Asymptotical Properties of Residual Bootstrap for Autoregressions

Jens-Peter Kreiss
Institut für Mathematische Stochastik
Technische Universität Braunschweig
Pockelsstrasse 14
D-38106 Braunschweig
Germany

April 17, 1997
Abstract

In this paper we deal with stationary autoregressive processes of finite or infinite but unknown order. Under fairly general assumptions we derive the asymptotic consistency of a usual residual bootstrap procedure for smooth functions of the empirical autocovariance and autocorrelation. Especially the order of the fitted autoregressive model is allowed to be data-dependent. Supplementary to the usual residual bootstrap we consider a wild bootstrap procedure. Some remarks concerning the asymptotic accuracy of the two proposed bootstrap procedures and a simulation study conclude the paper.
1. Introduction

Let us consider the following strictly stationary univariate autoregressive process $X = (X_t : t \in \mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\})$ which fulfills the following stochastic difference equation

$$X_t = \sum_{\nu=1}^{\infty} a_\nu X_{t-\nu} + \varepsilon_t , \ t \in \mathbb{Z},$$

where we assume for the white noise process $\varepsilon = (\varepsilon_t : t \in \mathbb{Z})$

(A1) $(\varepsilon_t : t \in \mathbb{Z})$ consists of independent and identically distributed (i.i.d.) random variables with $E\varepsilon = 0$ and $\text{Var} \varepsilon = E\varepsilon^2 = \sigma^2 > 0$. $F$ denotes the (cumulative) distribution function of $L(\varepsilon_1)$.

In order to obtain stationary solutions of (1.1) we assume for the parameter $\vartheta = (a_\nu : \nu \in \mathbb{N} = \{1, 2, \ldots\})$

(A2) $\sum_{\nu=1}^{\infty} |a_\nu| < \infty$ and $1 - \sum_{\nu=1}^{\infty} a_\nu z^\nu \neq 0$ for all complex $|z| \leq 1$.

The intention of this paper is to investigate asymptotical properties of residual-based bootstrap procedures for estimators based on empirical autocovariances, i.e.

$$\hat{\gamma}_T(h) = \frac{1}{T} \sum_{t=1}^{T-h} X_t X_{t+h} , \ h = 0, 1, 2, \ldots.$$ 

This class includes estimators for the theoretical autocorrelation, for the parameters of a fitted autoregression of given order $q$, say, for autoregressive spectral density estimates and smoothed periodogram estimators.

The main idea is to fit an autoregressive scheme of order $P$ to given observations $X_1, \ldots, X_T$ according to (1.1). The order $P$ may depend on the sample size $T$ as well as on the data themselves. Such an autoregressive fit supplies us with estimated residuals $\hat{\varepsilon}_t = X_t - \sum_{\nu=1}^{P} \hat{a}_\nu(P) X_{t-\nu}$. Denote by $\hat{F}_T^\gamma$ the corresponding centered empirical distribution function of the estimated residuals (cf. Section 2 for details). The usual residual bootstrap procedure is now defined as an autoregression with order $P$, parameters $\hat{a}_1(P), \ldots, \hat{a}_P(P)$ and i.i.d. innovations $\varepsilon_t^\gamma$ distributed according to $\hat{F}_T^\gamma$. This usual residual bootstrap turns out to be asymptotically consistent for estimators based on the empirical autocovariance as well as on the empirical autocorrelation function.

The idea of bootstrapping AR($\infty$)-processes was introduced in Kreiss (1988). It has been successfully applied to ARMA-processes of finite order (recall that invertible ARMA-processes have an autoregressive representation), cf. Kreiss and Franke (1992). Swanepoel and van Wyk (1986) applied this idea to spectral density estimates. In contrast to their paper we give a reasonable amount of theory for this specific situation in Section 6. Paparoditis and Streitberg (1992) applied the AR($\infty$)-bootstrap to order selection procedures in multivariate ARMA-models.
Recently Bühlmann (1995) considers this bootstrap proposal and compares it with the so-called blockwise bootstrap which was introduced by Künsch (1989). See also Bühlmann (1993) and Bühlmann and Künsch (1994). Bickel and Bühlmann (1995) consider in detail the probabilistic structure of this kind bootstrap. Especially, they obtain a new kind of mixing property for the bootstrap process.

In contrast to all mentioned papers we allow that the order of the fitted autoregressive model may depend on the data, i.e. it may be obtained from a usual order selection procedure. Moreover we consider different kinds of resampling schemes for our setup.

Franke and Härdle (1992) and Dahlhaus and Janas (1996) consider a completely different bootstrap approach for time series data, which is based on the periodogram. This approach yields consistent bootstrap approximations for the distribution of statistics which can be represented as functionals of the periodogram.

Künsch (1989), Bühlmann (1993) and Bühlmann and Künsch (1994) proposed the so-called blockwise-bootstrap for dependent observations. This approach does not need an autoregressive scheme for the data-generating process and is in this respect more flexible. On the other hand, if the underlying structure is of autoregressive type (including all invertible ARMA-structures) the proposal given in this paper is much easier to apply and theoretically much easier to handle.

Wu (1986) introduced the idea of wild bootstrap. In the context of autoregressive models it means to replace the complicated process structure of the data by a much simpler regression model with (conditionally) fixed design in the bootstrap world. In this case it suffices to use independent innovations \( \varepsilon_i \) distributed such that

\[
E^* \varepsilon_i = 0 \quad \text{and} \quad E^* (\varepsilon_i^k) = \hat{\varepsilon}_i^k (k = 2, 3).
\]

The situations in which this much simpler wild bootstrap works for autoregressions are more restricted and will be discussed in detail in Section 4.

The paper is organized as follows. Section 2 contains exact definitions of the bootstrap proposals which will be investigated in the paper. Asymptotic properties of the usual residual bootstrap are given in Section 3, while Section 4 is devoted to the wild bootstrap proposal. Some remarks concerning the accuracy of the bootstrap proposals will be given in Section 5. Section 6 deals with bootstrapping spectral density estimates. Using results from Section 3 and 4, we are able to show asymptotic consistency for such a proposal. Simulation results, which will be reported on in Section 7, demonstrate the finite sample properties of the bootstrap proposals for various estimators. Finally, Section 8 contains most of the proofs and some necessary auxiliary results.

2. The Bootstrap Procedures

Suppose that observations \( X_{1-p_{\max}(T)}, \ldots, X_T \) (\( p_{\max}(T) \) is defined below) of the underlying autoregressive process (1.1) are available. At first we fit an autoregression of order \( P(T) \equiv P \) to the data. \( P \) may depend on the data, e.g. the order \( P \) may be obtained from a usual
order selection criterion (AIC, BIC, Rissanen-Schwartz criterion and so on). To be as flexible as possible we merely assume throughout the whole paper that

\((A3)\) \(P = P(T) \in [p_{\text{min}}(T), p_{\text{max}}(T)]\), where the deterministic sequences 
p_{\text{min}}(T) and \(p_{\text{max}}(T)\) are assumed to fulfill \(p_{\text{max}}(T) \geq p_{\text{min}}(T) \to \infty\) and \(p_{\text{max}}(T)^7 \cdot (\log T)^7 / T^2 \to 0\) as \(T \to \infty\).

**Remark.** The assumptions on \(p_{\text{min}}(T)\) and \(p_{\text{max}}(T)\) are in full strength not really necessary for all asymptotic results. But we find it much more convenient to work with a single assumption on \(p_{\text{min}}(T)\) and \(p_{\text{max}}(T)\) throughout the whole paper. If for example \(X\) is an autoregression of finite order \(p_s\), say, (that is \(a_v = 0\) for all \(v > p_s\)) then it suffices that \(\liminf_{T \to \infty} p_{\text{min}}(T) \geq p_s\) holds.

Using the autoregressive fit of order \(P\) introduced above, we easily obtain estimates \(\hat{\varepsilon}_t = \hat{\varepsilon}_t(P)\) of the innovations \(\varepsilon_t\) according to

\[
\hat{\varepsilon}_t = X_t - \sum_{i=1}^{P} \hat{a}_i(P)X_{t-i}, \quad t = 1, \ldots, T.
\]

Here and in the following we use the parameter estimator

\[
\hat{\theta}_T(P) = (\hat{a}_1(P), \ldots, \hat{a}_P(P))^T,
\]

which is obtained from the well-known least-squares equations

\[
\hat{R}_T(P) \hat{\theta}_T(P) = \left( \sum_{i=1}^{T} X_iX_{i-1}, \ldots, \sum_{i=1}^{T} X_iX_{i-P} \right)^T.
\]

The \(P \times P\)-matrix \(\hat{R}_T\) is defined as

\[
\hat{R}_T = \left( \sum_{i=1}^{T} X_{i-i}X_{i-j} : i, j = 1, \ldots, P \right).
\]

Of course this least-squares estimator is closely related to the famous Yule-Walker estimator. More exactly the difference of both quantities is of order \(O_P(p_{\text{max}}(T)/T)\) uniformly in \(P(T) \in [p_{\text{min}}(T), p_{\text{max}}(T)]\), i.e. is asymptotically negligible.

In principle we could use any other \(\sqrt{T}\)–consistent procedure, but at several places it is quite convenient to work with this least-squares estimator because of its simple structure. Center the estimated innovations \(\hat{\varepsilon}_t\) around zero by subtracting their sample mean and denote the empirical distribution of the centered innovations by \(\hat{F}_T\).

The usual residual bootstrap process \(X^*_t = (X^*_t : t \in \mathbb{Z})\) is defined as an autoregression of order \(P\), parameter value \(\hat{\theta}_T(P)\) and i.i.d. innovations \((\varepsilon^*_t : t \in \mathbb{Z})\) distributed according to \(\hat{F}_T\), i.e.

\[
X^*_t = \sum_{i=1}^{P} \hat{a}_i(P)X^*_{t-i} + \varepsilon^*_t = \sum_{\nu=0}^{\infty} \hat{a}_\nu(P)\varepsilon^*_{t-\nu}, \quad t \in \mathbb{Z}.
\]
The moving average coefficients $\hat{\alpha}_\nu(P)$, which are formally defined as the power series coefficients of $\left(1 - \sum_{\nu=1}^P \hat{\alpha}_\nu(P)z^\nu\right)^{-1}$, can be computed recursively as follows ($\hat{\alpha}_0(P) = 1$)

$$
\hat{\alpha}_\nu(P) = \sum_{\mu=1}^{\nu\wedge P} \hat{\alpha}_\mu(P)\hat{\alpha}_{\nu-\mu}(P), \, \nu = 1, 2, \ldots .
$$

As we will see in Section 4 it is not necessary to define the bootstrap innovations $\varepsilon_t^*$ in such a complicated way. It suffices to ensure that they have on average the correct behaviour. Having the wild bootstrap idea in mind one could think of bootstrap innovations

$$
(2.5) \quad \varepsilon_t^* = \hat{\varepsilon}_t \cdot \eta_t, \, t = 1, \ldots , T
$$

for i.i.d. random variables ($\eta_t$) with zero mean and second and third moment equal to one. Then the bootstrap process is defined as follows ($X_t^s = 0$ for $s < 0$)

$$
(2.6) \quad X_t^+ = \sum_{\nu=1}^P \hat{\alpha}_\nu(P)X_{t-\nu}^+ + \varepsilon_t^+ = \sum_{\nu=0}^{t-1} \hat{\alpha}_\nu(P)\varepsilon_{t-\nu}^+, \, t = 1, \ldots , T
$$

So far both bootstrap proposals really generate processes in the bootstrap world. This is not the case for the following third proposal, namely applying the complete idea of wild bootstrap to autoregressive processes. On the basis of estimated innovations $\hat{\varepsilon}_t$ and bootstrap innovations $\varepsilon_t^*$ as above we define the wild bootstrap observations $X_t^+$ according to

$$
(2.7) \quad X_t^+ = \sum_{\nu=1}^P \hat{\alpha}_\nu(P)X_{t-\nu} + \varepsilon_t^+, \, t = 1, \ldots , T
$$

(2.7) is nothing else but a regression model with conditionally fixed design. We will see in Section 4 that this proposal is usually not consistent. The wild bootstrap yields a consistent procedure only if the true underlying order of the autoregression coincides with $P$ and if we are interested in the distribution of the parameter estimates. Even if we are interested in the distribution of the parameter estimate for an AR(1)-fit to data generated by an autoregression of order greater than one the wild bootstrap does not work. The reason for that is that the distribution which we intend to mimic by the bootstrap and even the asymptotic distribution really depends on the whole dependence structure of the process and this is not captured by the wild bootstrap.

3. Asymptotics for the usual Residual Bootstrap

Let us begin the considerations in this section with a proof of the consistency of the usual residual bootstrap (cf. (2.4)) for the empirical autocovariance and autocorrelation. For this purpose we will frequently make use of the metric $d_p$, which for two probability measures $P$ and $Q$ on $\mathbb{R}^k$ equipped with the Borel $\sigma-$field is defined as

$$
(3.1) \quad d_p(P,Q) = \inf \left( E\|X - Y\|^p \right)^{1/p}, \, p > 0
$$
where the infimum is over all \( \mathbb{R}^{2k} \)-random vectors \((X, Y)\) such that \( X \sim P \) and \( Y \sim Q \). An excellent reference for properties of \( d_p \) is Bickel and Freedman (1981). \( d_2 \) is called Mallow’s metric. A key result for the following is

**Proposition 3.1** For \( r = 2, 4 \) we have under the assumptions \((A1), (A2)\) and \((A3)\) together with \( E \varepsilon_1^{2r} < \infty \)

\[
d_r \left( \hat{F}^c_T, F \right) = \text{op}(1) .
\]

We now can state the asymptotic validity of the usual residual bootstrap with respect to empirical autocovariances.

**Theorem 3.1** Assume that \( E \varepsilon_1^8 < \infty \) and \((A1), (A2)\) and \((A3)\). Then we have

\[
d_2 \left( \mathcal{L} \left( \sqrt{T} (\hat{\gamma}_T^* (h) - \gamma^* (h))_{h=0}^K \right), \mathcal{L} \left( \sqrt{T} (\hat{\gamma}_T (h) - \gamma (h))_{h=0}^K \right) \right) = \text{op}(1) .
\]

\( \hat{\gamma}_T, \gamma \) denote the empirical and the theoretical autocovariance function of the underlying process, while \( \hat{\gamma}_T^* (h) = \frac{1}{T} \sum_{t=1}^{T-h} X_t^* X_{t+h}^* \), \( \gamma^* (h) = E X_t X_{t+h}^* \) denote the corresponding quantities for the bootstrap process \( X^* \).

Moreover we have the following asymptotic normality of the bootstrapped autocovariances

\[
\sqrt{T} (\hat{\gamma}_T^* (h) - \gamma^* (h) : h = 0, \ldots, K) \Rightarrow \mathcal{N}(0, V_K) \text{ in probability} .
\]

The asymptotic covariance matrix

\[
V_K = \left[ \left( \frac{E \varepsilon_1^4}{\sigma^4} - 3 \right) \gamma(i) \gamma(j) + \sum_{k=-\infty}^{\infty} (\gamma(k) \gamma(k-i+j) + \gamma(k+j) \gamma(k-i)) \right]_{i,j=0}^K
\]

is the same as for the empirical autocovariances (cf. Brockwell and Davis (1991), Proposition 7.3.2).

To this theorem we have the following immediate corollary, which we state without proof, because the arguments for the application of the so-called \( \Delta \)-method (cf. Brockwell and Davis (1991), Proposition 6.4.3) are completely routine.

**Corollary 3.1** Under the same assumptions as in Theorem 3.1 we have the following asymptotic normality for the bootstrapped autocorrelations

\[
\hat{r}_h^* = \hat{\gamma}_T^* (h) / \hat{\gamma}_T^* (0) , \ h = 0, 1, 2, \ldots .
\]

\[
\sqrt{T} (\hat{r}_h^* - r_h^* : h = 1, 2, \ldots, K) \Rightarrow \mathcal{N}(0, W_K) \text{ in probability} .
\]

Notice that \( r^* (h) := \gamma^* (h) / \gamma^* (0) = \hat{r}_h + \text{op}(T^{-1/2}) \), the empirical autocorrelation, for all \( h \). The covariance matrix \( W_K \) is of course equal to the asymptotic covariance of the usual empirical autocorrelations and can be found in Brockwell and Davis (1991), Theorem 7.2.1.
Remark. It is of some importance later on, that the covariance matrix $W_K$ of the empirical autocorrelations does not depend on the distribution of the innovations $\varepsilon_t$ at all. This is in contrast to the asymptotic covariance matrix $V_K$ of the empirical autocovariances, which depend on the second and fourth moments of the innovations.

We have seen in this section that the usual residual bootstrap for autoregressions with infinite order behaves as it should, at least with respect to smooth functions of the empirical autocovariances and the probably more relevant autocorrelations. We like to conclude this section with some remarks concerning parameter estimation.

Suppose that we are interested in estimates of the parameters of an autoregressive fit of given order $q$ to the data. The data themselves still may arise from an infinite order underlying model. The theoretical parameters $b_\mu$ of the fitted model are given by the following smooth function of the autocorrelations

$$
\begin{pmatrix}
    b_1 \\
    \vdots \\
    b_q
\end{pmatrix}
= \begin{pmatrix}
    \cdots & r_{1-j} \\
    \vdots & \ddots & \ddots \\
    r_{q-1} & \cdots & r_{q-j}
\end{pmatrix}^{-1}
\begin{pmatrix}
    r_1 \\
    \vdots \\
    r_q
\end{pmatrix}.
$$

If we consider plug-in estimates $\hat{b}_\mu$, which are defined as stated in (3.5) with $r_h$ replaced by the empirical autocorrelation $\hat{r}_h$, then the usual residual bootstrap is consistent for such parameter estimates. Moreover, such a result holds for any smooth function of the autocovariance or the autocorrelation function. To be precise, we obtain for the parameter estimates of an AR($q$)-fit as described above

**Corollary 3.2** Under the same assumptions as in Theorem 3.1 we have that the bootstrap distribution

$$
L \left( \sqrt{T} \left[ \left( \hat{b}_1, \ldots, \hat{b}_q \right) - (b_1, \ldots, b_q) \right] \middle| X_{1_{-\text{max}(T)}}, \ldots, X_T \right)
$$

converges weakly in probability to the same multivariate normal distribution as

$$
L \left( \sqrt{T} \left( \hat{b}_\mu - b_\mu : \mu = 1, \ldots, q \right) \right).
$$

Here

$$
\begin{pmatrix}
    \hat{b}_1 \\
    \vdots \\
    \hat{b}_q
\end{pmatrix}^T = \begin{pmatrix}
    \cdots & \hat{r}_{1-j} \\
    \vdots & \ddots & \ddots \\
    \hat{r}_{q-1} & \cdots & \hat{r}_{q-j}
\end{pmatrix}^{-1}
\begin{pmatrix}
    \hat{r}_1 \\
    \vdots \\
    \hat{r}_q
\end{pmatrix},
$$

whereas $(\hat{b}_1, \ldots, \hat{b}_q)$ are similarly defined (replace $\hat{r}_h$ by $r_h$ , cf. Corollary 3.1). We may replace $(\hat{b}_1, \ldots, \hat{b}_q)$ by $(\hat{b}_1, \ldots, \hat{b}_q)$, since the difference of both quantities is of smaller order than $T^{-1/2}$.

**Proof:** This result is a direct consequence of Corollary 3.1 and the so-called $\Delta-$method (cf. Brockwell and Davis (1991), Proposition 6.4.3).

In Section 6 of the paper we will discuss in more detail a much more complicated functional of the whole autocorrelation function. But before let us focus our attention to a wild bootstrap proposal for autoregressions.
4. Asymptotics for the Wild Bootstrap

The principal idea of a wild bootstrap procedure has been stated already in Section 2 of the paper. Let us begin this section with a negative result, namely that the wild bootstrap proposal usually leads to an inconsistent procedure. Since the wild bootstrap observations $X_t^+$ are derived from a regression model it makes no sense to consider autocorrelation structure. That is why we focus on parameter estimates, only. Assume that we intend to approximate the distribution of the parameter estimates of an AR($q$)-fit as was described at the end of the previous section.

A fitted model of order $q$ in the wild bootstrap world is defined by the following theoretical parameter values which we intend to estimate in the bootstrap world.

$$
\arg\min_{b_1, \ldots, b_q} E^+ \sum_{t=1}^{T} \left( X_t^+ - \sum_{\nu=1}^{q} b_\nu X_{t-\nu} \right)^2
$$

(4.1) \[ = \left( \frac{1}{T} \sum_{t=1}^{T} X_{t+1 - \nu} X_{t-\mu} : \nu, \mu = 1, \ldots, q \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} E^+ X_t^+ X_{t-\nu} : \nu = 1, \ldots, q \right)^T \]

Here $E^+ X_t^+ = \sum_{\nu=1}^{p} \hat{a}_\nu(P) X_{t-\nu}$. Since the $\hat{a}_\nu(P)$ are usual least squares estimators, cf. (2.3), we obtain by simple algebra that the $\hat{b}_\nu$, as defined above, coincide with the least-squares estimators of order $q$, i.e.

$$
\left( \hat{b}_1, \ldots, \hat{b}_q \right)^T = \left( \sum_{t=1}^{T} X_{t+1 - \nu} X_{t-\mu} \right)^{-1} \left( \sum_{t=1}^{T} X_t X_{t-\nu} \right)_{\nu=1, \ldots, q},
$$

whenever $P(T) \geq q$. Let us consider estimators for the $\tilde{b}_\nu$ in the wild bootstrap world which are defined as follows

$$
\left( \tilde{b}_1^+, \ldots, \tilde{b}_q^+ \right)^T := \left( \sum_{t=1}^{T} X_{t+1 - \nu} X_{t-\mu} \right)^{-1} \left( \sum_{t=1}^{T} X_t^+ X_{t-\nu} \right)_{\nu=1, \ldots, q}.
$$

The following proposition gives us the asymptotic distribution of these quantities.

**Proposition 4.2** Assume $Ez^4 < \infty$ and (A1)-(A3). Then we have in probability

$$
\mathcal{L} \left( \sqrt{T} \left[ \left( \tilde{b}_1^+, \ldots, \tilde{b}_q^+ \right) - \left( \hat{b}_1, \ldots, \hat{b}_q \right) \right] \bigg| X_1, \ldots, X_T \right) \Rightarrow \mathcal{N} \left( \mathbf{0}, \sigma^2 \cdot \Gamma(q)^{-1} \right),
$$

where $\Gamma(q) = \left( EX_{t+1} X_{t+1-\nu} \right)_{\nu, \mu=1, \ldots, q}$.

**Proof:** Let us denote $\tilde{\Gamma}_T(q) = \left( \frac{1}{T} \sum_{t} X_{t+1 - \nu} X_{t-\mu} : \nu, \mu = 1, \ldots, q \right)$. Then we have from (4.1)

$$
\sqrt{T} \left[ \left( \tilde{b}_1^+, \ldots, \tilde{b}_q^+ \right) - \left( \hat{b}_1, \ldots, \hat{b}_q \right) \right] = \tilde{\Gamma}_T^{-1}(q) \left( \frac{1}{\sqrt{T}} \sum_{t} \left( X_t^+ - E^+ X_t^+ \right) X_{t-\nu} \right)_{\nu=1, \ldots, q}.
$$
A usual CLT for triangular arrays of independent random variables (cf. Gänssler and Stute (1977), Korollar 9.2.9) now yields the desired result, since we can show that

$$
E^+ \left( \frac{1}{\sqrt{T}} \sum_i (X^+_i - E^+X^+_i) X_{t-\mu} \right) \left( \frac{1}{\sqrt{T}} \sum_i (X^+_i - E^+X^+_i) X_{t-\mu} \right)
= \frac{1}{T} \sum_i E^+ (X^+_i - E^+X^+_i)^2 X_{t-\mu} \xi_t \xi_t + \frac{1}{T} \sum_i \xi_t^2 X_{t-\mu} X_{t-\mu}, \text{ cf. (2.5)}
\rightarrow_{T \to \infty} \sigma^2 E X_{t-\mu} X_{t-\mu}.
$$

The last convergence follows by direct computation from Lemma 8.2, 8.3 and the ergodic theorem.

\[ \square \]

**Remark.** The asymptotic variance given in Proposition 4.2 is equal to the asymptotic variance of the Yule-Walker or the least-squares parameter estimates in AR(q)-models, but it is not equal to the asymptotic variance of Yule-Walker parameter estimates of an AR(q)-fit to autoregression of higher order. Even in the case where the underlying autoregression is of finite but higher order than q the wild bootstrap as introduced above is not consistent. The positive message is, that if we intend to fit an autoregression of order q to an underlying process of exactly this order, then the wild bootstrap is consistent. In this specific situation the wild bootstrap has the advantage over the usual bootstrap, that it even works for models with conditional heteroskedasticity. Let us state this assertion as a seperate result.

**Theorem 4.2** Assume that the underlying autoregression is of order \( q \in \mathbb{N} \), (A1)-(A3), \( E\xi^2_1 = 1 \), \( E\xi^4_1 < \infty \) and that we have heteroskedasticity, e.g.

\[ (4.3) \quad X_t = \sum_{\nu=1}^q b_\nu X_{t-\nu} + \sigma(X_{t-1})\xi_t, \quad t \in \mathbb{Z}. \]

\( \sigma : \mathbb{R} \to (0, \infty) \) denotes a bounded function, which is also bounded away from zero. Then we have in probability

\[ \mathcal{L} \left( \sqrt{T} \left[ \left( \hat{b}^+_1, \ldots, \hat{b}^+_q \right) - (\hat{b}_1, \ldots, \hat{b}_q) \right] \right)
\Rightarrow \mathcal{N} \left( 0, \Gamma(q)^{-1} \left( E \sigma^2(X_0)X_{1-\nu}X_{1-\mu} \right)_{\nu,\mu=1,\ldots,q} \Gamma(q)^{-1} \right). \]

Since the asymptotic variance is equal to the asymptotic variance of the least-square parameter estimates, this means that the wild bootstrap proposal for AR(q)-parameter estimation is consistent for heteroskedastic autoregression of order q.

**Proof:** The arguments are quite similar to the ones given in the proof of Proposition 4.2, since

\[ \frac{1}{T} \sum_t \xi_t^2 X_{t-\nu} X_{t-\mu} = \frac{1}{T} \sum_t \xi_t^2 \sigma^2(X_{t-1})X_{t-\nu}X_{t-\mu} + o_T(1) \to_{T \to \infty} E \left( \sigma^2(X_0)X_{1-\nu}X_{1-\mu} \right), \]
in probability. To see the first equality we make use of (8.8) and the fact that the fourth moment of \( X_t \) is bounded. The last convergence is a usual ergodic property.

**Remark.** The assumptions on the function \( \sigma \) in Theorem 4.2 ensure strict stationarity and ergodicity of the model (cf. Masry and Tjøstheim (1995), Lemma 3.1). Without any doubt these assumptions can be relaxed. The asymptotic normality of the least-squares estimator in model (4.3) can be obtained from a usual CLT for martingale difference schemes (cf. Gänssler and Stute (1977), Satz 9.2.3) and the ergodic theorem.

If we want to modify the wild bootstrap in order to obtain a consistent procedure for more general situations we may proceed as follows. Let \( \varepsilon_t^\dagger \) be defined as in Section 2, cf. (2.5). Suppose now that bootstrap observations \( X_t^\dagger \) are given as

\[
X_t^\dagger = \sum_{\nu=1}^{P} \hat{\alpha}_\nu(P)X_{t-\nu}^\dagger + \varepsilon_t^\dagger .
\]

This bootstrap proposal preserves the whole dependence structure of the process. The bootstrap observations are no longer independent; for that reason one may hesitate to call the proposal a wild bootstrap procedure. Under suitable assumptions we obtain that this proposal is consistent for empirical autocorrelations (not for empirical autocovariances, which is perhaps of minor importance). This result ensures that the modified wild bootstrap proposal is an appropriate tool for approximating the distribution of a fitted autoregression of fixed order, \( q \) say, to the data even if the underlying model has a different autoregressive order, including infinity.

In the following we state rigorous results, which make these remarks precise. Let us start with an investigation of the empirical autocovariances for the modified wild bootstrap. As mentioned above, we do not obtain consistency for this bootstrap procedure. The main reason for this is that the asymptotic distribution of the empirical autocovariances depends on fourth moments of the innovations, which are not mimicked correctly by the construction of \( \varepsilon_t^\dagger \). Nevertheless we obtain as a corollary (stated as Theorem 4.3) the consistency of the modified wild bootstrap for empirical autocovariances.

**Proposition 4.3** Assume (A1)-(A3) and \( E\varepsilon_t^4 < \infty \). Then we have the following asymptotical distribution for the empirical autocovariances of the modified wild bootstrap process

\[
\left( \sqrt{T} \left( \hat{\gamma}_t^\dagger (h) - \gamma_t^\dagger (h) \right) : h = 0,1,2,\ldots,K \right) \Rightarrow \mathcal{N} \left( 0, V_K^\dagger \right) \text{ in probability,}
\]

where (recall the definition of \( X_t^\dagger \), cf. (2.6))

\[
\hat{\gamma}_t^\dagger (h) = \frac{1}{T} \sum_{t=1}^{T-h} E_t^\dagger X_t^\dagger X_{t+h}^\dagger = \frac{1}{T} \sum_{t=1}^{T-h} \sum_{\nu=0}^{i-1} \hat{\alpha}_\nu(P) \hat{\alpha}_{\nu+h}(P) \varepsilon_{t-\nu}^2 .
\]

The covariance matrix \( V_K^\dagger \) is given by the following entries \( r,s = 0,\ldots,K \)

\[
\left( E\gamma_t^h \left( \frac{E\varepsilon_t^4}{\sigma_t^4} - 2 \right) \gamma(r)\gamma(s) + \sum_{j=-\infty}^{\infty} (\gamma(j)\gamma(j-r+s) + \gamma(j+r)\gamma(j-s)) \right)
\]

9
Remark. \( V_k^t \) differs from the asymptotical covariance of the empirical autocovariances (cf. Theorem 3.1) by the factor \((E\eta_t^4 - 1)E\varepsilon_t^4/\sigma^4 - 2\) instead of \( E\eta_t^4/\sigma^4 - 3 \). Since \( E\eta_t^4 = 1 \) we have \( E\eta_t^4 \geq 1 \). Moreover, if we realize the distribution of \( \eta_t \) by the unique two-point distribution with \( E\eta_t = 0 \) and \( E\eta_t^3 = E\eta_t^3 = 1 \), then we obtain \( E\eta_t^4 = 2 \). In any case the modified wild bootstrap procedure is not consistent concerning autocovariances. But this is not a big problem, since for the much more relevant autocorrelations we obtain consistency for the modified wild bootstrap. This is the contents of the following theorem.

**Theorem 4.3** Under the same assumptions as in Proposition 4.3 we have the following asymptotical distribution for empirical autocorrelations \( \hat{r}_h^t = \hat{\gamma}_h^t(h)/\hat{\gamma}_h^t(0) \) of the modified wild bootstrap process (abbreviate \( \gamma_t^1(h)/\gamma_t^1(0) \) by \( \hat{r}_h^t \)).

\[
(4.8) \quad \left( \sqrt{T} \left( \hat{r}_h^t - r_h^t \right) : h = 1, 2, \ldots, K \right) = N(0, W_K) \quad \text{in probability.}
\]

\( W_K \) is defined in Corollary 3.1.

**Proof:** We obtain the assertion from Brockwell and Davis (1991), Proposition 6.4.3, since we have for the map \( g: IR^{K+1} \to IR^K, (x_0, x_1, \ldots, x_K) \to (x_1/x_0, \ldots, x_K/x_0) \) that \( \left( \hat{r}_1^t, \ldots, \hat{r}_K^t \right) = g \left( \hat{\gamma}_t^1(0), \ldots, \hat{\gamma}_t^1(K) \right) \).

Exactly along the lines of Section 3 (cf. Corollary 3.2) we easily obtain that the modified wild bootstrap is an appropriate tool for approximating the distribution of an fitted autoregression of fixed order, \( q \) say, to the data. More exactly we have the following corollary to Theorem 4.3.

**Corollary 4.3** Under the same assumptions as in Proposition 4.3 we have that the modified wild bootstrap distribution

\[
(4.9) \quad \mathcal{L} \left( \sqrt{T} \left[ \hat{b}_1^t, \ldots, \hat{b}_q^t \right] - \left( \hat{b}_1^t, \ldots, \hat{b}_q^t \right) \right) \bigg| X_{1-\max(T)}, \ldots, X_T
\]

converges weakly in probability to the same multivariate normal distribution as

\[
\mathcal{L} \left( \sqrt{T} \left( \hat{b}_\mu^t - \hat{b}_\mu : \mu = 1, \ldots, q \right) \right)
\]

Here

\[
\left( \hat{b}_1^t, \ldots, \hat{b}_q^t \right)^T = \left( \hat{r}_h^t \right)_{i,j=1}^{q-1}^{-1} \begin{pmatrix} \hat{r}_1^t \\ \vdots \\ \hat{r}_q^t \end{pmatrix}
\]

and \( (\hat{b}_1^t, \ldots, \hat{b}_q^t)^T \) is analogously defined by replacing \( \hat{r}_h^t \) by \( r_h^t \) (cf. Theorem 4.3 for a definition of \( r_h^t \)).
5. Accuracy of the Bootstrap Proposals

In a remarkable paper by Janas (Janas (1993 a,b)) we find conditions under which an Edgeworth expansion is valid for both the statistics of interest and their bootstrap counterparts. But note that Janas considered the so-called periodogram-based bootstrap procedure, which is quite different from the bootstrap proposals defined herein. Nevertheless, under the assumptions made in the paper by Janas, we obtain that we can approximate the following probability

\[ P \left\{ \sqrt{T} (\hat{\gamma}_T(0) - \gamma(0), \ldots, \hat{\gamma}_T(K) - \gamma(K)) \in C \right\} \]

(5.1)

uniformly over all measurable and convex subsets \( C \) of \( \mathbb{R}^{K+1} \) by an Edgeworth expansion \( \Psi_{T,3}(C) \) up to an error term of order \( o(T^{-1/2}) \).

Since autocorrelations are smooth functions of the sample autocovariances, e.g. \((\hat{r}_1, \ldots, \hat{r}_K) = (g_1(\hat{\gamma}_T(0), \ldots, \hat{\gamma}_T(K)), \ldots, g_K(\hat{\gamma}_T(0), \ldots, \hat{\gamma}_T(K)))\), we obtain for standardized expressions like

\[ \sqrt{T} V_{T,K}^{-1/2} (\hat{r}_1 - r_1, \ldots, \hat{r}_K - r_K) =: S_T, \]

(5.2)

where \( V_{T,K} \) denotes the \( K \times K \) covariance matrix of \( \sqrt{T}(\hat{r}_1 - r_1, \ldots, \hat{r}_K - r_K) \), the following Edgeworth approximation.

\[ \sup C \left| P \left\{ S_T \in C \right\} - \int_C \left( 1 + T^{-1/2} p_3(x) \right) d\Phi(x) \right| = o(T^{-1/2}) \]

(5.3)

where the coefficients of the polynomial \( p_3 \) depend continuously on the cumulants of order less than or equal to 3 of the following Taylor approximation (abbreviate \( G_T = \sqrt{T}(\hat{\gamma}_T(0) - \gamma(0), \ldots, \hat{\gamma}_T(K) - \gamma(K)) \))

\[ W_{T,3}^{(i)} = (\nabla g_i)_{(\gamma(0), \ldots, \gamma(K))} G_T^T + \frac{T^{-1/2}}{2!} G_T (D g_i)_{(\gamma(0), \ldots, \gamma(K))} G_T^T. \]

The supremum in (5.3) is over all measurable and convex sets \( C \) in \( \mathbb{R}^K \), cf. Janas (1993a), Lemma 5.5 or Bhatthacharya and Denker (1990), p. 22-32.

One main point in Janas (1993) was that the cumulants of \( W_{T,3}^{(i)} \) of order 3 or less do not depend on cumulants or moments of the innovations. Since our bootstrap proposals (usual bootstrap and modified wild bootstrap) are based on quite analogous autoregressive processes, we can compute formally similar edgeworth expansions of first order for these two bootstrap proposals. We obtain that the cumulants of order 3 or less of the corresponding Taylor approximation coincide (up to a sufficiently large order) with the cumulants of the non-bootstrapped quantities. This indicates that both bootstrap proposals catch the first higher order terms, which means that both bootstrap proposals outperform the normal approximation.

Since parameters of fitted autoregressive models of fixed order are smooth functions of the autocorrelation, a similar remark holds true for usual or modified wild bootstrap approximations of these statistics.

It should be stressed here that we do not have a rigorous proof of the validity of an Edgeworth expansion for our bootstrap proposals and that the results of Janas (1993) for bootstrap quantities do not cover any of our proposals.
6. Bootstrapping Spectral Density Estimators

In Section 5 we already mentioned the papers Janas (1993 a,b). Janas proposed a periodogram-based bootstrap procedure which is able to mimic the behaviour of statistical quantities which can be written as functionals of the periodogram. In this context we also refer to a paper by Franke and Härdle (1992) where the same bootstrap proposal as in Janas is applied to smoothed periodogram estimators of the spectral density. Let us discuss in this section how the bootstrap proposals of Section 3 and 4 apply to spectral density estimation. Most popular estimates of the spectral density are obtained from smoothing the so-called periodogram. As is argued for example in Brockwell and Davis (1991), Chapter 10.4, smoothed periodogram estimators are quite similar to so-called lag window estimators defined as follows

\[
\hat{\varphi}_T(\lambda) = \frac{1}{2\pi} \sum_{|h| \leq M} w \left( \frac{h}{M} \right) \hat{\gamma}_T(h) e^{-ih\lambda}, \quad 0 \leq \lambda \leq \pi,
\]

where \( w : [-1, 1] \to [0, \infty) \) denotes a continuous window function, which is assumed to satisfy

(W) \( 0 \leq w(u) \leq w(0) = 1, \quad w(u) = w(-u) \) for all \( |u| \leq 1 \) and \( \lim_{u \to 0} \frac{1}{|u|^q} w(u) > 0 \) for some \( q > 0 \)

Asymptotic normality of \( \hat{\varphi}_T \) is given in the monograph of Anderson (1971), Theorem 9.4.1. The result is as follows.

**Proposition 6.4** Assume (A1), (A2), \( E\varepsilon_1^4 < \infty, \sum_h h^2 |\gamma(h)| < \infty \) and that \( w : [-1, 1] \to [0, \infty) \) is continuous and satisfies (W). Then we have for \( M(T) \to \infty \) but \( M(T)/T \to 0 \) and \( T/M(T)^5 \to 0 \) as \( T \to \infty \)

\[
\sqrt{\frac{T}{M(T)}} (\hat{\varphi}_T(\lambda) - \varphi(\lambda)) \Rightarrow N(0, \tau^2(\lambda)) ,
\]

where

\[
\tau^2(\lambda) = \begin{cases} 
2\varphi^2(\lambda) \int_{-1}^{1} w^2(u) \, du & \text{if } \lambda = 0, \pi \\
\varphi^2(\lambda) \int_{-1}^{1} w^2(u) \, du & \text{if } 0 < \lambda < \pi
\end{cases}.
\]

\( \varphi : [0, \pi] \to IR \) denotes the underlying spectral density of the process, which has the following representation

\[
\varphi(\lambda) = \frac{E\varepsilon_1^2}{2 \pi} \left| 1 - \sum_{\nu=1}^{\infty} a_{\nu} e^{-i\nu\lambda} \right|^2 = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\lambda}.
\]

**Proof:** Cf. Anderson (1971), Theorem 9.4.1. The representation (6.4) of the spectral density \( \varphi \) for AR(\( \infty \))-processes can be found in Brockwell and Davis (1991), Corollary 4.3.2 and Theorem 4.4.1. \( \blacksquare \)
We have the following representation of $\hat{\phi}_T(\lambda) - \phi(\lambda)$ (cf. (6.1) and (6.4))

$$
\sqrt{\frac{T}{M}} (\hat{\phi}_T(\lambda) - \phi(\lambda)) \\
= \sqrt{\frac{T}{M}} \frac{1}{2\pi} \sum_{|i| \leq M} w\left(\frac{h}{M}\right) \left\{ \hat{\gamma}_T(h) - \gamma(h) \right\} e^{-ih\lambda} - \sqrt{\frac{T}{M}} \frac{1}{2\pi} \sum_{|i| \geq M} \left(1 - w\left(\frac{h}{M}\right)\right) \gamma(h) e^{-ih\lambda} \\
- \sqrt{\frac{T}{M}} \frac{1}{2\pi} \sum_{|i| > M} \gamma(h) e^{-ih\lambda}.
$$

From (W) and $\sum_h h^2 |\gamma(h)| < \infty$ we obtain that the second and third summand in the above representation converge to zero, i.e. the asymptotic distribution is completely determined by

$$(6.5) \quad S_T(\lambda) = \sqrt{\frac{T}{M}} \frac{1}{2\pi} \sum_{|i| \leq M} w\left(\frac{h}{M}\right) \left\{ \hat{\gamma}_T(h) - \gamma(h) \right\} e^{-ih\lambda}.$$ 

This strongly suggests to use

$$(6.6) \quad S_T^*(\lambda) = \sqrt{\frac{T}{M}} \frac{1}{2\pi} \sum_{|i| \leq M} w\left(\frac{h}{M}\right) \left\{ \hat{\gamma}_T^*(h) - \gamma^*(h) \right\} e^{-ih\lambda}$$

as a bootstrap approximation of $S_T$. Here we make use of the usual residual bootstrap introduced in Section 2 and 3. We have the following result.

**Theorem 6.4** Assume (A1)-(A3) and $\mathbb{E} \varepsilon_1^4 < \infty$. Then we have in probability for all $0 \leq \lambda \leq \pi$

$$d_2 \left( \mathcal{L}(S_T^*(\lambda)), \mathcal{L}(S_T(\lambda)) \right) \to 0 \text{ as } T \to \infty.$$ 

$d_2$ is defined in (3.1). Under the assumptions of Proposition 6.4 we have the same asymptotic normal distribution for $S_T^*(\lambda)$ as for the lag-window spectral density estimator (6.1).

### 7. Simulations

In this section we try to demonstrate the finite sample properties of the usual residual and the modified wild bootstrap. For this purpose let us consider the following three models.

Model I (AR(1)) : $X_t = 0.9 \cdot X_{t-1} + \varepsilon_t$

Model II (ARMA(2,1)) : $X_t = -1.29 \cdot X_{t-1} + 0.83 \cdot X_{t-2} + \varepsilon_t + 0.5 \cdot \varepsilon_{t-1}$

Model III (ARMA(2,2)) : $X_t = 0.1 \cdot X_{t-1} + 0.8 \cdot X_{t-2} + \varepsilon_t + 0.1 \cdot \varepsilon_{t-1} + 0.8 \cdot \varepsilon_{t-2}$

In model I we deal with estimators of an AR(1)-fit, i.e. with estimators of the underlying coefficient $b_1 = 0.9$ and the usual residual bootstrap, cf. Section 3. All procedures are based on a sample size of $T = 100$ of the underlying process. The innovations $(\varepsilon_t)$ are
assumed to be double-exponentially distributed.
In Figure 1a we compare the simulated density of the distribution $\sqrt{T}(\hat{b}_1 - b_1)$ (thick curve), cf. (3.5), with three independent usual residual bootstrap approximations (thin curves) based on different and randomly chosen samples of size 100 of the underlying model.

Insert Figure 1a around here

These bootstrap approximations have been computed using the order $P(T) = 1$ for the bootstrap autoregressive process.
Figure 1b contains similar densities, but now we use $P(T) = 3$, i.e. we use an overfitted AR(3)-model in the bootstrap world. But again we use the usual residual bootstrap procedure and the parameter estimator of an autoregressive fit of order 1.

Insert Figure 1b around here

Now let us turn to an underlying autoregression of infinite order, i.e. to model II. Figures 2a and 2b demonstrate the behaviour of the modified wild bootstrap, cf. Section 4. All simulated densities correspond to the distribution of estimators of the first autocorrelation $r_1 = -0.626$.

Insert Figures 2a and 2b around here

The thick curves in Figure 2a and 2b represent the density of the distribution of $\sqrt{T}(\hat{r}_1 - r_1)$, i.e. the distribution of the sample autocorrelation (sample size $T = 200$). The thin curves are independent modified wild bootstrap approximations. In Figure 2a we use an AR(2)-model in the bootstrap world, i.e. $P(T) = 2$, while Figure 2b contains similar results for $P(T) = 5$.

From the simulations one gets the impression that the selection of the order $P(T)$, i.e. the order of the autoregressive model in the bootstrap world, is not a very crucial parameter. Moreover, overfitting seems to be not very problematic.

Finally, we consider lag-window estimators of the spectral density in model III, cf. Section 6. As a window function we use the so-called triangular window, i.e. \( w(u) = (1 - |u|)I_{[-1,1]}(u) \). For the sample size $T = 200$ a truncation at $M = 10$ seems to be appropriate. We use the usual residual bootstrap procedure to produce a pointwise confidence band at level 0.95 for the underlying spectral density (cf. Figure 3a). The dashed line refers to the underlying spectral density, while the solid lines refer to the pointwise confidence band.
The underlying bootstrap autoregressive process is of order $P(T) = 4$.
In comparison, we report in Figure 3b a classical pointwise confidence band (level 0.95) for the spectral density based on a Chi-square approximation (cf. Brockwell and Davis (1991), Chapter 10.5).
Remark. All simulated densities are computed on the basis of 1000 Monte Carlo replications. For the smoothing we have used a usual Nadaraya-Watson kernel density estimator with kernel $K(u) = 3/4(1 - u^2)1_{[-1,1]}(u)$ and bandwidth $h > 0$.

8. Auxiliary Results and Proofs

At several places we heavily make use of the following inequalities, which mainly concern the behaviour of the moving average coefficients $\hat{a}_\nu(P)$ of the fitted autoregression of order $P$ to the given data set.

Let us consider the following two power series expansions. Both series converge at least for all complex $z$ with magnitude less than or equal to one.

$$A(z) = 1 - \sum_{\nu=1}^{\infty} a_\nu z^\nu \quad \text{and} \quad A_p(z) = 1 - \sum_{\nu=1}^{p} a_\nu(p) z^\nu.$$  

Here $(a_\nu : \nu \in \mathbb{N})$ denote the coefficients of the underlying autoregression (1.1), whereas $(a_\nu(p) : \nu = 1, \ldots, p)$ denote the coefficients of a (theoretically) best AR$(p)$-fit in $L_2$-distance, i.e. the coefficients $(a_1(p), \ldots, a_p(p))^T = \psi(p)$ are defined uniquely as the argmin of

$$E(X_t - \sum_{\nu=1}^{p} c_\nu X_{t-\nu})^2,$$  

which is equivalent to

$$\psi(p) = R(p)^{-1}(r_1, \ldots, r_p)^T.$$  

15
From assumption \( (A/2) \) we obtain the following moving average representation of the underlying process \( X \).

\[
X_t = \sum_{\nu=0}^{\infty} \alpha_{\nu} \varepsilon_{t-\nu}, \quad t \in \mathbb{Z},
\]

where \( (\alpha_{\nu} : \nu = 0, 1, 2, \ldots) \) denote the power series coefficients of \( (1 - \sum_{\nu} a_{\nu} z^{\nu})^{-1} \).

According to Theorem 2.2 of Baxter (1962) we have a constant \( C > 0 \) and \( p_o \in \mathbb{N} \) such that for all integers \( p \geq p_o \)

\[
\sum_{\nu=1}^{p} \left| a_{\nu}(p) - a_{\nu} \right| \leq C \cdot \sum_{\nu=p_o}^{\infty} \left| a_{\nu} \right|.
\]

This inequality easily implies that for all large \( p \) the polynomial \( A_p \), cf. (8.1), has no zeroes with magnitude less than or equal to one. Moreover we have the following result.

**Lemma 8.1** There exist \( \delta > 0 \) and \( p_o \in \mathbb{N} \), such that for all \( p \geq p_o \)

\[
\inf_{|\nu| \leq 1 + 1/p} |A_p(\nu)| \geq \delta.
\]

**Proof:** Suppose that the assertion of Lemma 8.1 is false. Then there exists a sequence \( p(k) \) of integers converging to infinity as \( k \to \infty \) and a sequence of complex numbers \( z_k \) with \( |z_k| \leq 1 + 1/p(k) \) such that \( A_{p(k)}(z_k) \to 0 \) as \( k \to \infty \). Without loss of generality we assume that \( z_k \to z_o \). Necessarily \( |z_o| \leq 1 \); more exactly \( |z_o| = 1 \). We will show that \( A(z_o) = 0 \), which is a contradiction to our assumptions. To this end observe

\[
A(z_o) = 1 - \sum_{\nu=1}^{\infty} a_{\nu} z_o^{\nu} = A_{p(k)}(z_k) + \sum_{\nu=1}^{p(k)} a_{\nu} (z_k^{\nu} - z_o^{\nu}) + \sum_{\nu=1}^{p(k)} (a_{\nu}(p(k)) - a_{\nu}) z_k^{\nu} + \sum_{\nu=p(k)+1}^{\infty} a_{\nu} z_o^{\nu} = o(1),
\]

because of our assumptions on \( z_k \), \( (A2), (8.4) \) and the Theorem of dominated convergence. Note that \( |z_k|^{p(k)} \leq |z_k|^{p(k)} \leq (1 + 1/p(k))^{p(k)} \leq 3 \) for all \( \nu \leq p(k) \). This implies \( A(z_o) = 0 \). \( \blacksquare \)

The coefficients \( a_{\nu} \) of the moving average representation (8.3) can formally be defined as follows

\[
1 + \sum_{\nu=1}^{\infty} a_{\nu} z^{\nu} = \left( 1 - \sum_{\nu=1}^{\infty} a_{\nu} z^{\nu} \right)^{-1}, \quad |z| \leq 1.
\]

Because of Lemma 8.1 we may define for all \( p \) large enough

\[
1 + \sum_{\nu=1}^{\infty} a_{\nu}(p) z^{\nu} = \left( 1 - \sum_{\nu=1}^{p} a_{\nu}(p) z^{\nu} \right)^{-1}, \quad |z| \leq 1 + 1/p.
\]

We obtain the following result.
Lemma 8.2 There exists a constant \( C > 0 \) such that for all large \( p \)

\[
\sum_{\nu=0}^{\infty} |\alpha_{\nu}(p) - \alpha_{\nu}| \leq C \cdot \sum_{\nu=p}^{\infty} |a_{\nu}| .
\]

Obviously, the right hand side converges to zero as \( p \to \infty \).

Proof: Since \( a = (1, -a_1, -a_2, \ldots) \in \ell_1 = \{(z_\nu : \nu \in \mathbb{N}) \subset \ell_1 \mid \sum_\nu |z_\nu| < \infty \} \) and \( 1 - \sum_\nu a_\nu z_\nu \neq 0 \) for all \( |z| = 1 \), \( a \) has a multiplicative inverse \( a^{-1} = (1, a_1, a_2, \ldots) \in \ell_1 \), i.e.

\[
a \cdot a^{-1} = \left( 1, a_n - \sum_{k=0}^{n-1} a_{n-k} a_k : n \in \mathbb{N} \right) = (1, 0, 0, \ldots) ,
\]
where \( a \cdot b \) is the convolution of \( a \) and \( b \). Cf. Zelazko (1973), Theorem 8.11 and Assertion 9.4.

The coefficients \( a_\nu \) are the power series coefficients of \( (1 - \sum_\nu a_\nu z_\nu)^{-1} \). Because of \( a^{-1} \in \ell_1 \) we have \( \sum_\nu |a_\nu| < \infty \).

Furthermore, for all large \( p \), \((1, -a_1(p), \ldots, -a_p(p), 0, 0, \ldots) \equiv a(p) \in \ell_1 \) and is invertible with inverse \( a(p)^{-1} = (1, a_1, a_2, \ldots) \), cf. (8.6).

Finally observe that

\[
\sum_{\nu=0}^{\infty} |\alpha_{\nu}(p) - \alpha_{\nu}| = |a(p)^{-1} - a^{-1}| = |a(p)^{-1}(a - a(p))a^{-1}|
\]

\[
\leq |a(p)^{-1} - a^{-1}| |a - a(p)| |a^{-1}| + |a^{-1}|^2 |a - a(p)| ,
\]
which leads to

\[
|a(p)^{-1} - a^{-1}| \leq \frac{|a^{-1}|^2 |a - a(p)|}{1 - |a^{-1}| |a - a(p)|} .
\]

\(|a - a(p)| = \sum_{\nu=1}^{p} |a_\nu - a_\nu(p)| + \sum_{\nu=p+1}^{\infty} |a_\nu| \) together with (8.4) implies the assertion of Lemma 8.2.

Remark. If \( a_\nu = 0 \) for all \( \nu \geq p_0 \), i.e. \( X \) is a finite order autoregression, then \( \sum_{\nu=0}^{\infty} |\alpha_{\nu}(p) - \alpha_{\nu}| = 0 \) for all \( p \geq p_0 \).

Finally we need a bound for \( |\hat{\alpha}_\nu(p) - \alpha_\nu(p)| \). Here we denote by \( \hat{\alpha}_\nu(p) \) the coefficients of the series \((1 - \sum_{\nu=1}^{p} \hat{a}_\nu(p) z_\nu)^{-1} \). Recall that \( \hat{\psi}_T(p) = (\hat{a}_1(p), \ldots, \hat{a}_p(p))^T \) denotes the least squares parameter estimator, cf. (2.2). From Hannan and Kavalieris (1986), Theorem 2.1, we obtain under our assumptions

\[
\max_{1 \leq \nu \leq p} |\hat{a}_\nu(p) - a_\nu(p)| = \mathcal{O}_P \left( \sqrt{\frac{\log T}{T}} \right) \quad (8.8)
\]

uniformly in \( p \leq p_T , \) \( p_T^2 = o(T/\log T) \). This yields, again uniformly in \( p \leq p_T \),

\[
\sum_{\nu=1}^{p} |\hat{a}_\nu(p) - a_\nu(p)| \left( 1 + \frac{1}{p} \right)^\nu = \mathcal{O}_P \left( p \sqrt{\frac{\log T}{T}} \right) = o_P(1) . \quad (8.9)
\]

17
Because of (8.9) and Lemma 8.1 we have with large probability that the polynomial
\[ 1 - \sum_{\nu=1}^{p} \hat{a}_\nu(p) z^\nu \] also has no zeroes with magnitude less than or equal to \( 1 + 1/p \).
Because of this we may apply Cauchy’s inequality for holomorphic functions in order to obtain the following result.

**Lemma 8.3** We have uniformly in \( p \leq p_T \), \( p_T^2 = o(T/\log T) \), and uniformly in \( \nu \in \mathbb{N} \)
\begin{equation}
|\hat{a}_\nu(p) - a_\nu(p)| \leq \frac{p}{(1 + \frac{1}{p})^\nu} \mathcal{O}_p \left( \sqrt{\frac{\log T}{T}} \right).
\end{equation}

**Proof:** Cauchy’s inequality (cf. Ahlfors (1966) or Rudin (1987), Theorem 10.26) together with (8.8), (8.9) and Lemma 8.1 yields
\[
|\hat{a}_\nu(p) - a_\nu(p)| \leq \frac{1}{(1 + \frac{1}{p})^\nu} \max_{|z| = 1 + 1/p} \left| \left( 1 - \sum_{\nu} a_\nu(p) z^\nu \right)^{-1} - \left( 1 - \sum_{\nu} \hat{a}_\nu(p) z^\nu \right)^{-1} \right|
= \frac{p}{(1 + \frac{1}{p})^\nu} \mathcal{O}_p \left( \sqrt{\frac{\log T}{T}} \right),
\]
which is the assertion.

**Proof of Proposition 3.1:** Let us denote by \( \hat{F}_T \) the empirical distribution of the unobservable innovations \( \varepsilon_1, \ldots, \varepsilon_T \). Because of Lemma 8.4 in Bickel and Freedman (1981) it suffices to show that \( d_r \left( \hat{F}_T^\nu, \hat{F}_T^\xi \right) \) converges to zero in probability. To see this let \( J \) be Laplace distributed on \( \{1, \ldots, T\} \) and define random variables \( Y_1, Y_2 \) with marginals \( \hat{F}_T^\nu, \hat{F}_T^\xi \), respectively, as follows
\[
Y_1 = \varepsilon_J, \quad Y_2 = \hat{\varepsilon}_J - \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t.
\]
For the case \( r = 2 \) we obtain
\[
d_2^2 \left( \hat{F}_T^\nu, \hat{F}_T^\xi \right) \leq E \left( Y_1 - Y_2 \right)^2
= \frac{1}{T} \sum_j \left( \hat{\varepsilon}_j - \varepsilon_j - \frac{1}{T} \sum_i \hat{\varepsilon}_i \right)^2
\leq \frac{6}{T} \sum_j (\hat{\varepsilon}_j - \varepsilon_j)^2 + 3 \left( \frac{1}{T} \sum_i \varepsilon_i \right)^2
= \frac{6}{T} \sum_j \left( \sum_{\nu=1}^{P} \left( \hat{a}_\nu(P) - a_\nu \right) X_{j-\nu} - \sum_{\nu=P+1}^{\infty} a_\nu X_{j-\nu} \right)^2 + o_P(1).
\]
The proof can be concluded from the following three results.

\[
E \left( \sum_{\nu=1}^{\infty} |a_{\nu}| X_{j-\nu} \right)^2 \leq E X_1^2 \cdot \left( \sum_{\nu=\text{pmin}}^{\infty} |a_{\nu}| \right)^2 = o(1),
\]

since \( p_{\text{min}} \to T \to \infty \infty \).

\[
\frac{1}{T} \sum_{j} \left( \sum_{\nu=1}^{p} (\hat{a}_{\nu}(P) - a_{\nu}(P)) X_{j-\nu} \right)^2 \leq \max_{1 \leq p \leq \text{pmax}(T)} \max_{1 \leq \nu \leq p} |\hat{a}_{\nu}(p) - a_{\nu}(p)|^2 \frac{1}{T} \sum_{j} \left( \sum_{\nu=1}^{\text{pmax}} |X_{j-\nu}| \right)^2 = O_p \left( \frac{\text{pmax}(T) \log T}{T} \right) = o_p(1),
\]
because of (8.8).

\[
\begin{align*}
\frac{1}{T} \sum_{j} \left( \sum_{\nu=1}^{p} (a_{\nu}(P) - a_{\nu}) X_{j-\nu} \right)^2 &= \sum_{p=\text{pmin}}^{\text{pmax}} \left( (a_{\nu}(P) - a_{\nu}) (a_{\mu}(P) - a_{\mu}) \right) \frac{1}{T} \sum_{j} X_{j-\nu} X_{j-\mu} \\
&\leq \sum_{p=\text{pmin}}^{\text{pmax}} \sum_{\nu,\mu=1}^{p} |a_{\nu}(P) - a_{\nu}| |a_{\mu}(P) - a_{\mu}| \frac{1}{T} \sum_{j} X_{j-\nu} X_{j-\mu} - \gamma(\nu - \mu) \\
&\quad + \sum_{\nu,\mu=1}^{p} |a_{\nu}(P) - a_{\nu}| |a_{\mu}(P) - a_{\mu}| \cdot |\gamma(\nu - \mu)| \\
&= \frac{1}{T} \sum_{p=\text{pmin}}^{\text{pmax}} \left( \sum_{\nu=1}^{p} |a_{\nu}(P) - a_{\nu}| \right)^2 + E X_1^2 \cdot \left( \sum_{\nu=1}^{p} |a_{\nu}(P) - a_{\nu}| \right)^2 \\
&\leq \left( O_p(T^{-1/2}) \cdot \text{pmax}(T) + E X_1^2 \right) \cdot \left( \sum_{\nu=\text{pmin}}^{\infty} |a_{\nu}| \right)^2, \text{ cf. (8.4)} \\
&= o_p(1), \text{ by assumption (A2) and (A3).}
\end{align*}
\]

To see equality (*) observe that \( E \left( \frac{1}{T} \sum_{j=1}^{T} \{ X_{\nu} X_{\mu} - \gamma(\nu - \mu) \} \right)^2 \) is uniformly bounded in \( \nu \) and \( \mu \) and that \( |\gamma(\nu - \mu)| \leq E X_1^2 \).

**Proof of Theorem 3.1:** For reasons of not too complicated notation, we restrict our attention to the case \( K=0 \), only.

Denote by \( ((\varepsilon_1^i, \varepsilon_i) : i \in \mathbb{Z}) \) a sequence of i.i.d. bivariate random vectors with arbitrary dependence structure between \( \varepsilon_1^i \) and \( \varepsilon_1 \) up to the given marginals \( \varepsilon_1 \sim F_1^\varepsilon \) and \( \varepsilon_1 \sim F \).

The square of (3.2) is bounded through

\[
\frac{1}{T} \inf E \left( \sum_{t=1}^{T} \left[ (X_t^2 - E X_t^2) - (X_t^2 - E X_t^2) \right]^2 \right)
\]
Compute the first expectation on the right hand side of (8/1/2/) and obtain the following

\[
\left( \sum_{\nu} \alpha_{\nu} e_{t-\nu}^* - \sum_{\nu} \alpha_{\nu} (E_{\nu} e_{t-\nu}^* - E_{\nu} e_{t-1}) \right)^2
\]

Proposition 3/1 completes the argument.

It remains to consider /8/1/1/). A tedious but direct computation gives us the following right hand side of /8/1/2/). This bound also vanishes in probability /.

It suffices to consider the following two expectations.

\[
E \left( \sum_{t} \left( \left( \sum_{\nu} \alpha_{\nu} e_{t-\nu}^* \right)^2 - \sum_{\nu} \alpha_{\nu} (E_{\nu} e_{t-\nu}^* - E_{\nu} e_{t-1}) \right) \right)^2
\]

and

\[
E \left( \sum_{t} \left( \sum_{\nu} \left( \hat{\alpha}_{\nu}(P) - \alpha_{\nu} \right) e_{t-\nu}^* \right)^2 - \sum_{\nu} \left( \hat{\alpha}_{\nu}(P) - \alpha_{\nu} \right)^2 E_{\nu} e_{t-1}^* \right)^2
\]

\[
\leq 2 \cdot E \left( \sum_{t} \left( \sum_{\nu} \left( \hat{\alpha}_{\nu}(P) - \alpha_{\nu} \right) e_{t-\nu}^* - \hat{\alpha}_{\nu}(P) - \alpha_{\nu} \right)^2 E_{\nu} e_{t-1}^* \right)^2
\]

\[
+ 8 \cdot E \left( \sum_{t} \left( \sum_{\nu} \left( \hat{\alpha}_{\nu}(P) - \alpha_{\nu} \right) e_{t-\nu}^* \sum_{\mu} \alpha_{\mu} e_{t-\mu}^* - \sum_{\nu} \left( \hat{\alpha}_{\nu}(P) - \alpha_{\nu} \right) \alpha_{\nu} E_{\nu} e_{t-1}^* \right)^2 \right)
\]

Compute the first expectation on the right hand side of (8/1/2) and obtain the following bound

\[
\sum_{t=1}^{T} \sum_{s=1}^{t} \sum_{\nu} \left( \hat{\alpha}_{\nu+s}(P) - \alpha_{\nu+s} \right)^2 \left( \hat{\alpha}_{\nu}(P) - \alpha_{\nu} \right)^2 \cdot \left( E_{\nu} e_{t}^* - 3 (E_{\nu} e_{t-1})^2 \right)
\]

\[
+ \sum_{t=1}^{T} \sum_{s=1}^{t} \sum_{\nu, \mu} \left| \hat{\alpha}_{\nu+s}(P) - \alpha_{\nu+s} \right| \left| \hat{\alpha}_{\nu+s}(P) - \alpha_{\nu+s} \right| \left| \hat{\alpha}_{\nu}(P) - \alpha_{\nu} \right| \left( E_{\nu} e_{t-1}^* \right)^2
\]

Use the fact that \( E_{\nu} e_{t}^* \) and \( E_{\nu} e_{t-1}^* \) are bounded in probability because of Proposition 3/1 in order to obtain with the help of Lemma 8/2 and 8/3 that the last expression can be bounded through

\[
O_p \left( \frac{\sqrt{T/2}(T)}{T} \log T + \sum_{\nu=\min(T)}^{\infty} |a_{\nu}| \right) = O_p(1).
\]

Exactly along the same lines we obtain a similar bound for the second expectation of the right hand side of (8/1/2). This bound also vanishes in probability.

It remains to consider (8/1/1). A tedious but direct computation gives us the following bound for this expectation

\[
O_p \left( \sqrt{E(e_{t}^* - e_1)^2} + \sqrt{E(e_{t-1}^* - e_1)^2} \right) = O_p \left( d_2(F_{\hat{\alpha}_T}^*, F) + d_4(F_{\hat{\alpha}_T}^*, F) \right).
\]

Proposition 3/1 completes the argument.
Proof of Proposition 4.3: For the sake of simplicity we restrict our attention again to the case $K=0$. Recall from (2.6) the definition of $X_t^\dagger$ and obtain

$$
\sqrt{T} \left( \gamma_t^\dagger(0) - \gamma_0(0) \right) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left( X_t^{12} - \gamma_t(0) \right)
$$

$$
= \frac{1}{\sqrt{T}} \sum_{t} \left[ \left( \sum_{\nu=0}^{t-1} \hat{\alpha}_{\nu}(P) \varepsilon_{t-\nu}^\dagger \right)^2 - \sum_{\nu=0}^{t-1} \hat{\alpha}_{\nu}^2(P) \cdot \varepsilon_{t-\nu}^2 \right].
$$

For fixed $m \in \mathbb{N}$ abbreviate

$$
U_{t,T}^m := \left( \sum_{\nu=0}^{m} \hat{\alpha}_{\nu}(P) \varepsilon_{t-\nu}^\dagger \right)^2 - \sum_{\nu=0}^{m} \hat{\alpha}_{\nu}^2(P) \varepsilon_{t-\nu}^2
$$

and observe that these are centered and $m$-dependent random variables. We may conclude the proof of Proposition 4.3 from Brockwell and Davis (1991), Proposition 6.3.9, if we can verify (8.14) - (8.16).

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} U_{t,T}^m \Longrightarrow \mathcal{N}(0, \tau_m^2) \quad \text{in probability},
$$

where

$$
\tau_m^2 = ((E\eta_1^4 - 1)E\varepsilon_1^4 - 2\sigma^4) \left( \sum_{\nu=0}^{m} \alpha_{\nu}^2 \right)^2 + 2\sigma^4 \left( \sum_{\nu=0}^{m} \alpha_{\nu}^2 \right)^2 + 2 \sum_{h=1}^{m} \left( \sum_{\nu=0}^{m-h} \alpha_{\nu} \alpha_{\nu+h} \right)^2.
$$

$$
\tau_m^2 \rightarrow_{m \rightarrow \infty} \tau^2 = (E\eta_1^4 - 1)E\varepsilon_1^4 - 2\sigma^4 \left( \sum_{\nu=0}^{\infty} \alpha_{\nu}^2 \right)^2 + 2\sigma^4 \left( \sum_{\nu=0}^{\infty} \alpha_{\nu}^2 \right)^2 + 2 \sum_{h=1}^{\infty} \left( \sum_{\nu=0}^{\infty} \alpha_{\nu} \alpha_{\nu+h} \right)^2
$$

$$
\frac{(E\eta_1^4 - 1)E\varepsilon_1^4 - 2\sigma^4}{\sigma^4} \cdot \gamma(0)^2 + 2 \left( \gamma(0)^2 + 2 \sum_{h=1}^{\infty} \gamma(h)^2 \right)
$$

and

$$
\lim_{m \rightarrow \infty} \lim_{T \rightarrow \infty} E^T \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left( U_{t,T}^m - U_{t,T}^m \right) \right)^2 = 0 \quad \text{in probability}.
$$

To see (8.13) we use the following Lemma 8.4. We have

$$
E^T U_{t+\tau,\tau}^m U_{t,T}^m = \sum_{\nu,\mu=0}^{m} \sum_{\tau,\delta=0}^{m} \hat{\alpha}_{\nu}(P) \hat{\alpha}_{\mu}(P) \hat{\alpha}_{\tau}(P) \hat{\alpha}_{\delta}(P) \ E^T \left( \varepsilon_{t+\nu-\mu,\tau+\nu-\mu,\tau+\delta-\nu+\delta,\tau+\delta-\nu} \right)
$$

$$
- \sum_{\nu,\tau=0}^{m-h} \hat{\alpha}_{\nu}^2(P) \hat{\alpha}_{\tau}^2(P) \varepsilon_{t+\nu-\tau,\tau}^2
$$

$$
= \sum_{\tau=0}^{m-h} \hat{\alpha}_{\tau}^2(P) \hat{\alpha}_{\tau}^2(P) \left[ E^T(\varepsilon_{t-\tau}^2)^4 - 3 \cdot \varepsilon_{t-\tau}^2 \right]
$$

$$
+ 2 \sum_{\tau,\delta=0}^{m-h} \hat{\alpha}_{\tau} \hat{\alpha}_{\tau+h}(P) \hat{\alpha}_{\delta} \hat{\alpha}_{\delta}(P) \varepsilon_{t-\tau}^2 \varepsilon_{t-\delta}^2.
$$
Observe that by definition of $\varepsilon_i^4$, cf. (2.5), we have $E^4(\varepsilon_i^4) = E \eta_1^4 \cdot \varepsilon_i^4$. From Proposition 3.1 and Lemma 8.3 of Bickel and Freedman (1981) we have for all $\tau, \delta = 0, \ldots, m$:

\begin{align*}
(i) & \quad \frac{1}{r_T} \sum_{\nu=1}^{r_T} \varepsilon_{i_{\nu}, -\tau} \rightarrow_{T \to \infty} E \varepsilon_1^4 \quad \text{in probability} \\
(ii) & \quad \frac{1}{r_T} \sum_{\nu=1}^{r_T} \varepsilon_{i_{\nu}, -\tau}^2 \varepsilon^2_{i_{\nu}, -\delta} \rightarrow \left\{ \begin{array}{l}
(E \varepsilon_1^4)^2 = \sigma^4, \quad \tau \neq \delta \\
E \varepsilon_1^4, \quad \tau = \delta
\end{array} \right. \quad \text{in probability.}
\end{align*}

$\{r_T\}$ denotes a sequence of integers converging to infinity and $t_1 < t_2 < \ldots < t_{r_T} \leq T$ for all $T \in \mathbb{N}$. We finally obtain from (i), (ii) and Lemma 8.2 and 8.3 for every fixed $m \in \mathbb{N}$

\[
\frac{1}{\sqrt{T}} \sum_{\nu=1}^{r_T} E^4 U_{i_{\nu} + h, T}^m U_{i_{\nu}, T}^m \rightarrow_{T \to \infty} c(h) \quad \text{in probability,}
\]

where

\[
c(h) = ((E \eta_1^4 - 1) E \varepsilon_1^4 - 2\sigma^4) \sum_{\tau=0}^{m-h} \alpha_\tau^2 \alpha_{\tau+h}^2 + 2\sigma^4 \left( \sum_{\tau=0}^{m-h} \alpha_\tau \alpha_{\tau+h} \right)^2.
\]

Simple algebra yields

\[
\tau_m^2 = c(0) + 2 \sum_{h=1}^{m} c(h)
\]

\[
= ((E \eta_1^4 - 1) E \varepsilon_1^4 - 2\sigma^4) \left( \sum_{\tau=0}^{m} \alpha_\tau^2 \right)^2 + 2\sigma^4 \left( \left( \sum_{\tau=0}^{m} \alpha_\tau^2 \right)^2 + 2 \sum_{h=1}^{m} \left( \sum_{\tau=0}^{m-h} \alpha_\tau \alpha_{\tau+h} \right)^2 \right) > 0.
\]

The following Lemma 8.4 now ensures (8.13), (8.15) is obvious. Finally we conclude (8.16) as follows.

\[
E^4 \left( \frac{1}{\sqrt{T}} \sum_{f=1}^{T} \left( U_{i_{f}, T}^{m} - U_{i_{f}, T}^{m} \right) \right)^2
\]

\[
= \frac{1}{T} \sum_{s,t=1}^{T} \sum_{\tau > m} \hat{\alpha}_{\tau+s-t}^2 (P) \hat{\alpha}_\tau^2 (P) \left\{ E^4 (\varepsilon_1^4)^4 - 3 \cdot \varepsilon_1^4 \right\}
\]

\[
+ \frac{2}{T} \sum_{s,t=1}^{T} \sum_{\tau \leq \delta > m} \hat{\alpha}_{\tau+s-t} (P) \hat{\alpha}_{\tau} (P) \hat{\alpha}_{\tau+s-t} (P) \hat{\alpha}_{\tau} (P) \varepsilon_1^2 \varepsilon_1^2.
\]

Repeated application of Lemma 8.2 and 8.3 yields that

\[
\limsup_{T \to \infty} E^* \left( \frac{1}{\sqrt{T}} \sum_{f=1}^{T} \left( U_{i_{f}, T}^{m} - U_{i_{f}, T}^{m} \right) \right)^2 = \mathcal{O} \left( \sum_{\tau=0}^{\infty} \alpha_\tau^2 \right) \to_{m \to \infty} 0
\]

by our assumptions.
In the above proof we make use of a version of the CLT for \( m \)-dependent random variables. This version is a slight modification of results known in the literature. This is especially true for a CLT for strictly stationary \( m \)-dependent sequences, cf. Brockwell and Davis (1991) Theorem 6.4.2. The following version allows for triangular arrays which are on average stationary, only (see Lemma 8.4 for precise definition).

**Lemma 8.4 CLT for \( m \)-dependent sequences**

Assume that for each \( T \in \mathbb{N} \) real-valued, centered and \( m \)-dependent random variables \( \{U_{i,T} : t = 1, \ldots, T\} \) are given. Further assume

For \( h \in \mathbb{N}_0 \), each sequence \( \{r_t : t \in \mathbb{N}\} \) of positive integers converging to infinity and arbitrary \( t_1, t_2, \ldots \) with \( r_{t_T} \leq T - h \):

\[
\frac{1}{r_T} \sum_{t=1}^{r_T} E(U_{i_t+h,T}U_{i_t,T}) \to_{T \to \infty} c(h) \quad h \in \mathbb{N}_0,
\]

where the function \( c \) fulfills \( c(0) + 2 \sum_{k=1}^{m} c(h) = \tau^2 > 0 \).

\[
\frac{1}{T^{1+\delta}} \sum_{t=1}^{T} E|U_{i,T}|^{2(1+\delta)} \to_{T \to \infty} 0 \quad \text{for some } \delta > 0.
\]

Then we have

\[
\lim_{T \to \infty} \text{Var} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} U_{i_t,T} \right) = \tau^2
\]

and

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} U_{i_t,T} \Rightarrow \mathcal{N}(0, \tau^2).
\]

**Remark.** The convergence part in assumption (8.17) of Lemma 8.4 is for example implied by the stricter assumption

\[
EU_{i_{T+h},T}U_{r_T,T} \to c(h) \quad \text{as } T \to \infty
\]

for every sequence \( \{t_T : T \in \mathbb{N}\} \) with \( t_T \in \{1, \ldots, T\} \) and \( t_T \to \infty \) as \( T \to \infty \). Because of this, Lemma 8.4 generalizes Theorem 6.4.2 of Brockwell and Davis (1991) slightly.

**Proof:** To see (8.19) compute for \( T > m \)

\[
\text{Var} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} U_{i_t,T} \right) = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} E(U_{s,T}U_{i_t,T}) = \\
= \frac{1}{T} \sum_{t=-(T-1)}^{T-1} \sum_{s=1}^{T-|t|} E(U_{s+|t|,T}U_{s,T}) = \sum_{t=-m}^{m} \frac{1}{T} \sum_{s=1}^{T-|t|} E(U_{s+|t|,T}U_{s,T}) \quad \text{m-dependence}
\]

\[
\to_{T \to \infty} \sum_{t=-m}^{m} c(|t|) = c(0) + 2 \sum_{t=1}^{m} c(t) > 0, \quad \text{because of (8.16)}.
\]
To prove (8.20) consider for \( k > 2m \) and \( r = r_T = \lceil T/k \rceil \)

\[
Y_{T,k} := \frac{1}{\sqrt{T}} \sum_{j=1}^{r} (U_{(j-1)k+1,T} + \cdots + U_{jk-m,T}) \cdot \frac{1}{Z_j,T} .
\]

The triangular array \( \{Z_{t,T} : t = 1, \ldots, r\}_{T \in \mathbb{N}} \), which consists of independent random variables in each row, fullfills (for the sake of simplicity we assume that \( T = kr - m \) for some integer \( k \))

\[
\frac{1}{T} E \left( \sum_{t=1}^{T} U_{t,T} - \sum_{j=1}^{r} Z_{j,T} \right)^2 = \frac{1}{T} \sum_{j=1}^{r-1} E (U_{jk-m+1,T} + \cdots + U_{jk,T})^2
\]

\[
= \frac{r-1}{T} \sum_{\nu=1}^{m} \sum_{\mu=1}^{m} \frac{1}{r-1} \sum_{j=1}^{r-1} E (U_{jk+\nu,T} U_{jk-\mu,T}) \rightarrow \nu(\mu-\delta) \text{ as } T \to \infty \text{ because of (8.17)}.
\]

\[
\to T \to \infty \quad \frac{1}{k} \sum_{|h| \leq m} \left( m - |h| \right) c(|h|) .
\]

Because of this we obtain for each \( \varepsilon > 0 \)

\[
\lim_{k \to \infty} \limsup_{T \to \infty} P \left( \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} U_{t,T} - Y_{T,k} \right| > \varepsilon \right) = 0 .
\]

According to Proposition 6.3.9, Brockwell and Davis (1991), it suffices to deal with \( Y_{T,k} \). We have in mind to apply the CLT for triangular arrays. To this end compute

\[
\frac{1}{T} \sum_{j=1}^{r} E Z_{j,T}^2 = \frac{1}{T} \sum_{j=1}^{r} E (U_{(j-1)k+1,T} + \cdots + U_{jk-m,T})^2
\]

\[
= \frac{1}{T} \sum_{j=1}^{r} \sum_{t=(k-m-1)}^{k-m-1} \sum_{s=1}^{k-m-|t|} E (U_{(j-1)k+s+|t|,T} U_{(j-1)k+s,T}) \to T \to \infty \quad \sum_{|h| \leq m} \frac{k-m-|h|}{k} \cdot c(|h|) \text{ because of (8.17)}.
\]

Since a Ljapunov-condition is fullfilled, i.e. for some \( \delta > 0 \)

\[
\frac{1}{T^{1+\varepsilon}} \sum_{j=1}^{r} E |Z_{j,T}|^{2(1+\varepsilon)} \leq \frac{1}{T^{1+\varepsilon}} \sum_{j} E \left( \sum_{s=1}^{k-m} |U_{(j-1)k+s,T}| \right)^{2(1+\varepsilon)} \leq \frac{k^{1+2\varepsilon}}{T^{1+\varepsilon}} \sum_{j} \sum_{s} E |U_{(j-1)k+s,T}|^{2(1+\varepsilon)} \leq k^{1+2\varepsilon} \cdot \frac{1}{T^{1+\varepsilon}} \sum_{t=1}^{T} E |U_{t,T}|^{2(1+\varepsilon)} \to T \to \infty 0 \text{ because of (8.18)},
\]

24
we obtain eventually from the CLT for triangular arrays (see for example Gänssler and Stute (1977), Korollar 9.2.9)

$$Y_{T,k} = \frac{1}{\sqrt{T}} \sum_{j=1}^{r} Z_{j,T} \Rightarrow \mathcal{N} \left(0, \sum_{|h| \leq m} \left(1 - \frac{m + |h|}{k}\right) c(|h|)\right).$$

Observe that \(\sum_{|h| \leq m} \left(1 - \frac{m + |h|}{k}\right) c(h) \to k \to \infty \sum_{|h| \leq m} c(h) = \tau^2 > 0\) to conclude the proof.

**Proof of Theorem 6.4**: We have (let \(\delta_{r,s} = 1\) if \(r = s\) and 0 otherwise)

$$
\gamma(h) = E \varepsilon_1^2 \sum_{\nu,\mu = 0}^{\infty} \alpha_{\nu} \alpha_{\mu} \delta_{\nu + h, \mu} \quad \text{and} \quad \gamma^*(h) = E^* X_i^* X_{i+h}^* = E^* (\varepsilon_1^*)^2 \sum_{\nu,\mu = 0}^{\infty} \hat{\alpha}_{\nu}(P) \hat{\alpha}_{\mu}(P) \delta_{\nu + h, \mu}.
$$

These expressions are a consequence of the moving average representation of \(X\), cf. (8.3), and of \(X^*\), cf. (2.4). From the definition of \(\hat{\gamma}_T(h)\) and \(\hat{\gamma}_T^*(h)\), c.f. (1.2) and Theorem 3.1, we obtain the following bound (\(\lambda \in [0, \pi]\))

$$
\frac{T M}{2} d_2^T (2\pi S_T^2(\lambda), 2\pi S_T(\lambda))
$$

$$
= \frac{T^2}{2} \inf E \left( \sum_{|h| \leq M} w(\frac{h}{M}) \left[ |\hat{\gamma}_T(h) - \gamma^*(h)| - |\hat{\gamma}_T(h) - \gamma(h)| \right] e^{-i h \lambda} \right)^2
$$

$$
\leq \inf \left\{ E \left( \sum_{|h| \leq M} w(\frac{h}{M}) e^{-i h \lambda} \sum_{i=1}^{T-1} \sum_{\nu,\mu = 0}^{\infty} \left| \hat{\alpha}_{\nu}(P) \hat{\alpha}_{\mu}(P) - \alpha_{\nu} \alpha_{\mu} \right| (\varepsilon_{i-\nu}^* \varepsilon_{i+h-\nu}^* - \delta_{\nu + h, \mu} E^*(\varepsilon_1^*)^2) \right)^2 \right\}
$$

$$
+ E \left( \sum_{|h| \leq M} w(\frac{h}{M}) e^{-i h \lambda} \sum_{i=1}^{T-1} \sum_{\nu,\mu = 0}^{\infty} \alpha_{\nu} \alpha_{\mu} (\varepsilon_{i-\nu}^* \varepsilon_{i+h-\mu}^* - \varepsilon_{i-\nu} \varepsilon_{i+h-\mu}^* - \delta_{\nu + h, \mu} (E^*(\varepsilon_1^*)^2 - E^*(\varepsilon_1^*)) \right)^2 \right),
$$

where the infimum is over all i.i.d. random vectors \((\varepsilon_t^*, \varepsilon_t)\), \(t \in \mathbb{Z}\), with given marginals \(\varepsilon_1 \sim F^*\) and \(\varepsilon_t^* \sim F_{t}^*\).

Bound the first expectation by

$$
\sum_{|h| \leq M} \sum_{s,t} \sum_{\nu,\mu} \sum_{\tau,\delta} \left| \hat{\alpha}_{\nu}(P) \hat{\alpha}_{\mu}(P) - \alpha_{\nu} \alpha_{\mu} \right| \left| \hat{\alpha}_{\tau}(P) \hat{\alpha}_{\delta}(P) - \alpha_{\tau} \alpha_{\delta} \right|
$$

$$
E \varepsilon_{i-\nu}^* \varepsilon_{i+h-\nu}^* \varepsilon_{s-\tau}^* \varepsilon_{s+k-\delta}^* - \delta_{\nu + h, \mu} \delta_{\tau + k, \delta} \left( E^*(\varepsilon_1^*)^2 \right)^2.
$$

Since

$$
E \varepsilon_{i_1}^* \varepsilon_{i_2}^* \varepsilon_{i_3}^* \varepsilon_{i_4}^* = \begin{cases} E^*(\varepsilon_1^*)^4, & \text{if } i_1 = i_2 = i_3 = i_4 \\ (E^*(\varepsilon_1^*))^2, & \text{two pairs of different indices} \\ 0, & \text{otherwise} \end{cases}
$$

(8.21)

25
and since
\begin{equation}
\hat{\alpha}_\nu(P) = \{\hat{\alpha}_\nu(P) - \alpha_\nu(P)\} + \{\alpha_\nu(P) - \alpha_\nu\} + \{\alpha_\nu\},
\end{equation}
cf. (8.6) for the definition of \(\alpha_\nu(P)\), we obtain by a careful computation from Lemma 8.2 and Lemma 8.3 that the first expectation converges to zero in probability.

Finally bound the second expectation by
\[
\sum_{\|\|k\|\leq M} \sum_{s,l=1}^{T-\|\|k\|} \sum_{\nu,\mu=0}^{\infty} \sum_{\tau,\rho=0}^{\infty} \{\delta_{l,\nu,\rho,\tau,\rho,\tau} \delta_{l+h,\nu,\rho,\tau,\rho,\tau} + \delta_{l,\nu,\rho,\tau,\rho,\tau} \delta_{l+h,\nu,\rho,\tau,\rho,\tau} \}
\]
\[
\left\{ \left( E(\varepsilon_1^*)^2 - 2 (E\varepsilon_1^*)^2 + (E\varepsilon_1^2) \right) + o_P(1) \right\}
\]

Another careful estimation of the various sums above yields that they are bounded in probability.

Since
\[
\left| \left( E(\varepsilon_1^*)^2 - (E\varepsilon_1^*)^2 + (E\varepsilon_1^2) \right) \right| \\
\leq \left( |E\varepsilon_1^*(\varepsilon_1^* - \varepsilon_1)| + |E\varepsilon_1(\varepsilon_1^* - \varepsilon_1)| \right) \cdot O_P(1) \\
\leq d_2(\varepsilon_1^*, \varepsilon_1) \cdot O_P(1) = o_P(1), \text{ cf. Proposition 3.1},
\]
the second expectation is also negligible. This concludes the proof of Theorem 6.4. \[\square\]

**References**


Name der Datei: Kreiss17.tex
Letzte Änderung: 19.03.97 (Kreiss)