590	0.35850	0.12800	1.09180	0.19960	0.05000
Model (5)	0.05697	0.02734	0.01632	0.29442	0.02789	0.01692	0.29384	0.04124	0.03091	0.28022	0.03010	0.02018	0.29158
Dexp													
Model (1)	8.20220	6.77840	3.40200	13.5430	6.90450	3.49500	13.8380	7.33720	4.48470	20.7600	5.29620	2.85210	16.5390
Model (2)	2.45700	2.19910	0.7963	0.69700	2.24740	0.87810	0.81100	2.29460	1.00430	1.03000	1.93520	0.73860	0.81500
Model (3)	6.69020	5.89770	1.85380	4.04700	6.21010	1.96690	4.08000	6.36670	2.64640	7.07300	5.41800	1.71310	4.53900
Model (4)	1.16280	1.34960	0.33320	0.14500	1.37380	0.3519	0.16800	1.10300	0.3725	0.14200	1.10550	0.25930	0.07000
Model (5)	0.06268	0.03604	0.02155	0.34943	0.03677	0.02122	0.34855	0.05695	0.03823	0.32615	0.03790	0.02438	0.34736
$\eta_4 = 5$													
Model (1)	8.41330	6.79210	3.67240	16.0470	6.66670	3.31220	13.9660	7.42710	4.88060	24.6740	5.32650	3.16860	19.5180
Model (2)	2.66670	2.31250	1.26110	1.70800	2.32990	1.08730	1.29000	2.44030	1.60290	2.60800	2.05270	1.38400	2.28300
Model (3)	7.22120	5.97840	2.06200	5.77500	6.17500	2.28470	6.28800	6.28710	3.05650	10.1680	4.80720	1.21360	7.29300
Model (4)	1.17760	1.36280	0.39360	0.18800	1.35340	0.35373	0.15500	1.10700	0.32960	0.11300	1.04730	0.24440	0.07600
Model (5)	0.07358	0.03312	0.02841	0.49451	0.03243	0.02645	0.49537	0.05240	0.04590	0.46909	0.03424	0.02545	0.49278

Table 3.1: Mean (Mean), standard deviation (Std) and mean square error (MSE) of the estimated standard deviation of the first order sample autocovariance $\sqrt{n}\widehat{\gamma}(1)$, where n is the sample size, for different models, different innovation distributions and using four different bootstrap methods, for a sample size of n = 100 observations.

Model	Estimated Exact	AR-Sieve IID			AR-Sieve Wild IID			AR-Dependent Wild			Block Bootstrap		
		Mean	**Std**	**MSE**	**Mean**	**Std**	**MSE**	**Mean**	**Std**	**MSE**	**Mean**	**Std**	**MSE**
Gauss													
Model (1)	7.92621	7.60977	2.28388	5.28900	7.70728	2.31367	5.37300	7.92084	2.82757	7.95600	6.41292	2.05057	6.47400
Model (2)	2.13337	2.11812	0.366848	0.13500	2.14081	0.37118	0.13800	2.12107	0.47839	0.22800	1.89677	0.32026	0.15900
Model (3)	6.02580	5.99896	0.93739	0.87600	5.98354	0.94345	0.88800	6.09474	1.24396	1.54500	5.34719	0.74877	1.01700
Model (4)	1.24049	1.44834	0.19001	0.07800	1.46480	0.19555	0.08700	1.23911	0.20369	0.04200	1.15961	0.11968	0.02070
Model (5)	0.05330	0.02942	0.01250	0.79903	0.02929	0.01276	0.79926	0.05165	0.03398	0.76078	0.03998	0.02126	0.78056
Logis													
Model (1)	8.25149	7.91738	2.40235	5.85300	7.91564	2.33567	5.54100	8.31661	3.60197	12.9120	6.60240	2.28735	7.92600
Model (2)	2.31922	2.29531	0.42158	0.17700	2.27297	0.41413	0.17400	2.35247	0.60137	0.36000	2.01385	0.39958	0.25200
Model (3)	6.44756	6.08314	1.03559	1.20000	6.30588	1.16290	1.36500	6.52256	1.75110	3.05700	5.71074	1.06608	1.67400
Model (4)	1.20048	1.44384	0.18845	0.09300	1.45059	0.21079	0.10500	1.20187	0.22222	0.04800	1.14055	0.13371	0.02130
Model (5)	0.05976	0.03161	0.01165	1.00720	0.03204	0.01213	1.00630	0.05585	0.02588	0.95966	0.04531	0.01894	0.98011
Dexp													
Model (1)	8.73231	8.13232	2.37984	5.99400	8.13284	2.31281	5.68200	8.34606	3.22802	10.5180	6.87226	2.41708	9.27300
Model (2)	2.51372	2.47458	0.54629	0.29700	2.45968	0.50974	0.26100	2.58664	0.78635	0.62100	2.17875	0.51442	0.37500
Model (3)	6.88733	6.38694	1.23028	1.75500	6.76556	1.35187	1.83300	6.84749	1.74747	3.03900	6.08625	1.25245	2.20200
Model (4)	1.19130	1.39967	0.219451	0.09000	1.41699	0.224127	0.10200	1.18732	0.278687	0.07800	1.11665	0.160561	0.03000
Model (5)	0.06885	0.03869	0.01436	1.33150	0.03907	0.01485	1.33070	0.06675	0.03255	1.26840	0.05447	0.02540	1.29580
$\eta_4 = 5$													
Model (1)	9.22438	8.7737	3.74071	14.1270	8.47631	3.01948	9.63000	8.98865	4.20074	17.6130	7.44297	3.17156	13.1820
Model (2)	2.75032	2.51095	0.73699	0.59700	2.45051	0.65991	0.52200	2.59686	1.00113	1.02000	2.31523	0.75829	0.76200
Model (3)	7.31237	6.56187	1.51762	2.85600	6.85147	1.79787	3.42900	7.35862	2.46488	6.04800	6.62752	2.19970	5.28300
Model (4)	1.21607	1.44037	0.27332	0.12300	1.45527	0.27141	0.12900	1.18143	0.22222	0.05100	1.11994	0.15363	0.03300
Model (5)	0.07448	0.03812	0.01854	1.56750	0.03846	0.01977	1.56670	0.06773	0.03714	1.49530	0.05489	0.02791	1.52630

Table 3.2: Mean (Mean), standard deviation (Std) and mean square error (MSE) of the estimated standard deviation of the first order sample autocovariance $\sqrt{n}\widehat{\gamma}(1)$, where n is the sample size, for different models, different innovation distributions and using four different bootstrap methods, for a sample size of n = 300 observations.

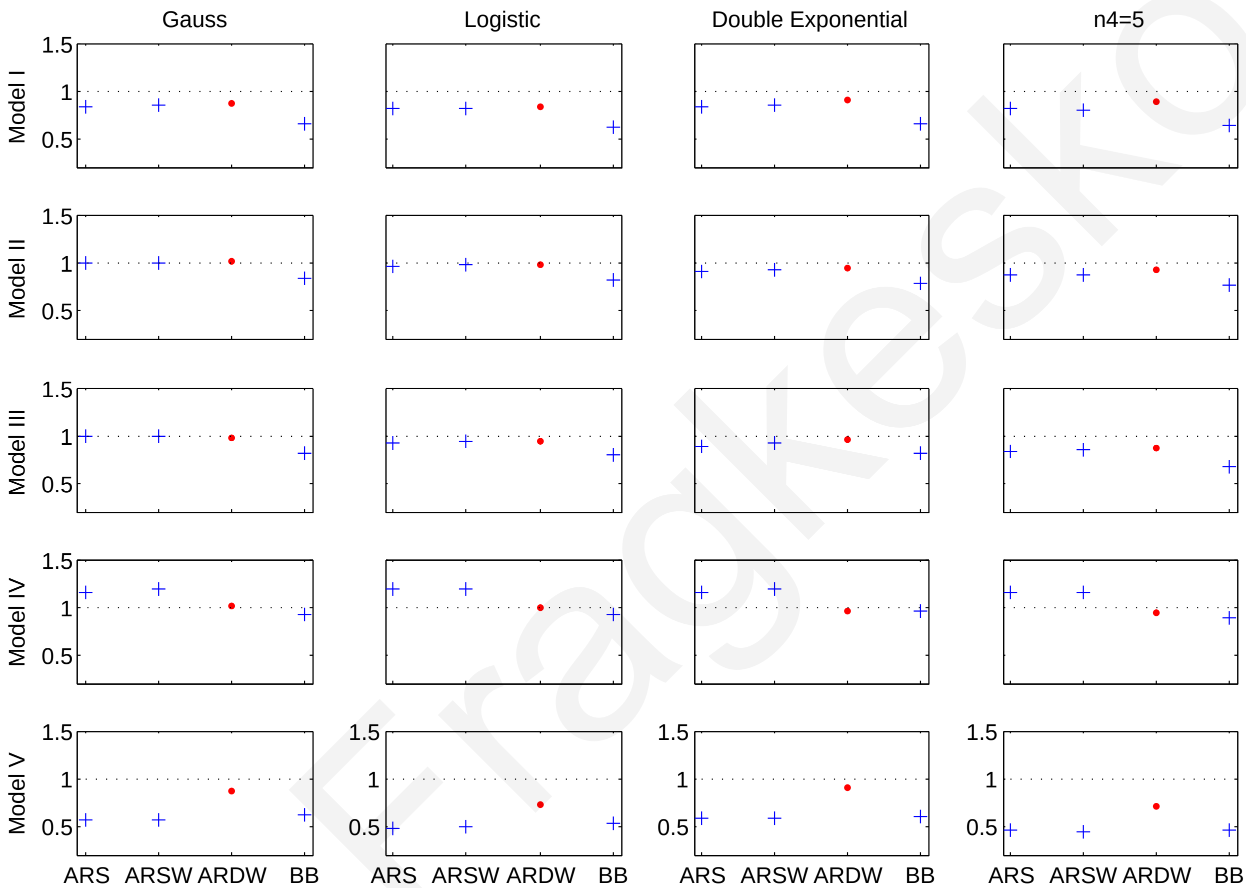

Figure 3.1: Ratio of the estimated standard deviation of the first order sample autocovariance function divided by the estimated exact standard deviation, for the different models, the different innovation distributions and using the autoregressive sieve bootstrap (ARS), the autoregressive wild bootstrap (ARSW), the autoregressive dependent wild bootstrap (ARDW) and the block bootstrap (BB), for a sample size of n = 100 observations.

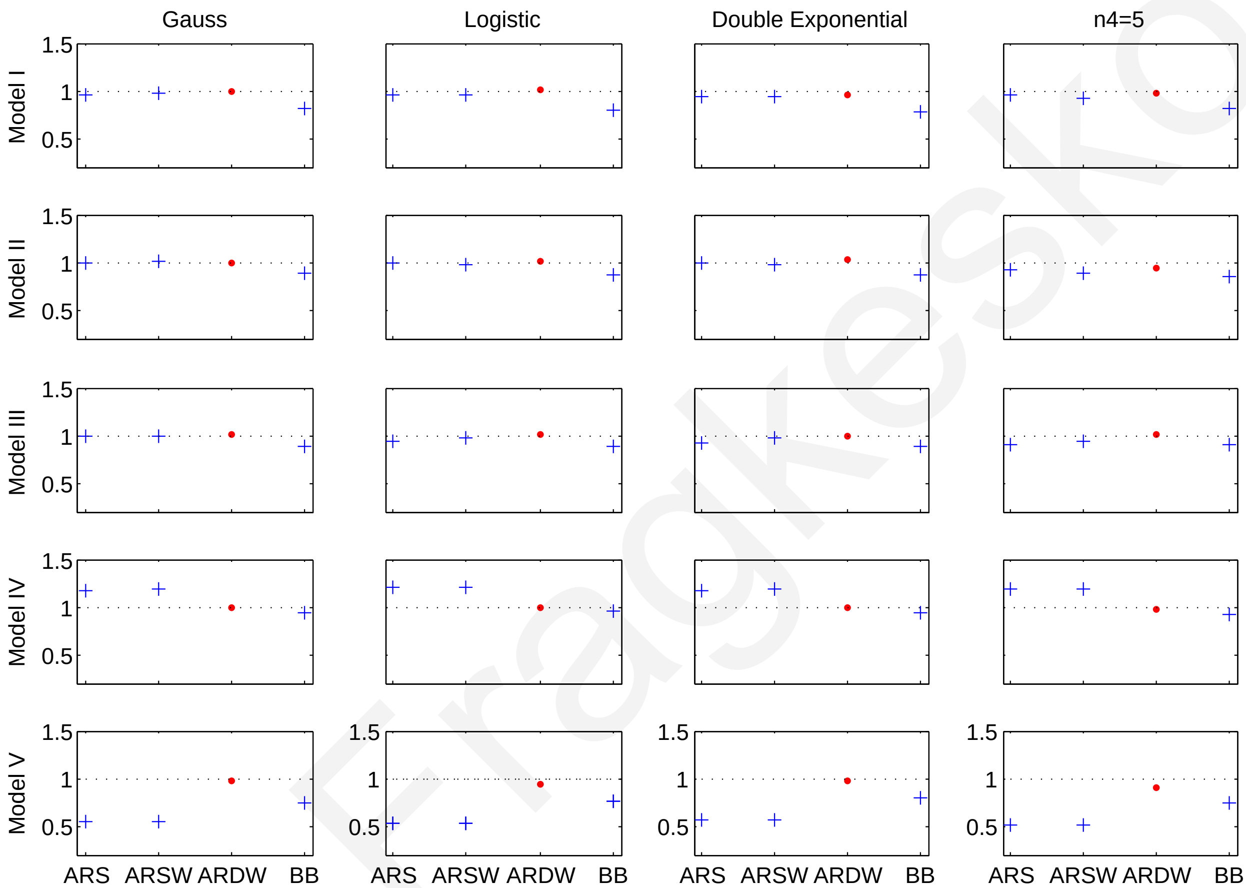

Figure 3.2: Ratio of the estimated standard deviation of the first order sample autocovariance function divided by the estimated exact standard deviation, for the different models, the different innovation distributions and using the autoregressive sieve bootstrap (ARS), the autoregressive wild bootstrap (ARSW), the autoregressive dependent wild bootstrap (ARDW) and the block bootstrap (BB), for a sample size of n = 300 observations.

3.5 Proofs

Lemma 3.5.1. *Suppose that Assumption 3.1, Assumption 3.2 with* $p_n = o(n/\log(n))^{1/4}$ *and Assumption 3.4 are satisfied. Then,*

(i) $\varepsilon_t^* \xrightarrow{D} \widetilde{\varepsilon}_t$ *in probability,*

(ii) $X_t^* \xrightarrow{D} \widetilde{X}_t$ *in probability.*

Proof: (i) Follows immediately since under the assumptions made

$$\widehat{\gamma}_e(0) \xrightarrow{P} \gamma_e(0) \quad \text{and} \quad \widehat{\eta}_{4,\varepsilon} \xrightarrow{P} \eta_{4,\varepsilon}.$$

(ii) Let $\widehat{\psi}_{j,p}$ be the coefficients of

$$\widehat{A}_p^{-1}(z) = \sum_{j=0}^{\infty} \widehat{\psi}_{j,p} z^j, \quad \widehat{\psi}_{0,p} = 1, \quad |z| \leq 1,$$

where

$$\widehat{A}_p(z) = 1 - \sum_{j=1}^{p} \widehat{a}_{j,p} z^j.$$

Let $e_t^* = \sqrt{\widehat{\gamma}_e(0)}\varepsilon_t^*$. For $M \in \mathbb{N}$ we write

$$X_t^* = \sum_{j=0}^{M} \psi_{j,p} e_{t-j}^* + U_t^* + V_t^*$$

where

$$U_t^* = \sum_{j=0}^{M} \left(\widehat{\psi}_{j,p} - \psi_{j,p}\right) \mathrm{e}_{t-j}^* \quad \text{and} \quad V_t^* = \sum_{j=M+1}^{\infty} \widehat{\psi}_{j,p} e_{t-j}^*.$$

Let $x \in \mathbb{R}$ be a continuity point of the distribution function of $\widetilde{X}_t$. By Slutsky's theorem and for $\gamma > 0$ we get

$$P^*(X_t^* \leq x) \leq P^*\left(\sum_{j=0}^{M} \psi_{j,p} e_{t-j}^* \leq x + \gamma\right) + P^*(|U_t^*| \leq \gamma/2) + P^*(|V_t^*| \leq \gamma/2).$$

Applying Lemma 5.1 and Lemma 5.2 of Bühlmann (1997) and using the fact that

$$E^*(\varepsilon_t^*)^2 \xrightarrow{P} E(\widetilde{\varepsilon}_t)^2, \quad \text{and} \quad \text{that} \quad \widehat{\gamma}_e(0) \xrightarrow{P} \gamma_e(0),$$

we can choose for any $k > 0$, a $M = M(\gamma, k)$, such that for n sufficiently large

$$P^*(|U_t^*| \leq \gamma/2) \leq k/2, \quad \text{in probability},$$

and

$$P^*\left(|V_t^*| \leq \gamma/2\right) \leq k/2, \text{ in probability.}$$

We then have, in probability, that

$$P^*\left(X_t^* \leq x\right) \leq P^*\left(\sum_{j=0}^{M} \psi_{j,p} e_{t-j}^* \leq x + \gamma\right) + k,$$

and, similarly,

$$P^*\left(X_t^* \leq x\right) \geq P^*\left(\sum_{j=0}^{M} \psi_{j,p} e_{t-j}^* \leq x - \gamma\right) - k.$$

Part (i) of the lemma, together with the i.i.d. property of $\{\widetilde{\varepsilon}_t, t \in \mathbb{Z}\}$ and of $\{\varepsilon_t^*, t \in \mathbb{Z}\}$ yield for n sufficiently large, that, for an arbitrary $l > 0$,

$$P^*\left(\sum_{j=0}^{M} \psi_{j,p} e_{t-j}^* \leq x + \gamma\right) \leq P\left(\widetilde{X}_t \leq x + \gamma + l\right) + 2k$$

and

$$P^*\left(\sum_{j=0}^{M} \psi_{j,p} e_{t-j}^* \leq x - \gamma\right) \geq P\left(\widetilde{X}_t \leq x - \gamma - l\right) - 2k.$$

Thus, for n sufficiently large

$$P^*\left(X_t^* \leq x\right) \leq P\left(\widetilde{X}_t \leq x + \gamma + l\right) + 3k,$$

and

$$P^*\left(X_t^* \leq x\right) \geq P\left(\widetilde{X}_t \leq x - \gamma - l\right) - 3k,$$

in probability, which concludes the proof. □

Proof of Theorem 3.2.1: A careful inspection of the proof of Theorem 3.1 in Kreiss et al. (2011), shows that to establish Theorem 3.2.1 it suffices to show that for every $r \in \mathbb{N}$,

$$\left(X_{t_1}^*, ..., X_{t_r}^*\right) \xrightarrow{D} \left(\widetilde{X}_{t_1}, ..., \widetilde{X}_{t_r}\right) \text{ in probability.} \tag{3.10}$$

For this, we decompose each $X_{t_i}^*$ as in the proof of Lemma 3.5.1 (ii) and proceed along the same lines as in the proof of the corresponding assertion to show that for any $c_1, c_2, ..., c_r$ with $c_i \in \mathbb{R}$, $i = 1, 2, ..., r$,

$$\sum_{i=1}^{r} c_i X_{t_i}^* \xrightarrow{D} \sum_{i=1}^{r} c_i \widetilde{X}_{t_i} \text{ in probability,}$$

which by the Cramér-Wold device establishes assertion (3.10). □

Proof of Theorem 3.3.1: Define a bootstrap time series $\widetilde{Y}_1^*, \widetilde{Y}_2^*, ..., \widetilde{Y}_N^*$, where

$$\widetilde{Y}_t^* = Y_t \quad \text{for} \quad t = 1, 2, \ldots, p,$$

and

$$\widetilde{Y}_t^* = \sum_{j=1}^{p} \widehat{b}_{j,p} \widetilde{Y}_{t-j}^* + \widetilde{V}_t^* \text{ for } \quad t = p+1, p+2, ..., N. \tag{3.11}$$

Here, $\widetilde{V}_t^*$ are dependent wild bootstrap generated observations, which are obtained as

$$\widetilde{V}_t^* = (V_{t,p} - \overline{\overline{V}}_n) W_t^*, \ t = p+1, p+2, ..., N, \qquad \overline{\overline{V}}_n = \frac{1}{N-p} \sum_{t=p+1}^{N} V_{t,p},$$

and $V_{t,p} = Y_t - \sum_{j=1}^{p} b_{j,p} Y_{t-j}, \ t = p+1, p+2, ..., N$. Notice that, in contrast to $Y_1^*, Y_2^*, ..., Y_N^*$, the random variables $\widetilde{Y}_1^*, \widetilde{Y}_2^*, ..., \widetilde{Y}_N^*$ are based on the true filtered time series $V_{t,p}, t = p+1, p+2, ..., N$. In the following, and in all related cases, we ignore the effect of the starting values.

We first show that

$$\frac{1}{\sqrt{N}} \sum_{t=1}^{N} \left(Y_t^* - \widetilde{Y}_t^* \right) = o_P(1), \tag{3.12}$$

that is, that the effect of estimating $V_{t,p}$ by $\widehat{V}_{t,p}$ is asymptotically negligible.

Let $\widehat{c}_{j,p}$ be the coefficients of the inverse polynomial

$$\left(1 - \sum_{j=1}^{p} \widehat{b}_{j,p} z^j \right)^{-1} = \sum_{j=0}^{\infty} \widehat{c}_{j,p} z^j, \quad \widehat{c}_{0,p} = 1, \quad |z| \leq 1.$$

Then, Y_t^* and $\widetilde{Y}_t^*$ can be expressed as

$$Y_t^* = \sum_{j=0}^{t-1} \widehat{c}_{j,p} (\widehat{V}_{t-j,p} - \overline{V}_n) W_{t-j}^* \quad \text{and} \quad \widetilde{Y}_t^* = \sum_{j=0}^{t-1} \widehat{c}_{j,p} (V_{t-j,p} - \overline{\overline{V}}_n) W_{t-j}^*$$

respectively. We have

$$\begin{aligned}
\frac{1}{\sqrt{N}} \sum_{t=1}^{N} \left(Y_t^* - \widetilde{Y}_t^* \right) &= \frac{1}{\sqrt{N}} \sum_{t=1}^{N} \sum_{j=0}^{t-1} \widehat{c}_{j,p} (\widehat{V}_{t-j,p} - V_{t-j,p}) W_{t-j}^* + (\overline{\overline{V}}_n - \overline{V}_n) \frac{1}{\sqrt{N}} \sum_{t=1}^{N} \sum_{j=0}^{t-1} \widehat{c}_{j,p} W_{t-j}^* \\
&= \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} \sum_{r=1}^{N-j} \widehat{c}_{j,p} (\widehat{V}_{r,p} - V_{r,p}) W_r^* + (\overline{\overline{V}}_n - \overline{V}_n) \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} \sum_{r=1}^{N-j} \widehat{c}_{j,p} W_r^* \\
&= - \sum_{j=0}^{N-1} \widehat{c}_{j,p} \sum_{j_1=1}^{p} \left(\widehat{b}_{j_1,p} - b_{j_1,p} \right) \frac{1}{\sqrt{N}} \sum_{r=1}^{N-j} Y_{r-j_1} W_r^* \\
&\quad + (\overline{\overline{V}}_n - \overline{V}_n) \sum_{j=0}^{N-1} \widehat{c}_{j,p} \frac{1}{\sqrt{N}} \sum_{r=1}^{N-j} W_r^* \\
&= T_{1,N} + T_{2,N},
\end{aligned}$$

with an obvious notation for $T_{1,N}$ and $T_{2,N}$. Observe that

$$\left|\sum_{j=0}^{N-1}\widehat{c}_{j,p}\frac{1}{\sqrt{N}}\sum_{r=1}^{N-j}W_r^*\right| \leq \sum_{j=0}^{\infty}|\widehat{c}_{j,p}|\left|\frac{1}{\sqrt{N}}\sum_{r=1}^{N-j}W_r^*\right| = O_P(\sqrt{l_n}), \tag{3.13}$$

since

$$Var^*\left(\frac{1}{\sqrt{N}}\sum_{r=1}^{N-j}W_r^*\right) = \frac{1}{N}\sum_{t,s=1}^{N-j}w\left(\frac{s-t}{l_n}\right)$$
$$= \frac{1}{N}\sum_{r=-(N-j-1)}^{N-j-1}(N-j-|r|)w\left(\frac{r}{l_n}\right) = O(l_n),$$

uniformly in j. Thus, from equation (3.13) and since

$$\overline{V}_n = O_P\left(1/\sqrt{n}\right) \quad \text{and} \quad \overline{\overline{V}}_n = O_P\left(1/\sqrt{n}\right)$$

we get that,

$$T_{2,N} = O_P\left(\sqrt{l_n/n}\right) \to 0 \text{ as } n \to \infty.$$

Regarding $T_{1,N}$ notice that,

$$|T_{1,N}| \leq \sum_{j=0}^{N-1}|\widehat{c}_{j,p}|\sum_{j_1=1}^{p}\left|\widehat{b}_{j_1,p} - b_{j_1,p}\right|\left|\frac{1}{\sqrt{N}}\sum_{r=1}^{N-j}Y_{r-j_1}W_r^*\right|$$

while,

$$\left|\frac{1}{\sqrt{N}}\sum_{r=1}^{N-j}Y_{r-j_1}W_r^*\right| = O_P(1)$$

uniformly in j and j_1. This follows because

$$E(E^*(\frac{1}{\sqrt{N}}\sum_{r=1}^{N-j}Y_{r-j_1}W_r^*)^2) = \frac{1}{N}\sum_{t,s=1}^{N-j}\gamma_Y(t-s)w\left(\frac{t-s}{l_n}\right) \leq \sum_{r=-\infty}^{\infty}|\gamma_Y(r)| < \infty.$$

Additionally, since

$$\sum_{j=0}^{\infty}|\widehat{c}_{j,p}| < \infty \quad \text{and} \quad \sum_{j=1}^{p}\left|\widehat{b}_{j,p} - b_{j,p}\right| = O_P\left(1/\sqrt{n}\right),$$

we conclude that

$$T_{1,N} = O_P\left(1/\sqrt{n}\right)$$

which yields equation (3.12).

Next, define random variables $Y_1^+, Y_2^+, ..., Y_N^+$, as $Y_j^+ = Y_j$ for $j = 1, 2, \ldots, p$ and

$$Y_t^+ = \sum_{j=1}^{p}b_{j,p}Y_{t-j}^+ + \widetilde{V}_t^* = \sum_{j=0}^{t-1}c_{j,p}(V_{t-j,p} - \overline{\overline{V}}_n)W_{t-j}^*, \text{ for } \quad t = p+1, p+2, ..., N.$$

We show that

$$\frac{1}{\sqrt{N}}\sum_{t=1}^{N}\left(\widetilde{Y}_t^* - Y_t^+\right) = o_P(1). \tag{3.14}$$

To see (3.14) write

$$\begin{aligned}\frac{1}{\sqrt{N}}\sum_{t=1}^{N}\left(\widetilde{Y}_t^* - Y_t^+\right) &= \frac{1}{\sqrt{N}}\sum_{t=1}^{N}\sum_{j=0}^{t-1}(\widehat{c}_{j,p} - c_{j,p})\left(V_{t-j,p} - \overline{\overline{V}}_n\right)W_{t-j}^* \\ &= \frac{1}{\sqrt{N}}\sum_{j=0}^{N-1}(\widehat{c}_{j,p} - c_{j,p})\sum_{r=1}^{N-j}V_{r,p}W_r^* - \overline{\overline{V}}_n\frac{1}{\sqrt{N}}\sum_{j=0}^{N-1}(\widehat{c}_{j,p} - c_{j,p})\sum_{r=1}^{N-j}W_r^* \\ &= T_{3,N} + T_{4,N},\end{aligned}$$

with an obvious notation for $T_{3,N}$ and $T_{4,N}$. Along the same lines as for the term $T_{2,N}$ and using $\overline{\overline{V}}_n = O_P\left(1/\sqrt{n}\right)$ and

$$\sum_{j=0}^{\infty}|\widehat{c}_{j,p} - c_{j,p}| = O_P\left(1/\sqrt{n}\right),$$

see Kreiss and Franke (1992), Lemma 2.2, we conclude that

$$T_{4,N} = O_P\left(\sqrt{l_n}/n\right).$$

Furthermore,

$$|T_{3,N}| \leq \sum_{j=0}^{N-1}|\widehat{c}_{j,p} - c_{j,p}|\left|\frac{1}{\sqrt{N}}\sum_{r=1}^{N-j}V_{r,p}W_r^*\right|,$$

where,

$$\left|\frac{1}{\sqrt{N}}\sum_{r=1}^{N-j}V_{r,p}W_r^*\right| = O_P(1)$$

uniformly in j, since

$$E(E^*(\frac{1}{\sqrt{N}}\sum_{r=1}^{N-j}V_{r,p}W_r^*)^2) = \frac{1}{N}\sum_{t,s=1}^{N-j}\gamma_V\left(t-s\right)w\left(\frac{t-s}{l_n}\right) \leq \sum_{r=-\infty}^{\infty}|\gamma_V\left(r\right)| < \infty,$$

due to Assumption 3.5, where $\gamma_V\left(r\right) = Cov\left(V_{t,p}, V_{t+r,p}\right)$, $r \in \mathbb{Z}$ is the autocovariance function of $\{V_{t,p}, t \in \mathbb{Z}\}$. Thus,

$$T_{3,N} = O_P(1/\sqrt{n})$$

which completes the proof of assertion (3.14).

Let,

$$L_n^* = \frac{1}{\sqrt{N}}\sum_{t=1}^{N}Y_t^*,$$

and

$$L_n = \frac{1}{\sqrt{N}} \sum_{t=1}^{N} Y_t^+ = \frac{1}{\sqrt{N}} \sum_{t=1}^{N} \sum_{j=0}^{t-1} c_{j,p} \widetilde{V}_{t-j}^*.$$

Notice that by (3.12) and (3.14) we have that

$$L_n^* = L_n + o_P(1).$$

In order to show that

$$L_n \xrightarrow{D} N(0, \sigma_\infty^2), \quad \text{as} \quad n \to \infty, \quad \text{where} \quad \sigma_\infty^2 = 2\pi f_Y(0),$$

it suffices by Proposition 6.3.9 of Brockwell and Davis (1991), to show that

(a) $L_{n,M} \xrightarrow{D} Z_M$ for M fixed where

$$L_{n,M} = \frac{1}{\sqrt{N}} \sum_{t=1}^{N} \sum_{j=0}^{M} c_{j,p} \widetilde{V}_{t-j}^*$$

and

$$Z_M \sim N(0, \sigma_M^2), \quad \sigma_M^2 = \left(\sum_{j=0}^{M} c_{j,p} \right)^2 2\pi f_Y(0) \left| B_p(0) \right|^2.$$

(b) $Z_M \xrightarrow{D} Z$ as $M \to \infty$, where $Z \sim N(0, \sigma_\infty^2)$.

(c) $\lim_{M\to\infty} \limsup_{n\to\infty} P\left(|L_n - L_{n,M}| > \varepsilon \right) = 0$ for every $\varepsilon > 0$.

To establish (a) write

$$L_{n,M} = \sum_{j=0}^{M} c_{j,p} \frac{1}{\sqrt{N}} \sum_{t=1}^{N} \widetilde{V}_t^* - \sum_{j=0}^{M} c_{j,p} \frac{1}{\sqrt{N}} \sum_{s=0}^{M-1} (M-s) \widetilde{V}_{n-s}^*. \tag{3.15}$$

The second term on the right-hand side of (3.15) is in absolute value bounded by

$$\sum_{j=0}^{M} |c_{j,p}| \frac{M^2}{\sqrt{N}} \frac{1}{M} \sum_{s=0}^{M-1} \left| \widetilde{V}_{n-s}^* \right| = \sum_{j=0}^{M} |c_{j,p}| \frac{M^2}{\sqrt{N}} O_p(1) = O_p\left(\frac{1}{\sqrt{n}} \right).$$

Thus, the limiting distribution of $L_{n,M}$ is identical to the limiting distribution of the first term on the right hand side of (3.15). Let,

$$L_{n,M}^{(1)} = \sum_{j=0}^{M} c_{j,p} \frac{1}{\sqrt{N}} \sum_{t=1}^{N} \widetilde{V}_t^*.$$

Notice that,

$$E^*\left(L_{n,M}^{(1)} \right) = 0,$$

and

$$Var^*\left(L_{n,M}^{(1)}\right) = \left(\sum_{j=0}^{M} c_{j,p}\right)^2 \sum_{h=-(N-1)}^{N-1} \frac{1}{N} \sum_{t=1\vee(1-h)}^{N\wedge(N-h)} \left(V_{t,p} - \overline{V}_n\right)\left(V_{t+h,p} - \overline{V}_n\right) w\left(\frac{h}{l_n}\right)$$
$$= \left(\sum_{j=0}^{M} c_{j,p}\right)^2 2\pi \widehat{f}_V(0).$$

Since

$$2\pi \widehat{f}_V(0) \xrightarrow{P} 2\pi f_V(0), \quad \text{as} \quad n \to \infty$$

and using the relation

$$f_V(0) = |B_p(0)|^2 f_Y(0), \quad \text{where} \quad B_p(0) = 1 - \sum_{j=1}^{p} b_{j,p},$$

we get that

$$Var^*\left(L_{n,M}^{(1)}\right) \to \sigma_M^2, \quad \text{as} \quad n \to \infty,$$

where

$$\sigma_M^2 = \left(\sum_{j=0}^{M} c_{j,p}\right)^2 2\pi f_Y(0) |B_p(0)|^2.$$

We next show that

$$\frac{1}{\sqrt{N}} \sum_{t=1}^{N} \widetilde{V}_t^* \xrightarrow{D} N(0, 2\pi f_V(0)).$$

Since

$$\frac{1}{\sqrt{N}} \sum_{t=1}^{N} \widetilde{V}_t^* = \frac{1}{\sqrt{N}} \sum_{t=1}^{N} V_{t,p} W_t^* + O_P\left(\sqrt{\frac{l_n}{n}}\right),$$

and W_t^* is l_n-dependent, we apply Theorem 1.21 of Kreiss and Paparoditis (2017). For this, it suffices to show that the following conditions are satisfied.

(i) $\Sigma_n = Var^*\left(N^{-1/2} \sum_{t=1}^{N} V_{t,p} W_t^*\right) \to 2\pi f_V(0)$, in probability.

(ii) $\sup_{a,r} A_{a,r} = \sup_{a,r} \frac{1}{r} E^*\left(N^{-1/2} \sum_{t=a}^{a+r-1} V_{t,p} W_t^*\right)^2 = O_P(1/N)$.

(iii) $\max_{1\le t\le N} E^*\left|N^{-1/2} V_{t,p} W_t^*\right|^{2+\delta} = O_P\left(N^{-1-\delta/2}\right)$, for some $\delta > 0$.

(iv) $l_n^{2+2/\delta}/N \to 0$, as $n \to \infty$.

To see (i) and (ii) observe that

$$\Sigma_n = \frac{1}{N} \sum_{t,s=1}^{N} V_{t,p} V_{s,p} w\left(\frac{s-t}{l_n}\right)$$

$$= \frac{1}{N} \sum_{h=-l_n}^{l_n} \sum_{t=1}^{N-h} w\left(\frac{h}{l_n}\right) V_{t,p} V_{t+h,p}$$

$$= 2\pi \widehat{f}_V(0) + O_P\left(\frac{l_n}{N}\right) \xrightarrow{P} 2\pi f_V(0),$$

and that

$$Var(A_{a,r}) = Var\left(\frac{1}{rN} \sum_{t,s=a}^{a+r-1} V_{t,p} V_{s,p} w\left(\frac{s-t}{l_n}\right)\right)$$

$$= \frac{1}{r^2 N^2} \sum_{t,s,h,g=a}^{a+r-1} w\left(\frac{s-t}{l_n}\right) w\left(\frac{h-g}{l_n}\right) \gamma_V(t-h)\gamma_V(s-g)$$

$$+ \frac{1}{r^2 N^2} \sum_{t,s,h,g=a}^{a+r-1} w\left(\frac{s-t}{l_n}\right) w\left(\frac{h-g}{l_n}\right) \gamma_V(t-g)\gamma_V(s-h)$$

$$+ \frac{1}{r^2 N^2} \sum_{t,s,h,g=a}^{a+r-1} w\left(\frac{s-t}{l_n}\right) w\left(\frac{h-g}{l_n}\right) k_V(s-t, h-t, g-t)$$

$$\leq \frac{2}{N^2}\left(\sum_{h=-\infty}^{\infty} |\gamma_V(h)|\right)^2 + \frac{1}{N^2}\left(\sum_{u,v,l=-\infty}^{\infty} |k_V(u,v,l)|\right)^2 = O\left(\frac{1}{N^2}\right).$$

Consider (iii). We have

$$\max_{1\leq t\leq N} E^* \left|N^{-1/2} V_{t,p} W_t^*\right|^{2+\delta} = \left(N^{-1/2}\right)^{2+\delta} \max_{1\leq t\leq N} |V_{t,p}|^{2+\delta} E^* |W_1^*|^{2+\delta} = O_p\left(N^{-1-\delta/2}\right).$$

Finally, (iv) follows directly from a corresponding assumption of the theorem. So far, we have shown that, as $n \to \infty$,

$$L_{n,M} \xrightarrow{D} Z_M \quad \text{where} \quad Z_M \sim N(0, \sigma_M^2),$$

which concludes the proof of (a).

Consider assertion (b). This assertion follows since as $M \to \infty$,

$$\sigma_M^2 \to \left(\sum_{j=0}^{\infty} c_{j,p}\right)^2 2\pi f_Y(0) |B_p(0)|^2 = |B_p(0)|^{-2} 2\pi f_Y(0) |B_p(0)|^2 = 2\pi f_Y(0) = \sigma_\infty^2.$$

To verify assertion (c) we proceed as follows. Write,

$$L_n - L_{n,M} = \frac{1}{\sqrt{N}} \sum_{t=1}^{N} \left\{\sum_{j=0}^{t-1} c_{j,p} \widetilde{V}_{t-j}^* - \sum_{j=0}^{M} c_{j,p} \widetilde{V}_{t-j}^*\right\}$$

$$= \frac{1}{\sqrt{N}} \sum_{t=M+2}^{N} \sum_{j=M+1}^{t-1} c_{j,p} \widetilde{V}_{t-j}^*$$

$$= \frac{1}{\sqrt{N}} \sum_{j=M+1}^{N-1} c_{j,p} \sum_{r=1}^{N-j} \widetilde{V}_r^*.$$

By Markov's inequality

$$
\begin{aligned}
P^{*}\left(|L_n - L_{n,M}| > \varepsilon\right) &\leq \frac{E^{*}(L_n - L_{n,M})^2}{\varepsilon^2} \\
&= \frac{1}{\varepsilon^2 N} \sum_{j,l=M+1}^{N-1} c_{j,p} c_{l,p} \sum_{r=1}^{N-j} \sum_{s=1}^{N-l} w\left(\frac{r-s}{l_n}\right) \left(V_{r,p} - \overline{\overline{V}}_n\right) \left(V_{s,p} - \overline{\overline{V}}_n\right) \\
&= T_{5,N} \text{ (say).}
\end{aligned}
$$

Notice that,

$$
\begin{aligned}
|E(T_{5,N})| &\leq \frac{1}{\varepsilon^2 N} \sum_{j,l=M+1}^{N-1} |c_{j,p}|\,|c_{l,p}| \sum_{r,s=1}^{N} \left| E\left(V_{r,p} - \overline{\overline{V}}_n\right) \left(V_{s,p} - \overline{\overline{V}}_n\right) \right| \\
&\leq \frac{1}{\varepsilon^2 N} \sum_{j,l=M+1}^{N-1} |c_{j,p}|\,|c_{l,p}| \sum_{r,s=1}^{N} |\gamma_V(r-s)| \\
&\quad + \frac{2}{\varepsilon^2 (N-p)} \sum_{j,l=M+1}^{N-1} |c_{j,p}|\,|c_{l,p}| \sum_{r=p+1}^{N} \sum_{s=1}^{N} |\gamma_V(r-s)| \\
&\quad + \frac{N}{\varepsilon^2 (N-p)^2} \sum_{j,l=M+1}^{N-1} |c_{j,p}|\,|c_{l,p}| \sum_{r,s=p+1}^{N} |\gamma_V(r-s)| \\
&= T_{6,N} + T_{7,N} + T_{8,N},
\end{aligned}
$$

with an obvious notation for $T_{6,N}$, $T_{7,N}$ and $T_{8,N}$. The term $T_{6,N}$ is bounded by

$$
\frac{1}{\varepsilon^2} \left(\sum_{j=M+1}^{N-1} |c_{j,p}| \right)^2 \sum_{h=-\infty}^{\infty} |\gamma_V(h)|,
$$

which converges to zero as M and $n \to \infty$, since the sums

$$
\sum_{j=0}^{\infty} |c_{j,p}| \quad \text{and} \quad \sum_{h=-\infty}^{\infty} |\gamma_V(h)|
$$

are finite. Similar arguments yield that the terms $T_{7,N}$ and $T_{8,N}$ converge also to zero as M and $n \to \infty$. This together with the relation

$$
L_n^{*} = L_n + o_P(1)
$$

concludes the proof that

$$
L_n^{*} \xrightarrow{D} N(0, \sigma_\infty^2)
$$

in probability. The assertion of the theorem follows then by an application of the δ-method and taking into account Assumption 3.3. □

Chapter 4

Conclusions and Future Research

4.1 Conclusions

In the first part of this thesis, we have investigated the problem of estimating the rescaled fourth order cumulant of the unobserved innovations of a linear time series. An existing nonparametric estimator of this parameter has been investigated. It has been shown how the behavior of this estimator is affected by the autocorrelation structure of the underlying process. An improved nonparametric estimator of the same parameter has been proposed which is based on pre-whitening the time series by means of an autoregressive filter. The parameter of interest is estimated using the filtered time series and an inverse-transformation is not required. This is due to an invariance property of the parameter of interest with respect to linear filtering.

The asymptotic properties of the new estimator have been investigated and its superiority has been shown for large classes of stochastic processes. Some simulations demonstrated that this theoretical superiority is also valid in finite sample situations. Our findings indicate, that the gains in terms of variance and bias reduction obtained by using the new estimator could be very impressive, especially for strongly correlated time series.

In the second part, we have proposed two modifications of the autoregressive-sieve respectively of the autoregressive bootstrap. First, an AR-sieve bootstrap procedure has been proposed where the pseudo innovations are not obtained by i.i.d. resampling from the empirical distribution of the estimated residuals but from some appropriate distribution ensuring that the generated i.i.d. pseudo-innovations imitate asymptotically correct also the rescaled fourth order cumulant of the true innovations. Next, a new version of the AR-bootstrap applied to an appropriately transformed time series together with a dependent-wild type generation of pseudo-innovations has been proposed. We show that this AR-bootstrap procedure is asymptotically valid for a wide range of weakly dependent processes and for large classes of statistics. A fully data driven procedure to select the parameters involved in both bootstrap procedures has been proposed. Extensive simulations and comparisons show a good finite sample behavior of the new bootstrap procedures proposed.

4.2 Future Research

4.2.1 Locally Stationary Processes

Stochastic processes with time varying characteristics have attracted considerable interest during the last decades. An important approach for the development of an asymptotic theory for such processes has been put forward by the concept of locally stationary processes introduced by Dahlhaus (1997). Loosely speaking, a stochastic process is locally stationary if it can be locally (in time) approximated by some stationary process. More precisely, a triangular array of sequences of random variables $\{X_{t,n} : t = 1, ..., n, n \in \mathbb{N}\}$ is called locally stationary, if it satisfies the following set of conditions:

(a) $X_{t,n}$ has the representation

$$X_{t,n} = \sum_{j=-\infty}^{\infty} \psi_{t,n}(j)\, \varepsilon_{t-j}, \quad t = 1, ..., n, \quad n \in \mathbb{N}$$

where the $\{\varepsilon_t\}$ are i.i.d with $E(\varepsilon_t) = 0, E(\varepsilon_t^2) = 1$ and $\kappa_{4,\varepsilon} = E(\varepsilon_t^4) - 3$.

(b)

$$\sup_{t=1,...,n} |\psi_{t,n}(j)| \leq \frac{k}{l^2(j)} \quad \text{for} \quad \text{all} \quad j \in \mathbb{Z},$$

where k is a non-negative constant independent of n and $\{l(j) : l \in \mathbb{Z}\}$ is a positive sequence satisfying

$$\sum_{j=-\infty}^{\infty} |j| l^{-1}(j) < \infty.$$

(c) Functions $\psi_j(\cdot) : (0,1] \to \mathbb{R}$ with

$$\sup_{u\in[0,1]} |\psi_j(u)| \leq \frac{k}{l(j)}$$

and

$$\sup_{u\in[0,1]} \left| \frac{\partial \psi_j(u)}{\partial u} \right| \leq \frac{k}{l(j)}$$

exist such that

$$\sup_{1\leq t\leq n} \left| \psi_{t,n}(j) - \psi_j\left(\frac{t}{n}\right) \right| \leq \frac{k}{nl(j)},$$

where $l(j)$, $j \in \mathbb{Z}$ and k are as above.

Consider the problem of estimating the rescaled fourth order cumulant of the unobserved innovations $\{\varepsilon_t, t \in \mathbb{Z}\}$, driving the above locally stationary linear process, that is of

$$\eta_{4,\varepsilon} = \kappa_{4,\varepsilon}/\sigma_\varepsilon^4 \quad \text{where} \quad \kappa_{4,\varepsilon} = \text{cum}_4(\varepsilon_t) = E(\varepsilon_t^4) - 3\sigma_\varepsilon^4$$

is the fourth order cumulant of ε_t. Toward this, recall for $X_{t,n}$ the local approximating linear process

$$X_t(u) = \sum_{j=-\infty}^{\infty} \psi_j(u)\, \varepsilon_{t-j}, \quad \text{where} \quad u \in [0,1].$$

Let $c(u,k) = Cov(X_t(u), X_{t+k}(u))$ and $c_2(u,k) = Cov\left(X_t^2(u), X_{t+k}^2(u)\right)$ be the local autocovariances of the process $X_t(u)$ and of the squared process $X_t^2(u)$ respectively. Straightforward calculations yield

$$Cov\left(X_t^2(u), X_{t+k}^2(u)\right) = \kappa_{4,\varepsilon} \sum_{j=-\infty}^{\infty} \psi_j^2(u)\, \psi_{j+k}^2(u) + 2Cov^2(X_t(u), X_{t+k}(u)).$$

Taking the sum over all $k \in \mathbb{Z}$, using the fact that

$$\sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j^2(u)\, \psi_{j+k}^2(u) = c^2(u,0)$$

and integrating both sides over the interval $[0,1]$, we end-up with the expression

$$\eta_{4,\varepsilon} = \frac{\sum\limits_{k=-\infty}^{\infty} \left(\int_0^1 c_2(u,k)\, du - 2\int_0^1 c^2(u,k)\, du \right)}{\int_0^1 c^2(u,0)\, du}.$$

The above expression motivates the following estimator of $\eta_{4,\varepsilon}$

$$\widehat{\eta}_{4,\varepsilon} = \frac{\sum_{k=-M_n}^{M_n} w\left(k/M_n\right)\left(\widehat{c}_2\left(k\right) - 2\widehat{c}^2\left(k\right)\right)}{\widehat{c}^2\left(0\right)}, \tag{4.1}$$

where

$$\widehat{c}\left(k\right) = \frac{1}{LN}\sum_{j=1}^{L}\sum_{t=1}^{N-k}\left(X_{[u_jn]+t-N/2,n} - \overline{X}_N(j,t)\right)\left(X_{[u_jn]+t-N/2+k,n} - \overline{X}_N(j,t)\right),$$

$$\widehat{c}_2\left(k\right) = \frac{1}{LN}\sum_{j=1}^{L}\sum_{t=1}^{N-k}\left(X^2_{\left[u_jn\right]+t-N/2,n} - \overline{X}_{2,N}(j,t)\right)\left(X^2_{\left[u_jn\right]+t-N/2+k,n} - \overline{X}_{2,N}(j,t)\right),$$

$$\overline{X}_N(j,t) = \frac{1}{N_s}\sum_{r=1}^{N_s} X_{[u_jn]+t-N/2+r,n} \quad \text{and} \quad \overline{X}_{2,N}(j,t) = \frac{1}{N_s}\sum_{r=1}^{N_s} X^2_{\left[u_jn\right]+t-N/2+r,n},$$

$N_s = \sum_{s=1}^{N} \mathcal{I}([u_jn] + t - N/2 + r + s \in \{1, 2, ..., n\})$, are sample estimates of the corresponding unknown quantities appearing in (4.1) and $\mathcal{I}\left(\cdot\right)$ is the indicator function. Furthermore, w is a lag-window, $M_n < n$ is a truncation parameter, N is the local window width and $u_j = [N(j-1) + N/2]/n$, $j = 1, ..., L$ are rescaled time points in the interval $[0, 1]$. A similar, frequency domain based estimator has been proposed by Kreiss and Paparoditis (2015).

A probably improved estimator of $\eta_{4,\varepsilon}$ can be obtained by using locally the idea of pre-whitening, i.e., by fitting locally to the time series a p^{th} order autoregressive process. To elaborate, suppose that the local spectral density $f\left(u, \lambda\right)$ is continuous in λ and satisfies

$$\inf_{u\in[0,1]}\inf_{\lambda\in[o,\pi]} f\left(u, \lambda\right) \geq c_f \quad \text{where} \;\; c_f > 0.$$

Then, the local approximating process $\{X_t\left(u\right), t \in \mathbb{Z}\}$ has for every $u \in [0, 1]$ the autoregressive representation

$$X_t\left(u\right) = \sum_{r=1}^{\infty} b_r\left(u\right) X_{t-r}\left(u\right) + \varepsilon_t, \tag{4.2}$$

where

$$\sum_{r=1}^{\infty} r\left|b_r\left(u\right)\right| < \infty \quad \text{and} \quad 1 - \sum_{r=1}^{\infty} b_r\left(u\right) z^r \neq 0$$

for all complex z with $|z| \leq 1$. The minimization of the local quadratic deviation

$$\frac{1}{N-p}\sum_{j=p}^{N-1}\left(X_{[un]-N/2+j,n} - \sum_{i=1}^{p} c_i\left(u\right) X_{[un]-N/2+j-i,n}\right)^2$$

with respect to $c_i(u), 1 = 1, 2, ..., p$ leads to the estimates

$$\widehat{b}_u(p)^\top = \left(\widehat{b}_1(u), \widehat{b}_2(u), ..., \widehat{b}_p(u)\right),$$

satisfying the system of equations

$$\widehat{R}_u(p)\,\widehat{b}_u(p) = \widehat{r}_u(p),$$

where

$$\widehat{R}_u(p) = \frac{1}{N-p}\sum_{j=p}^{N-1} X_j(u,p)\,X_j(u,p)^\top, \quad \widehat{r}_u(p) = \frac{1}{N-p}\sum_{j=p}^{N-1} X_j(u,p)\,X_{[un]-N/2+j,n}$$

and

$$X_j(u,p)^\top = \left(X_{[un]-N/2+j,n}, X_{[un]-N/2+j-1,n}, ..., X_{[un]-N/2+j-p,n}\right).$$

Furthermore, let

$$\widehat{\sigma}_p^2(u) = \frac{1}{N-p}\sum_{j=p}^{N-1} X^2_{[un]-N/2+j-p,n} - \widehat{b}_u(p)^\top\,\widehat{r}_u(p),$$

be the estimated variance of the residuals of the local autoregressive fit. The alternative estimator of $\eta_{4,\varepsilon}$ we propose can then be obtained as follows.

Step 1: For $t/n \in [0,1]$, fit locally an autoregressive model of order p to the observations $X_{1,n}, X_{2,n}, ..., X_{n,n}$, calculate the estimated parameters

$$\widehat{b}_{t/n}(p)^\top = \left(\widehat{b}_1(t/n), \widehat{b}_2(t/n), ..., \widehat{b}_p(t/n)\right)$$

and the residual variance $\widehat{\sigma}_p^2(t/n)$. Consider, then the rescaled residuals

$$\widehat{U}_{t,n} = \frac{1}{\widehat{\sigma}_p(t/n)}\left(X_{t,n} - \sum_{i=1}^{p}\widehat{b}_i(t/n)\,X_{t-i,n}\right), \quad t = p+1, ..., n.$$

Step 2: Using $\widehat{U}_{p+1,n}, \widehat{U}_{p+2,n}, ..., \widehat{U}_{n,n}$, calculate the estimator

$$\widehat{\eta}_4^S = \frac{1}{\widehat{\gamma}_U^2(0)}\sum_{h=-(N_p-1)}^{N_p-1} w\left(\frac{h}{M_n}\right)\left(\widehat{\gamma}_{2,U}(h) - 2\widehat{\gamma}_U^2(h)\right), \quad N_p = n-p;$$

see (2.7). Here,

$$\widehat{\gamma}_U(h) = \frac{1}{N_p}\sum_{t=p+1}^{n-|h|}\left(\widehat{U}_{t,n} - \overline{U}_n\right)\left(\widehat{U}_{t+|h|,n} - \overline{U}_n\right), \quad \overline{U}_n = \frac{1}{N_p}\sum_{t=p+1}^{n}\widehat{U}_{t,n},$$

$$\widehat{\gamma}_{2,U}(h) = \frac{1}{N_p}\sum_{t=p+1}^{n-|h|}\left(\widehat{U}^2_{t,n} - \overline{U}_{2,n}\right)\left(\widehat{U}^2_{t+|h|,n} - \overline{U}_{2,n}\right), \ \overline{U}_{2,n} = \frac{1}{N_p}\sum_{t=p+1}^{n}\widehat{U}^2_{t,n}.$$

and $w(\cdot)$ is a lag-window; see Assumption 2.2.

The motivation behind this procedure is the following. The innovations

$$\varepsilon_t = X_t(u) - \sum_{r=1}^{\infty} b_r(u) X_{t-r}(u),$$

do not depend on $u \in [0,1]$. Therefore, filtering the time series locally by an autoregressive process, intends to obtain local residuals $\widehat{U}_{t,n}$ which (asymptotic) will behave like ε_t. Thus, $\widehat{U}_{t,n}$, $t = p+1, ..., n$, can be used to estimate the rescaled fourth order cumulant of the unobserved innovations. Asymptotic properties of the above estimator can be investigated. Furthermore, the finite sample behavior of these estimators can be numerically compared by means of simulations with alternative estimators like the one proposed by Kreiss and Paparoditis (2015).

4.2.2 Multivariate Processes

An important but probably difficult to solved problem of future research is the estimation of the fourth order cumulant of the unobserved innovations for a multivariate linear time series. Solving this problem, also is important of extending the range of validity of the multivariate AR-sieve bootstrap; see Meyer and Kreiss (2015).

To elaborate, let $\{\underline{X}_t : t \in \mathbb{Z}\}$ be a $m-$dimensional stochastic process generated by

$$\underline{X}_t = (X_{1,t}, X_{2,t}, ..., X_{m,t})^\top = \sum_{j=-\infty}^{\infty} \Psi_j \underline{\varepsilon}_{t-j},$$

where the $\Psi_j = (\psi_j(v,s))_{v,s=1,2,...,m}$, $j \in \mathbb{Z}$, are $m \times m$ coefficient matrices, and $\underline{\varepsilon}_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, ..., \varepsilon_{m,t})^\top, t \in \mathbb{Z}$ is a zero mean, $m-$dimensional i.i.d innovation process with finite fourth moment, i.e, $E\left(\varepsilon_{i,t}^4\right) < \infty$ for $i = 1, 2, ..., m$. It seems difficult to obtain a consistent estimator of the fourth order cumulant

$$\eta_{i,j,r,s} = Cum\left(\varepsilon_{i,t}, \varepsilon_{j,t}, \varepsilon_{r,t}, \varepsilon_{s,t}\right), \quad i,j,r,s \in \{1,2,..,m\},$$

along the lines used in the univariate context. An alternative strategy will be to apply the AR-dependent wild bootstrap to a multivariate setting.

Recall that the AR-bootstrap with dependent wild innovations, proposed in this thesis works by fitting an autoregressive model of order p, not to the observed time series $X_1, X_2, \ldots, X_n$ itself, but to the time series of transformed random variables $Y_1, Y_2, \ldots, Y_{n-m+1}$, where Y_t is given by $Y_t = g(X_t, X_{t+1}, \ldots, X_{t+m-1})$. An interesting multivariate extension could be to consider the transformed process

$$Y_t^{(i,j)} = X_{i,t} X_{j,t+h}, \quad i,j = 1, ..., m, \quad t \in \mathbb{Z}, \tag{4.3}$$

where $X_{i,t}$ and $X_{j,t}$ are the i^{th} and j^{th} components, respectively ,of the vector process $\{\underline{X}_t : t \in \mathbb{Z}\}$ with $\underline{X}_t = (X_{1,t}, X_{2,t}, ..., X_{m,t})^\top$. Denoting by $\gamma_{ij}(h) = Cov(X_{i,t}, X_{j,t+h})$ and $\rho_{ij}(h) = \gamma_{ij}(h) / \sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}$ the cross-covariance and cross-correlation function respectively, the problem is to estimate the distribution of the sample quantities

$$\widehat{\gamma}_{ij}(h) = \frac{1}{n} \sum_{t=1}^{n-h} \left(X_{i,t} - \overline{X}_{i,n}\right)\left(X_{j,t+h} - \overline{X}_{j,n}\right) \quad \text{and} \quad \widehat{\rho}_{ij}(h) = \frac{\widehat{\gamma}_{ij}(h)}{\sqrt{\widehat{\gamma}_{ii}(0)\widehat{\gamma}_{jj}(0)}},$$

where $\overline{X}_{i,n} = (1/n)\sum_{t=1}^{n} X_{i,t}$. The application of the autoregressive bootstrap with depended wild innovations to the transformed process $Y_t^{(i,j)}$ given in (4.3) will allow for the estimation of the distribution of $\sqrt{n}(\widehat{\gamma}_{ij}(h) - \gamma_{ij}(h))$ or $\sqrt{n}(\widehat{\rho}_{ij}(h) - \rho_{ij}(h))$ which is an interesting problem in multivariate time series analysis.

Bibliography

[1] Anderson, T.W. (1971). *The Statistical Analysis of Time Series.* Wiley, NewYork.

[2] Auestad, B. and Tjøstheim, D. (1990). Identification of Nonlinear Time Series: First Order Characterization and Order Determenation. *Biometrika* **77**, 669-687.

[3] Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control.* San Francisco, CA: Holden-Day.

[4] Brockwell, P. J. and Davis, R. A. (1988). Applications of Innovation Representations in Time Series Analysis. In *Probability and Statistics* (ed. J. N. Srivastara). Amsterdam: North-Holland.

[5] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* Springer, New York. Second Edition.

[6] Bühlmann, P. (1997). Sieve Bootstrap for Time Series. *Bernoulli* **3**, 123-148.

[7] Bühlmann, P. (2002). Bootstraps for Time Series. *Statistical Science* **17**, 52-72.

[8] Cavaliere, J., Georgiev, I. and R. Taylor (2013). Wild Bootstrap of the Sample Mean in the Infinite Variance Case. *Econometric Reviews* **32**, 204-219.

[9] Dahlhaus, R. (1985). Asymptotic Normality of Spectral Estimates. *Journal of Multivariate Analysis* **16**, 412-431.

[10] Dahlhaus, R. (1997). Fitting Time Series Models to Nonstationary Processes. *The Annals of Statistics* **25**, 1-37.

[11] Dahlhaus, R. and Janas, D. (1996). A Frequency Domain Bootstrap for Ratio Statistics in Time Series Analysis. *The Annals of Statistics* **24**, 1934-1963.

[12] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer Verlag, New York.

[13] Franke, J. and Härdle, W. (1992). On Bootstrapping Kernel Spectral Estimates. *The Annals of Statistics* **20**, 121-145.

[14] Goncalves, S. and Kilian, L. (2004). Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form. *Journal of Econometrics* **123**, 89-120.

[15] Goncalves, S. and Kilian, L. (2007). Asymptotic and Bootstrap Inference for AR(∞) Processes with Conditional Heteroskedasticity. *Econometric Reviews* **26**, 609-641.

[16] Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series.* Wiley, New York.

[17] Hannan, E. J. and Kavalieris, L. (1986). Regression, Autoregression Models. *Journal of Time Series Analysis* **7**, 27-49.

[18] Hansen, B. E. (2000). Testing for Structural Change in Conditional Models. *Journal of Econometrics* **97**, 93-115.

[19] Janas, D. and Dahlhaus, R. (1994). A Frequency Domain Bootstrap for Time Series. In: *Computationally Intensive Statistical Methods.* Proc. *26th Symposium on the Interface* (J. Stal and A. Lehman, eds.), 423-425.

[20] Keenan, D. M. (1987). Limiting Behaviour of Functionals of Higher-Order Sample Cumulant Spectra. *The Annals of Statistics* **15**, 134-151.

[21] Kreiss, J. P. (1988). *Asymptotical Inference for a Class of Stochastic Processes.* Habilitationsschrift, Univ. Hamburg.

[22] Kreiss, J. P. and J. Franke (1992). Bootstrapping Stationary Autoregressive Moving Average Models. *Journal of Time Series Analysis* **13**, 297-317.

[23] Kreiss, J. P. and Paparoditis, E. (2011). Bootstrap for Dependent Data: A Review, (with Discussion, and a rejoinder), *Journal of the Korean Statistical Society*, **40**, 357-378 and 393-395.

[24] Kreiss, J. P., Paparoditis, E. and Politis, D. N. (2011). On the Range of Validity of the Autoregressive Sieve Bootstrap. *The Annals of Statistics* **39**, 2103-2130.

[25] Kreiss, J. P. and Paparoditis, E. (2012). The Hybrid Wild Bootstrap for Time Series. *Journal of the American Statistical Analysis* **107**, 1073-1084.

[26] Kreiss, J. P. and Paparoditis, E. (2015). Bootstrapping Locally Stationary Processes. *Journal of the Royal Statistical Society, Series B,* **77**, 267-290.

[27] Kreiss, J. P. and Paparoditis, E. (2017). *Bootstrap for Time Series: Theory and Applications.* Manuscript.

[28] Künsch, H. R. (1989). The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics* **17**, 1217-1241.

[29] Meyer, M. and Kreiss, J. P. (2015). On the Vector Autoregressive Sieve Bootstrap. *Journal of Time Series Analysis* **36**, 377-397.

[30] Paparoditis, E. (2002). Frequency Domain Bootstrap for Time Series. In: *Empirical Process Techniques for Dependent Data.* (Eds. Dehling, H., Mikosch, T. and Sørensen, M.), Birkhäuser, Boston, pp. 365-381.

[31] Paparoditis, E. and Politis, D. N. (2001). Tapered Block Bootstrap. *Biometrika* **88**, 1105-1119

[32] Paparoditis, E. and Streitberg, B. (1991). Order Identification Statistics in Stationary Autoregressive Moving Average Models: Vector Autocorrelations and the Bootstrap. *Journal of Time Series Analysis* **13**, 415-435.

[33] Politis, D. N. (2003). The Impact of Bootstrap Methods on Time Series Analysis: Silver Anniversary of the Bootstrap. *Statistical Science* **18**, 219-230.

[34] Pourahmadi, M. (2001). *Foundations of Time Series Analysis and Prediction Theory.* Wiley, New York.

[35] Press, H. and Tukey, J. W. (1956). Power Spectral Methods of Analysis and their Application to Problems in Airplane Dynamics. *Flight Test Manual, NATO Advisory Group for Aeronautical Research and Development.* Chapter IV–c, 1-41.

[36] Rosenblatt, M. (1985). *Stationary Sequences and Random Fields.* Birkhäuser, Boston.

[37] Shao, X. (2010). The Dependent Wild Bootstrap. *Journal of the American Statistical Analysis* **105**, 218-235.

[38] Shumway, R. H. and Stoffer, D. S. (2006). *Time Series Analysis and Its Applications.* Springer, New York. Second Edition.

[39] Taniguchi, M. (1982). On Estimation of the Integrals of the Fourth-Order Cumulant Spectral Density. *Biometrika* **69**, 117-122.