SECOND ORDER CORRECTNESS OF THE BLOCKWISE BOOTSTRAP FOR STATIONARY OBSERVATIONS

F. Götze

Fakultät für Mathematik
University of Bielefeld
D-33501 Bielefeld 1, Germany

H.R. Künsch

Seminar für Statistik
ETH Zürich
CH-8092 Zürich, Switzerland

Abstract

We show that the blockwise bootstrap approximation for the distribution of a studentized statistic computed from dependent data is second order correct if we choose the variance estimators such that their biases are similar. For the proof we extend the results of Götze and Hipp (1983) on the validity of the formal Edgeworth expansion for a sum to the studentized mean. The paper concludes with a small simulation example.

Keywords. Edgeworth expansion, studentization, time series, dependent data, strong mixing, resampling.

AMS 1980 Subject classifications. Primary 62M10; secondary 62G15, 62E20.

Running head. Second order correctness.

1 Introduction

Efron's (1979) bootstrap is distribution free, but the assumption of independence is crucial since it ignores the time order of the observations. For serially dependent observations, Künsch (1989) has proposed a blockwise bootstrap which samples blocks of length l of consecutive observations with replacement. He has shown that this procedure estimates the asymptotic variance and the asymptotic normal distribution consistently if the statistic considered is the arithmetic mean or a function thereof. He assumes only suitable mixing and moment conditions for the observations and an increase of $l = l(n)$ to infinity, but at a slower speed than n. However for a simple statistic like a function of an arithmetic mean there are other methods to estimate the asymptotic variance which require fewer computations, e.g the blockwise jackknife or an estimate of the spectrum at zero for the estimated influence function (see also Künsch, 1989). Hence in these cases the additional effort required for the bootstrap is only justified if the resulting approximation to the distribution of the statistic is better than the one relying on asymptotic normality. With i.i.d. data Efron's bootstrap has indeed this property, see e.g. Singh (1981), Hall (1988). Here we show that this is also true for the blockwise bootstrap as announced some time ago in an abstract (Götze and Künsch, 1990). The essential reason for this result is that the blockwise bootstrap distribution has asymptotically the correct skewness (cf. Künsch, 1989, Formula (3.19)). This gain in accuracy is however covered by the error in the bootstrap variance which is in our case of the order $O(l^{-1})+O_p((l/n)^{1/2})$ and thus always larger than the order $O(n^{-1/2})$ of the skewness term. Thus for a better than normal approximation we have to make the true and the bootstrap variance equal, e.g. by studentizing the statistic. When implementing this idea it turns out that it is crucial how the variances are estimated, see Davison and Hall (1993). For our result the essential requirement is that the variance estimators for the original and the bootstrap sample have the same bias.

The most difficult part in proving our result is to show the validity of the Edgeworth expansion for the studentized statistic for dependent observations. Götze and Hipp (1983) have shown the validity of the Edgeworth expansion for the arithmetic mean under a set of conditions covering a broad range of situations. We are going to use the same conditions. The difficulty in passing from the arithmetic to the studentized mean comes from the fact that the estimated variance is the sum of l different arithmetic means, namely the sample covariances for lags $0, \dots, l-1$, and l increases with n. Lahiri (1991) has shown that the blockwise bootstrap is second order correct if the asymptotic variance is known and we use it to standardize the mean. In the case where the observations are m-dependent with m known, his Theorem 2.2 can be used also for the studentized mean, but with a different way of studentization. In this way he avoids the problem with increasing l, but the result is of limited use since typically both the asymptotic variance and m are unknown.

Our result is somewhat surprising because dependence makes the estimation of the distribution of a statistic inherently more difficult. Note that we do not assume a specific type of dependence like an AR-model where second-order correctness can be obtained by applying Efron's bootstrap for the estimated innovations, see Bose (1988). Nevertheless our bootstrap approximation is correct in the $n^{-1/2}$-term like for i.i.d. situations. The effect of dependence is felt only in the following term of the expansion which is at best of the order $O(l/n)$ instead of the usual $O(n^{-1})$

This paper is organized as follows. In Section 2 we set up notations and discuss the heuristics of our result by considering formal Edgeworth expansions. This will also explain why some other ways of studentizing and bootstrapping do not give the desired second order result. In Section 3 we give a rigorous formulation and proof of our result. Finally in Section 4 we present a simulation example.

2 Notations and heuristics

Suppose that we have observed a stationary process $(Y_j, j \in \mathbb{Z})$ with values in $\mathbb{R}^k$ for $j = 1, \dots, n$. Let us denote $\mathbb{E}[Y_j]$ by $\mu, n^{-1}\sum_{j=1}^n Y_j$ by $\overline{Y}_n$ and $n^{1/2}(\overline{Y}_n - \mu)$ by S_n.

We consider a statistic

$$U_n = H(\overline{Y}_n) = H(\mu + n^{-1/2} S_n)$$

as an estimator of $H(\mu)$ where $H : \mathbb{R}^k \to \mathbb{R}$ is a given function. This framework is not as narrow as it may seem at first glance. We can replace the original observations Y_j by

$$Y_j' \triangleq \varphi(Y_j, Y_{j+1}, \dots, Y_{j+m-1}) \tag{1}$$

with a fixed m and φ and apply everything to the transformed observations Y_j'. So for instance the following version of the sample autocovariances

$$\hat{\gamma}_m = n^{-1} \sum_{j=1}^n Y_j Y_{j+m} - n^{-1} \sum_{j=1}^n Y_j n^{-1} \sum_{j=1}^n Y_{j+m}$$

fits into our framework. By taking nonlinear functions of $\hat{\gamma}_m$'s, we arrive at autocorrelations, partial autocorrelations and Yule-Walker estimators in autoregressive processes.

If H is differentiable and the gradient DH does not vanish at μ, then $n^{1/2}(U_n - H(\mu))$ is asymptotically equivalent to

$$DH(\mu)^T S_n = n^{-1/2} \sum_{i=1}^n Z_i \triangleq T_n$$

where

$$Z_i \triangleq DH(\mu)^T (Y_i - \mu).$$

Moreover under standard assumptions about moments of (Z_i) and the decay of mixing coefficients, T_n is asymptotically $\mathcal{N}(0, \sigma_\infty^2)$- distributed with

$$\sigma_\infty^2 \triangleq \sum_{j=-\infty}^{\infty} \mathbb{E}[Z_0 Z_j] = \lim_n \sigma_n^2 \tag{2}$$

where

$$\sigma_n^2 \triangleq Var(T_n) = \sum_{j=-n}^{n} (1 - |j|/n) \mathbb{E}[Z_0 Z_j].$$

For studentizing U_n we need an estimator of σ_∞^2. We estimate the covariances $\mathbb{E}[Z_0 Z_j]$ and then truncate the infinite sum corresponding to (2) with some lag weights w_k such that $w_0 = 1$:

$$\hat{\sigma}_n^2 = DH(\overline{Y}_n)^T \sum_{k=0}^{l-1} w_k \; n^{-1} \sum_{j=1}^{n-l} (Y_j - \overline{Y}_n)(Y_{j+k} - \overline{Y}_n)^T DH(\overline{Y}_n). \tag{3}$$

Note that $\hat{\sigma}_n$ can be computed using only the observations $Y_1, \ldots, Y_n$. If $w_k \to 2$ for all $k \neq 0$, $l \to \infty$ and $ln^{-1} \to 0$, then $\hat{\sigma}_n^2$ is consistent under mild conditions on (Y_j). The studentized statistic is now

$$U_{n,stud} = n^{1/2}(U_n - H(\mu))\hat{\sigma}_n^{-1}$$

which is asymptotically $\mathcal{N}(0, 1)-$ distributed.

Our aim is to produce a better than normal approximation of the distribution of $U_{n,stud}$ with the help of the bootstrap. The blockwise bootstrap of Künsch (1989) goes as follows. It depends on a block length which we choose equal to the width l of the lag window for the estimated variance $\hat{\sigma}_n^2$ in (3). The reason for this will be explained shortly. We assume for simplicity also that $n = bl$ with $b \in \mathbb{N}$. Then we generate bootstrap samples by resampling from all l-tuples of consecutive observations in the original sample. This means that the bootstrap sample $(Y_1^*, \ldots, Y_n^*)$ is

$$Y^*_{(j-1)l+i} = Y_{N_j+i} \quad (1 \leq j \leq b, 1 \leq i \leq l)$$

where the starting points of the blocks $N_1, \ldots, N_b$ are i.i.d. uniform on $\{0, 1, \ldots, n-l\}$. The bootstrap statistic is then

$$U_n^* = H(\overline{Y}_n^*).$$

We will denote the conditional expectation of any function of the Y_j^*'s for given $Y_1, \ldots, Y_n$ by $\mathbb{E}^*$. Because the $N_1, \ldots, N_b$ have the same distribution, a simple calculation similar to Lemma 3.1 of Künsch (1989) shows that

$$\mu_n^* \triangleq \mathbb{E}^*[\overline{Y}_n^*] = l^{-1}\mathbb{E}^*[Y_1^* + \ldots + Y_l^*] = (n-l+1)^{-1} \sum_{j=0}^{n-l} l^{-1} \sum_{i=1}^{l} Y_{j+i}$$

$$= (n-l+1)^{-1} \sum_{j=1}^{n} \min(j/l, 1, (n+1-j)/l) Y_j = \overline{Y}_n + O_p(l^{1/2} n^{-1}).$$

Hence $n^{1/2}(\overline{Y}_n^* - \overline{Y}_n)$ has a (conditional) bias of order $O(l^{1/2}n^{-1/2})$ which is larger than the first term in the Edgeworth expansion of $n^{1/2}(\overline{Y}_n - \mu)$. So for second order correctness we must center U_n^* at $H(\mu_n^*)$ instead of $H(\overline{Y}_n)$.

Let us introduce the quantities

$$\begin{aligned} A_j &\triangleq l^{-1/2} \sum_{i=1}^{l} (Y_{j+i} - \mu_n^*), \\ S_n^* &\triangleq n^{1/2}(\overline{Y}_n^* - \mu_n^*) = b^{-1/2} \sum_{j=1}^{b} A_{N_j}. \end{aligned}$$

By a Taylor expansion of H at μ_n^* we see that the linear part of $n^{1/2}(U_n^* - H(\mu_n^*))$ is

$$T_n^* \triangleq DH(\mu_n^*)^T S_n^* = b^{-1/2} \sum_{j=1}^{b} B_{N_j}$$

where $B_j \triangleq DH(\mu_n^*)^T A_j$. By the independence of $N_1, \ldots, N_b$ it follows that

$$\sigma_n^{*2} \triangleq Var^*[T_n^*] = Var^*(B_{N_1}) = (n-l+1)^{-1} \sum_{j=0}^{n-l} B_j^2.$$

A straightforward computation shows that up to boundary effects, σ_n^{*2} is equal to $\hat{\sigma}_n^2$ of (3) with $w_k = 2(1-k/l)$. This is the reason why we choose the window width in (3) and the block length of the bootstrap to be equal. However our results could be obtained also without this special choice. Asymptotic normality of the linear part T_n^* can be proved using Lindeberg's condition. So essentially first order correctness of the bootstrap follows from the convergence of σ_n^{*2} to σ_∞^2, see Künsch (1989). Because $\sigma_n^{*2} - \sigma_n^2$ is always at least of the order $O_p(n^{-1/3})$ (except when the Z_j's are uncorrelated), removing the error due to the wrong variance is of interest.

The obvious estimate of σ_n^{*2} is

$$\hat{\sigma}_n^{*2} \triangleq b^{-1} \sum_{j=1}^{b} \hat{B}_{N_j}^2 \tag{4}$$

where

$$\hat{B}_j \triangleq l^{-1/2} \sum_{i=1}^{l} DH(\overline{Y}_n^*)^T (Y_{j+i} - \overline{Y}_n^*).$$

Then the studentized bootstrap statistic is

$$U_{n,stud}^* = n^{1/2}(U_n^* - H(\mu_n^*))/\hat{\sigma}_n^*.$$

Note that in using $\hat{B}_j$ instead of B_j we have mimicked the uncertainty of estimating μ by $\overline{Y}_n$ in (3). However in other aspects $\hat{\sigma}_n^{*2}$ does not correspond exactly to $\hat{\sigma}_n^2$. Namely we can write

$$\hat{\sigma}_n^{*2} = n^{-1} DH(\overline{Y}_n^*)^T \sum_{[i/l]=[j/l]} (Y_i^* - \overline{Y}_n^*)(Y_j^* - \overline{Y}_n^*) DH(\overline{Y}_n^*) \tag{5}$$

where $[x]$ denotes the integer part of x. So for $\hat{\sigma}_n^{*2}$ we only take pairs in the same blocks whereas for $\hat{\sigma}_n^2$ we take all pairs at lag distance $< l$ with a lag weight w_k. The reason for this difference is that for H linear $\mathbb{E}^*[\hat{\sigma}_n^{*2}] = \sigma_n^{*2}(1 - b^{-1})$ whereas the analogue of (5) with the original observations would have a bias of the order $O(l^{-1})$. For second order correctness we will need both l and the bias of $\hat{\sigma}_n^2$ to be $o(n^{-1/2})$, see below.

We now derive formal Edgeworth expansions for $U_{n,stud}$ and $U_{n,stud}^*$. First we approximate $U_{n,stud}$ and $U_{n,stud}^*$ by quadratic statistics. A Taylor expansion of the numerator and denominator of $U_{n,stud}$ gives

$$n^{1/2}(U_n - H(\mu)) = T_n + n^{-1/2}\frac{1}{2} S_n^T D^2 H(\mu) S_n + n^{-1}\xi_{n,1}$$

and

$$\hat{\sigma}_n^2 = \sigma_n^2 + n^{-1/2}(V_n + S_n^T d) + (\tau_n^2 - \sigma_n^2) + n^{-1}\xi_{n,2}$$

where

$$\begin{aligned}
\tau_n^2 &\triangleq \sum_{k=0}^{l} w_k \mathbb{E}[Z_0 Z_k], \\
V_n &\triangleq \sum_{k=0}^{l} w_k n^{-1/2} \sum_{j=1}^{n-l} (Z_j Z_{j+k} - \mathbb{E}[Z_0 Z_k]) \\
d &\triangleq 2D^2 H(\mu) \sum_{k=-\infty}^{\infty} \mathbb{E}[(Y_0 - \mu)(Y_k - \mu)^T] DH(\mu).
\end{aligned}$$

Hence

$$\begin{aligned}
U_{n,stud} &= T_n \sigma_n^{-1} + n^{-1/2}(P(S_n) - \frac{1}{2} T_n V_n \sigma_n^{-3}) - \frac{1}{2} T_n(\tau_n^2 - \sigma_n^2)\sigma_n^{-3} + n^{-1}\xi_{n,3} \\
&\triangleq E_n - \frac{1}{2} T_n (\tau_n^2 - \sigma_n^2)\sigma_n^{-3} + n^{-1}\xi_{n,3} \quad \text{say,}
\end{aligned} \tag{6}$$

where

$$P(S_n) \triangleq \frac{1}{2} S_n^T D^2 H(\mu) S_n \sigma_n^{-1} - \frac{1}{2} T_n S_n^T d\ \sigma_n^{-3}.$$

Because $S_n = O_p(1)$ and $V_n = O_p(l^{1/2})$, the remainder terms are $\xi_{n,1} = O_p(1), \xi_{n,2} = O_p(l)$ and $\xi_{n,3} = O_p(l)$. Hence we need at least $l = o(n^{1/2})$ in order that the remainder terms do not contribute to an expansion up to $O(n^{-1/2})$.

The difference $\tau_n^2 - \sigma_n^2$ reflects the bias of $\hat{\sigma}_n^2$ as an estimator of σ_n^2 when H is linear. Its order depends on the weights w_k. The lowest order is achieved if $w_k \equiv 2$ for $0 < k < l$. Then $\tau_n^2 - \sigma_n^2$ is of the order $O(\sum_{k=l}^{\infty} |\mathbb{E}[Z_0 Z_k]|) + O(n^{-1})$. This is $O(n^{-1})$ under the assumptions of Section 3 provided l increases at least as fast as $c\log(n)$ with a big enough constant c. However with these weights $\hat{\sigma}_n^2$ is not necessarily positive. Weights guaranteeing positivity behave like $w_k \sim 2(1 - c(k/l)^2)$ with the constant $c > 0$. Then $\tau_n^2 - \sigma_n^2$ is $O(l^{-2})$, so l has to increase faster than $n^{1/4}$, preferably close to $n^{1/3}$.

Similarly we have

$$\begin{aligned} U_{n,stud}^* &= T_n^* \sigma_n^{*-1} + n^{-1/2}(P^*(S_n^*) - \frac{1}{2} l^{1/2} T_n^* V_n^* \sigma_n^{*-3}) + b^{-1}\xi_{n,3}^* \\ &\triangleq E_n^* + b^{-1}\xi_{n,3}^* \quad \text{say,} \end{aligned}$$

where

$$\begin{aligned} P^*(S_n^*) &\triangleq \frac{1}{2} S_n^{*T} D^2 H(\mu^*) S_n^* \sigma_n^{*-1} - \frac{1}{2} T_n^* S_n^{*T} d^* \sigma_n^{*-3}, \\ d^* &\triangleq 2D^2 H(\mu_n^*) \mathbb{E}^*[A_{N_1} A_{N_1}^T] DH(\mu_n^*), \\ V_n^* &\triangleq b^{-1/2} \sum_{j=1}^{b} (B_{N_j}^2 - \sigma_n^{*2}). \end{aligned}$$

The remainder term $\xi_{n,3}^*$ is $O_p^*(1)$ in probability.

For the formal Edgeworth expansion of $U_{n,stud}$ we introduce

$$\kappa_n \triangleq n^{1/2}\mathbb{E}[T_n^3]$$

and

$$\alpha_n \triangleq \mathbb{E}[T_n V_n] = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n-l} \sum_{k=0}^{l} w_k \mathbb{E}[Z_i Z_j Z_{j+k}].$$

Furthermore let Σ_n denote the covariance matrix of the $(k+1)$- dimensional vector $(Z_j, Y_j^T)^T$ and c_α the coefficients of the polynomial $P(S_n)$ with respect to the variables S_n and T_n. The formal Edgeworth expansion $\psi_n(a)$ of $U_{n,stud}$ is the same as the one for E_n. It is defined by its characteristic function

$$\begin{aligned} \hat{\psi}_n(t) &= (1 + n^{-1/2}(\frac{1}{6}\kappa_n \sigma_n^{-3/2}(it)^3 - \frac{1}{2}\alpha_n \sigma_n^{-3/2} it)) \exp(-t^2/2) \\ &\quad + n^{-1/2} it \sum_{\alpha} c_\alpha (-1)^{|\alpha|} D_w^\alpha \exp(-w^T \Sigma_n w/2) \mid_{w=(t,0,\ldots,0)} . \end{aligned} \tag{7}$$

Similarly the formal Edgeworth expansion of $U_{n,stud}^*$ (conditional on the observations) is defined by its characteristic function

$$\begin{aligned} \hat{\psi}_n^*(t) &= (1 + n^{-1/2}\kappa_n^* \sigma_n^{*-3/2}(\frac{1}{6}(it)^3 - \frac{1}{2}it)) \exp(-t^2/2) \\ &\quad + n^{-1/2} it \sum_{\alpha} c_\alpha^* (-1)^{|\alpha|} D_w^\alpha \exp(-w^T \Sigma_n^* w/2) \mid_{w=(t,0,\ldots,0)} . \end{aligned}$$

Here

$$\kappa_n^* \triangleq n^{1/2}\mathbb{E}^*[T_n^{*3}] = l^{1/2}\mathbb{E}^*[T_n^* V_n^*] = l^{1/2}\mathbb{E}^*[B_{N_1}^3]$$

and Σ_n^* and c_α^* are defined in analogy to Σ_n and c_α. Hence the two formal Edgeworth expansions are close if all coefficients are close, i.e. the moments up to order 3 should be close. From Künsch (1989) and the discussion above it follows that

$$\begin{aligned}
\mu_n^* &= \mu + O_p(n^{-1/2}), \\
\Sigma_n &= \Sigma_\infty + O(n^{-1}), \\
\Sigma_n^* &= \Sigma_\infty + O(l^{-1}) + O_p(b^{-1/2}).
\end{aligned}$$

Moreover it is easily seen that

$$\alpha_n = \kappa_n + O(n^{-1}) = \kappa_\infty + O(n^{-1})$$

where

$$\kappa_\infty \triangleq \sum_{i,j=-\infty}^{\infty} \mathbb{E}[Z_0 Z_i Z_j].$$

Finally

$$\kappa_n^* = l^{1/2}(n-l+1)^{-1}\sum_{j=0}^{n-l} B_j^3.$$

Hence because B_j is a standardized sum of length l we have

$$\mathbb{E}[\kappa_n^*] = \kappa_\infty + O(l^{-1}).$$

In Section 3 we will show that

$$\kappa_n^* - \mathbb{E}[\kappa_n^*] = O_p(l^{1/2}b^{-1/2}) = O_p(ln^{-1/2}).$$

Taking all this together we obtain the order of the difference between the two Edgeworth expansions:

$$\sup_a |\psi_n(a) - \psi_n^*(a)| = O(l^{-1}n^{-1/2}) + O_p(ln^{-1}). \tag{8}$$

The error between the Edgeworth expansions and the distribution functions will be shown to be

$$\sup_a |\mathbb{P}[U_{n,stud} \le a] - \psi_n(a)| = O(ln^{-1+\epsilon}) + O(\tau_n^2 - \sigma_n^2) \tag{9}$$

and

$$\sup_a |\mathbb{P}^*[U_{n,stud}^* \le a] - \psi_n^*(a)| = O_p(ln^{-1+\epsilon}) \tag{10}$$

for any $\epsilon > 0$ provided all moments exist. Together equations (8) - (10) give the desired result that the studentized bootstrap is second order correct. Proofs will be given in the next section.

We close this section with a brief discussion of the choice of l. We have already discussed how the term $O(\tau_n^2 - \sigma_n^2)$ can be made small. In order to get a minimal order for all the other terms, the optimal choice is $l = O(n^{1/4})$. Note that this is different from the optimal order for σ_n^{*2} which is $l = O(n^{1/3})$.

3 Rigorous results and proofs

3.1 Edgeworth expansion for the studentized statistic

We shall assume that the sequence of random vectors $R_j \stackrel{\triangle}{=} (Z_j, Y_j) \in \mathbb{R} \times \mathbb{R}^k$ satisfies the following conditions used in Götze and Hipp (1983) which will be denoted by GH for short in the sequel.

(A1)

$$\mathbb{E}R_j = 0, \quad j = 1, 2, \ldots .$$

(A2)

$$\beta_s \stackrel{\triangle}{=} \mathbb{E}||R_j||^s < \infty \quad \text{for some} s \geq 5.$$

(A3) There exists a sequence $\mathcal{D}_k$, $k \in \mathbb{Z}$ of sub σ-fields of $\mathcal{A}$ and a constant $d > 0$ such that for $j, m = 1, 2, \ldots$ with $m > d^{-1}$ the rv. R_j can be approximated by a $\mathcal{D}_{j-m,j+m} \stackrel{\triangle}{=} \sigma(\mathcal{D}_p : |p - j| \leq m)$ measurable random vector $\bar{Y}_{j,m}$ with

$$\mathbb{E}||R_j - \bar{Y}_{j,m}|| \leq d^{-1} \exp[-dm].$$

(A4) There exists a $d > 0$ such that for all $m, j = 1, 2, \ldots$ $A \in \mathcal{D}_{-\infty,j}, B \in \mathcal{D}_{j+m,\infty}$

$$|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \leq d^{-1} \exp[-dm].$$

(Rosenblatt mixing)

(A5) There exists a $d > 0$ such that for all $m, j = 1, 2, \ldots, d^{-1} < m < j$ and $|t| \geq d$

$$\mathbb{E}|\mathbb{E}(\exp[it(Z_{j-m} + \ldots + Z_{j+m})]|\mathcal{D}_l : l \neq j)| \leq \exp[-d]$$

and

$$\liminf_n Var(Z_1 + \ldots + Z_n)/n > 0$$

(A6) There exists a $d > 0$ such that for all $m, j, p = 1, 2, \ldots$ and $A \in \mathcal{D}_{j-p,j+p}$

$$\mathbb{E}|\mathbb{P}(A|\mathcal{D}_l : l \neq j) - \mathbb{P}(A|\mathcal{D}_l : 0 < |l - j| \leq m + p)| \leq d^{-1} \exp[-dm].$$

Moreover for the function H which defines our statistic U_n we make the following assumption.

(A7) $H : \mathbb{R}^k \to \mathbb{R}$ is three times differentiable, $DH(\mu) \neq 0$, and there are constants $C, A > 0$ such that

$$\|D^3H(x)\| \leq C(1 + \|x\|^A) \quad \text{for every } x \in \mathbb{R}^k$$

Condition (A1) means no restriction since we can always put $\mu = 0$. Our main result is now

Theorem 3.1 *Under the conditions (A1)-(A7) the Edgeworth approximation for $U_{n,stud}$ defined in (7) holds, i. e. for $s \geq 8$, $l \leq n^{1/3}$ we have*

$$\sup_a |\mathbb{P}(U_{n,stud} \leq a) - \psi_n(a)| = O(ln^{-1+2/s})$$

provided $\omega_j \equiv 2$. If $\omega_j \sim 2(1-(j/l)^2)$ the error contains an additional $O(l^{-2})$-term.

Remark 3.1 *A direct application of the multivariate result in GH for stochastic expansions is not very desirable although it yields expansions up to arbitrary degree. But we would have to check the conditional Cramér condition (A5) for the vector (S_n, V_n) which can be very difficult even for simple time series models, in particular when the statistic U_n is constructed from transformed observations as in (1). For examples where (A5) is verified, see Bose (1988) and Götze and Hipp (1992).*

Remark 3.2 *The proof of this result can be carried out in the case of random fields under obvious reformulations of the conditions $(A1-A4)$ without substantial modifications, for example the independence $(A3)$ reads*

$$|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \leq ce^{-dm} \tag{11}$$

for $A \in \sigma(I_1), B \in \sigma(I_2), I_1, I_2 \subset \mathbb{Z}^2$ with $p(I_1, I_2) \triangleq \min_{u \in I_1, v \in I_2} ||u-v|| \geq m$, where $||\cdot||$ denotes the supremum norm. For details see Jensen (1991). Thus we have an analogous result for statistics $U_n \triangleq H(n^{-2}\sum_{i,j=1}^n Z_{ij})$

Proof. In order to limit the moment conditions we have to use truncations as in the proofs of GH. Define $T : \mathbb{R}^{k+1} \to \mathbb{R}^{k+1}$ by

$$T(x) \triangleq xn^\beta \psi(||x||n^{-\beta})/||x||$$

where $\psi \in C^\infty(0,\infty)$ satisfies $\psi(r) \triangleq r$, for $r \leq 1$, ψ is increasing and $\psi(r) \equiv 2$ for $r \geq 2$. The value of $\frac{1}{2} > \beta > 0$ will be determined later. Define $R_j^\dagger \triangleq (Z_j^\dagger, Y_j^\dagger) \triangleq T(R_j)$. Let $U_{n,stud}^\dagger$ denote the statistic $U_{n,stud}$ applied to $R_j^\dagger, j = 1, \ldots, n$.

Note that

$$0 = \mathbb{E}R_j = \mathbb{E}R_j 1_{\{||R_j||<n^\beta\}} + \mathbb{E}R_j 1_{\{||R_j|| \geq n^\beta\}} = J_1 + J_2,$$

say, where $||J_2|| = O(n^{-(s-1)\beta})$ by Chebyshev's inequality. Hence we obtain

$$\begin{aligned} ||\mathbb{E}R_j^\dagger|| &= O(n^{-(s-1)\beta}), \\ ||\mathbb{E}S_n^\dagger|| &= O(n^{1/2-(s-1)\beta}) = O(n^{-1-\varepsilon}) \\ ||\mathbb{E}V_n^\dagger|| &= O(ln^{\frac{1}{2}-(s-1)\beta}) = O(n^{-1-\varepsilon}) \end{aligned} \tag{12}$$

choosing β such that

$$(s-1)\beta > 3/2 + \varepsilon, \quad s\beta > 2 + \varepsilon. \tag{13}$$

Then we have $\mathbb{P}(\exists \ 1 \le j \le n, \quad ||R_j|| > n^\beta) = O(n^{1-\beta s}) = o(n^{-1})$. These estimates together with the arguments in Lemma 3.30 of GH finally lead to

$$|\kappa_p(a^T S_n) - \kappa_p(a^T S_n^\dagger)| = O(n^{1-p/2-(s-p)\beta})$$

for the cumulants of order $p = 2, 3, \ldots, s$. By expansion of moments in terms of cumulants we get similar as on p. 224 and in Lemma 3.18 of GH or Bulinskii and Zhurbenko (1976)

$$\sup_n \mathbb{E}||S_n||^s < \infty, \quad \sup_n \mathbb{E}||S_n^\dagger||^p \le c(p) \tag{14}$$

for every $p \ge 2$. Furthermore,

$$\mathbb{E}|V_n|^{s/2} = O(l^{s/4}) \quad \text{and} \quad \mathbb{E}|V_n^\dagger|^p = O(l^{p/2}) \tag{15}$$

for any p. Thus we obtain for the Edgeworth expansion $\psi_n^\dagger(a)$ for $U_{n,stud}^\dagger$ defined similarly as $\psi_n(a)$ by standard technical arguments (see Bhattacharya and Rao (1986))

$$\sup_a |\psi_n(a) - \psi_n^\dagger(a)| = O(n^{-1})$$

and therefore

$$\begin{aligned} \Delta_n \ &\triangleq \ \sup_a |\mathbb{P}(U_{n,stud} \le a) - \psi_n(a)| \\ &\le \sup_a |\mathbb{P}(U_{n,stud} \le a) - \psi_n^\dagger(a)| + O(n^{-1}) \\ &\le \sup_a |\mathbb{P}(U_{n,stud}^\dagger \le a) - \psi_n^\dagger(a)| + O(n^{-1}). \end{aligned} \tag{16}$$

Next we argue that the Edgeworth expansions for $U_{n,stud}^\dagger$ and for $E_{n,\tau}^\dagger \triangleq E_n^\dagger + T_n^\dagger \sigma_n^{-1}(\tau_n^2 - \sigma_n^2)\sigma_n^{-2}$ agree up to the order $O(ln^{-1+\epsilon})$, i.e. we can neglect the remainder term $n^{-1}\xi_{n,3}$ in the Taylor expansion (6). By standard arguments (see e.g. Chibisov,

1972) this is true if $\mathbb{P}(|\xi^{\dagger}_{n,3}| > ln^{\epsilon}) = O(ln^{-1+\epsilon})$. This follows from the structure of the remainder terms in the Taylor expansions and Chebychev's inequality since $S_n^{\dagger}$ and $V_n^{\dagger}$ have moments of all orders with bounds given by (14 – 15).Finally, since the Edgeworth expansion of $E_n^{\dagger}$ and of $E^{\dagger}_{n,\tau}$, say $\psi^{\dagger}_{n,\tau}$, differ by $O(\tau_n^2 - \sigma_n^2))$ we obtain

$$\Delta_n \leq \sup_a |\mathbb{P}(E^{\dagger}_{n,\tau} \leq a) - \psi^{\dagger}_{n,\tau}(a)| + O(\tau_n^2 - \sigma_n^2)) + O(ln^{-1+\epsilon}). \tag{17}$$

In order to simplify the notation we shall replace in the following $E^{\dagger}_{n,\tau}$ by $E_n^{\dagger}$ and $\psi^{\dagger}_{n,\tau}$ by $\psi_n^{\dagger}$ since both statistics behave exactly similar in terms of estimating their c.f. By the Berry-Esseen Lemma we can estimate the error in the Edgeworth expansion by characteristic functions. Let $\varphi_n^{\dagger}(t) \triangleq \mathbb{E}\exp(itE_n^{\dagger})$. Then (15) together with (17) leads to

$$\begin{aligned}\Delta_n &\leq c\Big(\int_{|t|\leq n^{\varepsilon}} + \int_{n^{\varepsilon}<|t|<n^{1-2/s}} |\varphi_n^{\dagger}(t) - \hat{\psi}_n^{\dagger}(t)||t|^{-1}dt + O(ln^{-1+\varepsilon})\\ &= I_1 + I_2, \quad \text{say.}\end{aligned} \tag{18}$$

Estimation of I_1:

By Lemma 3.33 of G.H. we have

$$|\mathbb{E}\exp(itT_n^{\dagger}) - \hat{\psi}^{\dagger}_{n,lin}(t)| \leq cn^{-1}(1+\beta_s)(1+|t|^{12})\exp[-ct^2] \tag{19}$$

for some absolute constant $c > 0$, where

$$\psi^{\dagger}_{n,lin} \triangleq \exp[-\frac{1}{2}t^2](1 + n^{-\frac{1}{2}}(it)^3\frac{\kappa_3^{\dagger}}{6}/\sigma_n^{-3/2}).$$

Now expanding $\mathbb{E}\exp[itE_n^{\dagger}]$ in terms of $tn^{-1/2}(P(S_n^{\dagger}) - T_n^{\dagger}V_n^{\dagger}/2)$ we obtain

$$\begin{aligned}\mathbb{E}\exp[itE_n^{\dagger}] &= \mathbb{E}\exp[itT_n^{\dagger}] + itn^{-\frac{1}{2}}\mathbb{E}\exp[itT_n^{\dagger}](P(S_n^{\dagger}) - \frac{1}{2}T_n^{\dagger}V_n^{\dagger})\\ &\quad + O(t^2n^{-1})\left(\mathbb{E}P(S_n^{\dagger})^2 + \mathbb{E}(T_n^{\dagger}V_n^{\dagger})^2\right)\\ &\triangleq \mathbb{E}\exp[itT_n^{\dagger}] + J_{1,t} + J_{2,t} + R_t, \quad \text{say.}\end{aligned} \tag{20}$$

By (14) we obtain $\sup_n \mathbb{E}P(S_n^{\dagger})^2 \leq c < \infty$ and expanding $\mathbb{E}T_n^{\dagger 2}V_n^{\dagger 2}$ into individual summands we obtain by condition (A3) and (A4) similar as in Bulinskii and Zhurbenko (1976) by counting multiplicities.

$$\mathbb{E}T_n^{\dagger 2}V_n^{\dagger 2} = O(l). \tag{21}$$

Furthermore, let $T_I^\dagger \triangleq \sum_{j\in I} Z_j^\dagger n^{-1/2}$, $T_{n,I}^\dagger = T_n^\dagger - T_I^\dagger$ and $H_{p,\nu,j}^\dagger \triangleq \frac{1}{2}\sigma_n^{-3/2}\omega_\nu\ (Z_p^\dagger Z_{p+\nu}^\dagger - \mathbb{E} Z_p Z_{p+\nu}) Z_j^\dagger$. We have

$$J_{2,t} = \frac{-it}{n^{3/2}\sigma_n} \sum_{p,\nu,j}{}^{*} H_{p,\nu,j}^\dagger \exp[itT_I^\dagger]\exp[itT_{n,I}^\dagger]$$

where $\sum^*$ denotes the summation over $1 \le j \le n$, $1 \le p \le n-l$, and $1 \le \nu \le l$ and where I denotes the set of indices in $< 1, n >$ which have distance less than m from $\{j, p, p+\nu\}$ with $m \triangleq K \log n$, for some K to be chosen later. Using an expansion technique introduced by Tikhomirov (1980) we obtain by expanding $\exp[itT_I^\dagger]$ up to second order

$$\begin{aligned} J_{2,t} &= \frac{-it}{n^{3/2}} \sum_{p,\nu,j}{}^{*} \mathbb{E} H_{p,\nu,j}^\dagger (1 + itT_I^\dagger + \frac{(it)^2}{2} T_I^{\dagger 2}) \exp[itT_{n,I}^\dagger] \\ &\quad + O\left(\frac{t^4}{n^{3/2}} \sum_{p,\nu,j}{}^{*} \mathbb{E} |H_{p,\nu,j}^\dagger T_I^{\dagger 3}| \right) \\ &\triangleq J_{3,t} + J_{4,t} + J_{5,t} + R_{6,t}. \end{aligned}$$

In the term $R_{6,t}$ we have to estimate ln^2m^3 summands separately. Thus $R_{6,t} = O(lm^3t^4n^{-1})$. For the other terms we shall employ the approximate independence of e.g. $H_{p,\nu,j}^\dagger$ and $\exp[itT_{n,I}^\dagger]$ using conditions (A3) and (A4) in order to factor $J_{3,t}$ as a product of expectations plus an error of order $O(n^C \exp[-dm])$. In the terms $J_{4,t}$ and $J_{5,t}$ we have to split the set I into two resp. three parts according to the distance to I^c, e.g. write $I = I_0 \cup I_1 \cup I_2$ where

$$I_p \triangleq \{ j \in I : \frac{3-p}{3} m \ge |j - I^c| > \frac{2-p}{3} \}, \; p = 0, 1, 2$$

and split the term $J_{4,t}$ resp. $J_{5,t}$ according to $T_I^\dagger = \sum_j T_{I_j}^\dagger$ resp. $T_I^{\dagger 2} = \sum_{j,j'} T_{I_j}^\dagger T_{I'_j}^\dagger$. The new terms may be for example of the form $H_{p,\nu,j}^\dagger T_{I_j}^\dagger T_{I'_j}^\dagger \exp[itT_{n,I}^\dagger]$ and can *always* be written as a product of two factors which are approximately independent with an error of order $O(n^C \exp[-dm/3])$ due to a 'gap' of length $m/3$. (Compare GH Lemma 3.17, p.228). Thus we get $J_{4,t} = O(\frac{t^2}{n} m + n^C \exp[-dm/3])$ and $J_{5,t} = O(t^3 n^{-1} m^2 + n^C \exp[-dm/3])$. These estimates together imply

$$\begin{aligned} J_{2,t} &= \frac{-it}{2n^{3/2}\sigma_n^{3/2}} \left(\sum_{\mu,\nu} \omega_\nu \mathbb{E} Z_1^\dagger Z_{1+\nu}^\dagger Z_{1+\nu+\mu}^\dagger \right) \mathbb{E}\exp[itT_n^\dagger] \\ &\quad + O(\left(\frac{t}{n^{3/2}} \sum_{p,\nu,j}{}^{*} \mathbb{E} \left| H_{p,\nu,j}^\dagger (\mathbb{E}\exp[itT_n^\dagger] - \mathbb{E}\exp[itT_{n,I}^\dagger]) \right| \right) \\ &\quad + O(\frac{t^4}{n} m^3 l) \end{aligned} \qquad (22)$$

where the summations extend over $|\nu| \leq m$ and $|\mu| \leq m$. Finally the second line in (22), say $J_{7,t}$, can be handled easily by estimating $\exp[itT_l^\dagger] - 1$ using (14)-(15) . This leads to $J_{7,t} = O(t^2 n^{-1} m^{1/2})$.

Furthermore, in order to estimate $J_{1,t}$ we obtain by Lemma 3.33 of GH for the multivariate expansion of $\mathbb{E}\exp[ia^T S_n^\dagger]$ in terms of the multivariate c.f. expansion $\hat{\psi}_{n,mult}(a)$ of length 2 with first component of a being equal to t:

$$\begin{aligned}
& |\mathbb{E} P(S_n^\dagger)\exp[itT_n^\dagger] - \sum_\alpha c_\alpha D^\alpha \hat{\psi}_{n,mult}(a)| \\
\leq\ & \sum_\alpha |c_\alpha||D^\alpha(\mathbb{E}|\exp[ia^T S_n^\dagger] - \hat{\psi}_{n,mult}(a))| \\
\leq\ & c(1+\sup_j \mathbb{E}||R_j^\dagger||^4)(1+|t|^{6+|\alpha|})e^{-ct^2} n^{-\frac{1}{2}-\varepsilon},
\end{aligned} \tag{23}$$

where $c > 0$ denotes an absolute constant and the partial derivatives are taken at the point $(t, 0, 0, \ldots, 0)$. The coefficients c_α correspond to the monomial coefficients of the polynomial $P(S_n^\dagger)$. Collecting the expansion results $(19-23)$ we get by definition of the expansion $\hat{\psi}_n^\dagger(t)$

$$\begin{aligned}
|\varphi_n^\dagger(t) - \hat{\psi}_n^\dagger(t)| \ \leq\ & |\mathbb{E}\exp[itT_n^\dagger] - \hat{\psi}_{n,lin}^\dagger(t)| \\
& + n^{-\frac{1}{2}}|t||\mathbb{E}P(S_n^\dagger)\exp[itT_n^\dagger] - \sum_\alpha c_\alpha D^\alpha \hat{\psi}_{n,mult}(a)| \\
& + \frac{1}{2}|t| n^{-\frac{1}{2}}|\alpha_3^\dagger||\mathbb{E}\exp[itT_n^\dagger] - \exp[-\frac{1}{2}t^2]| \\
& + O(\frac{t^2}{n}m) + O(\frac{t^2}{n}l)
\end{aligned} \tag{24}$$

where $\alpha_3^\dagger \triangleq \sigma_n^{-1} \sum_\mu \sum_\nu \omega_\nu \mathbb{E} Z_1^\dagger Z_{1+\nu}^\dagger Z_{1+\nu+\mu}^\dagger$.

Hence we obtain by integration over $|t| \leq n^\varepsilon$

$$I_1 = O(l n^{-1+2\varepsilon}).$$

Estimation of I_2:

Let N denote a number such that $n \geq N \geq m$ where $m \triangleq K\log n$ and (K is chosen sufficiently large below) such that for $n^\varepsilon < |t| < n^{1-\varepsilon}$

$$N \triangleq [(\frac{n}{t^2}+1)m^2]. \tag{25}$$

The reasons for this choice will become apparent later in the proof.

Now we replace the random vector $R_j \triangleq (Z_j, Y_j)$ by some $\mathcal{D}_{j-m,j+m}$-measurable random vectors $R_j^{\dagger\dagger} = (Z_j^{\dagger\dagger}, Y_j^{\dagger\dagger})$ with

$$\mathbb{E}||R_j - R_j^{\dagger\dagger}|| < d^{-1}\exp[-dm]$$

using condition (A3). Hence also

$$\mathbb{E}||T(R_j) - T(R_j^{\dagger\dagger})|| < d^{-1}\exp[-dm].$$

Denote by $E_n^{\dagger\dagger}$ the statistic E_n for truncated and $\mathcal{D}_{j-m,j+m}$ measurable vectors $T(R_j^{\dagger\dagger})$. We obtain by the truncation of random variables

$$\begin{aligned} &|\mathbb{E}\exp[itE_n^{\dagger}] - \mathbb{E}\exp[itE_n^{\dagger\dagger}]| \\ &\leq n^A \sup\nolimits_j \mathbb{E}||T(R_j) - T(R_j^{\dagger\dagger})|| = O(n^{-2}) \end{aligned} \tag{26}$$

for some $A > 0$ provided K in the definition of m is chosen sufficiently large. Thus we may assume w.l.g. that all r.v. are truncated by n^β and are $\mathcal{D}_{j-m,j+m}$-measurable. In order to simplify the notation we shall from now on drop the †-notation and write $R_j \triangleq (Z_j, Y_j), S_n, T_n, V_n$ again.

We decompose

$$(S_n, V_n) = (S_N, V_N) + (S_{n-N}, V_{n-N}),$$

where

$$(S_N, V_N) \triangleq \sum_{j=1}^{N} n^{-\frac{1}{2}}(Z_j, Y_j, \sum_{\nu=0}^{l} \omega_\nu(Z_jZ_{j+\nu} - \mathbb{E}Z_jZ_{j+\nu}))$$

and (S_{n-N}, V_{n-N}) denotes the complementary part of the sums. Let $Q(S_n, V_n) \triangleq \mathbb{P}(S_n) - \frac{1}{2}T_nV_n$.

In the next step we modify an argument used in Lemma 3.43 of GH. We expand

$$Q(S_n, V_n) = Q(S_{n-N}, V_{n-N}) - \frac{1}{2}T_nV_N + \sum_{j=1}^{r} S_N^j \cdot Q_j(S_{n-N}, V_{n-N}),$$

where Q_j denotes an appropriate vector of polynomials and the dot indicates a multilinear form. Thus we have by expansion of exponential terms depending on

V_N and S_N in powers of $Z_j Z_{j+\nu}$ resp. Z_j and Y_j

$$
\begin{aligned}
|\mathbb{E}\exp[itE_n]| \quad \le \quad & |\mathbb{E}\exp[itT_n]\exp[itn^{-\frac{1}{2}}Q(S_{n-N},V_{n-N})] \\
& \times \sum_{\alpha,\beta}^{*} c_{\alpha,\beta} Z^{\alpha} Y^{\beta} Q_{\alpha\beta}(S_{n-N},V_{n-N})| \\
& + \frac{1}{p!}\mathbb{E}|Q(S_n,V_n) - Q(S_{n-N},V_{n-N})|^p (|t|n^{-\frac{1}{2}})^p
\end{aligned}
\tag{27}
$$

where the sum $\sum^*$ extends over all tuples

$$
\alpha \triangleq (\alpha_1,\ldots,\alpha_{N+l},0,\ldots,0), \quad \beta \triangleq (\beta_1,\beta_2,\ldots,\beta_N,0,\ldots,0)
$$

with $|\alpha| + |\beta| \le r(p-1)$, r denotes the degree of P and $Q_{\alpha,\beta}$ denote polynomials of the vector (S_{n-N}, V_{n-N}).

Let us estimate first the last remainder term, say I in (27). By (14), (15) and (25) we have

$$
\begin{aligned}
I &= O(l^{p/2}(|t|N^{\frac{1}{2}}/n)^p) \\
&= O((mn^{-\frac{1}{2}}l^{\frac{1}{2}})^p) \quad \text{for} \quad |t| \le \sqrt{n} \\
&= O((m|t|n^{-1}l^{\frac{1}{2}})^p) \quad \text{for} \quad \sqrt{n} < |t| < n^{1-\varepsilon}l^{-\frac{1}{2}}.
\end{aligned}
\tag{28}
$$

Thus $I = o(n^{-1})$ for $|t| \le n^{1-\varepsilon}l^{-1/2}$ provided that

$$
s\varepsilon > 2 \tag{29}
$$

In order to evaluate the expansion terms in (27) we may proceed as in the proof of Lemma 3.43 of GH, p. 233, bottom -235. Surrounding the at most $r(p-1)$ indices j where $\alpha_j > 0$ or $\beta_j > 0$ and the block $< N+1,\ldots,n >$ by an 'independence neighborhood' I of size $3m$ define

$$
\{j_1^0,\ldots,j_{r(p-1)}^0\} \triangleq \{j : \alpha_j > 0 \quad \text{or} \quad \beta_j > 0\}
$$

and

$$
I \triangleq \{j \in \{1,\ldots,N-m\} : |j - j_\nu^0| \ge 3m, \nu = 1,\ldots,r(p-1)\}.
$$

Divide I into blocks $A_1, B_1, \ldots, A_q, B_q$ as follows. Define $j_1,\ldots,j_q$

$$
\begin{aligned}
j_1 &\triangleq \inf I \\
j_{\nu+1} &= \inf\{j \ge j_\nu + 7m : j \in I\}
\end{aligned}
\tag{30}
$$

and let q denote the smallest integer for which the inf is undefined. Write

$$\begin{aligned} A_\nu &\stackrel{\triangle}{=} \prod\{\exp[itn^{-\frac{1}{2}}Z_j] : j \in I, |j - j_\nu| \le m\} \\ B_\nu &\stackrel{\triangle}{=} \prod\{\exp[itn^{-\frac{1}{2}}Z_j] : j \in I, j_\nu + m + 1 \le j \le j_{\nu+1} - m - 1\}, \end{aligned} \tag{31}$$

where $\nu = 1, \ldots, q-1$. Furthermore put

$$R \stackrel{\triangle}{=} Z^\alpha Y^\alpha \prod_{j \notin I} \exp[itZ_j n^{-\frac{1}{2}}] \exp[itQ(S_{n-N}, V_{n-N})].$$

Then

$$\exp[it(T_n + n^{-\frac{1}{2}}Q(S_{n-N}, V_{n-N}))] = \prod_1^q A_\nu B_\nu R,$$

where $|A_\nu| \le 1$, $|B_\nu| \le 1, |R| \le n^{\beta(p-1)r}$ and A_ν is $\mathcal{D}_{j_\nu - 2m, j_\nu + 2m}$ measurable, B_ν is $\mathcal{D}_{j_\nu+1, j_{\nu+1}-1}$ measurable and R is measurable with respect to

$$\{\mathcal{D}_l : \exists j \notin I \quad \text{with} \quad |l - j| \le m\}.$$

We have

$$\begin{aligned} & |\mathbb{E}R\prod_1^q A_\nu B_\nu - \mathbb{E}R\prod_1^q \mathbb{E}(A_\nu | \mathcal{D}_j : |j - j_\nu| \le 3m)B_\nu| \\ \le & \sum_{s=1}^q |\mathbb{E}R\prod_{\nu=1}^{s-1} A_\nu B_\nu (A_s - \mathbb{E}(A_s|\mathcal{D}_j : |j - j_s| \le 3m)) \\ & \times \prod_{\mu=s+1}^q \mathbb{E}(A_\mu|\mathcal{D}_j : |j - j_\mu| \le 3m)B_\mu|. \end{aligned} \tag{32}$$

Here we may replace A_s by $\mathbb{E}(A_s|\mathcal{D}_j : j \ne j_s)$ since the product of all remaining factors in (32) is measurable with respect to $\mathcal{D}_j : j \ne j_s$ (i. e. is constant with respect to this conditioning by our construction). Thus invoking condition (A6) we obtain that (32) can be bounded from above by $n^B d^{-1}\exp[-dm] = o(n^{-A})$ for arbitrary large $A > 0$ provided K in the definition of m is chosen large enough.

Recall that the functions $\mathbb{E}(A_\nu|\mathcal{D}_j : 0 < |j - j_\nu| \le 3m)$, $\nu = 1, \ldots, q$ are weakly dependent since $j_{\nu+1} - j_\nu \ge 7m, \nu = 1, \ldots, q-1$. Using condition (A4) and the bounds for R, A_ν and B_ν we obtain

$$\begin{aligned}
& |\mathbb{E}R\left[\prod_1^q B_\nu \mathbb{E}(A_\nu|\mathcal{D}_j : 0 < |j - j_\nu| \leq 3m)\right]| \\
\leq \;& n^{\beta(p-1)r}\mathbb{E}\prod_1^q |\mathbb{E}(A_\nu|\mathcal{D}_j : 0 < |j - j_\nu| \leq 3m)| \\
\leq \;& n^B \prod_1^q \mathbb{E}|\mathbb{E}(A_\nu|\mathcal{D}_j : 0 < |j - j_\nu| \leq 3m)| + 4n^B q d^{-1}\exp[-dm]
\end{aligned} \tag{33}$$

Condition (A5) implies that for $|t| \geq d$

$$\mathbb{E}|\mathbb{E}(A_\nu|\mathcal{D}_j : j \neq j_\nu)| \leq \exp[-d].$$

So by Lemma 3.2 of GH and assumption (A6)

$$\begin{aligned}
& \mathbb{E}|\mathbb{E}(A_\nu|\mathcal{D}_j : |j - j_\nu| \leq 3m)| \\
& \quad \leq \; \mathbb{E}|\mathbb{E}(A_\nu|\mathcal{D}_j : |j - j_\nu| \neq 0)| + O(n^B d^{-1}\exp[-dm]) \\
& \quad \leq \; \max(\exp[-dt^2n^{-1}], \exp[-d]) + O(n^B d^{-1}\exp[-dm]).
\end{aligned}$$

We finally obtain with $q \geq cN/m$, for some $c > 0$ using (25)

$$\begin{aligned}
|\mathbb{E}\exp[itE_n]| \;&\leq\; n^A\left(\max(\exp[-dt^2n^{-1}], \exp[-d])\right)^{N/m} + O(n^{-2}) \\
&=\; n^A\exp[-d'm] = O(n^{-2})
\end{aligned} \tag{34}$$

for some K sufficiently large and $d' > 0$, thus completing the proof.

3.2 Edgeworth expansions for the studentized Bootstrap

Assume that $(Z_j, Y_j) \in \mathbb{R}^{k+1}, j = 1, \ldots, n$ denotes a sequence of dependent r.v. satisfying condition (A1)-(A4). Let A_j, B_j, $N_1 \ldots N_b$ and $U^*_{n,stud}$ be as defined in Section 2. Then we have

Theorem 3.2 *Let $\psi_n(a)$ denote the Edgeworth-expansion defined in (7) and assume that $K\log(n) \leq l \leq n^{1/3}$ with K large enough and that conditions (A1)-(A4) hold with s replaced by qs, $q \geq 3$ and $s \geq 8$. Then*

$$\sup_a |\mathbb{P}(U^*_{n,stud} \leq a) - \psi_n(a)| = O_P(n^{-1+\varepsilon}l) + O(l^{-1}n^{-1/2})$$

for $\varepsilon \triangleq \frac{2}{s}$ and $l \leq n^{\frac{1}{3}}$.

Proof. We show (10) by using the same arguments as in Theorem 1 with $R_j, \quad j = 1, \ldots n$ replaced by $C_{N_j} \triangleq (B_{N_j}, A_{N_j}^T)^T, \quad j = 1, \ldots b$. So we have to check that the conditions (A1)-A(6) hold conditionally on $Y_1, \ldots Y_n$, uniformly for all $Y_1, \ldots Y_n$ in a set whose probability tends to one for n going to infinity. Condition (A1) holds by the definition of A_j and B_j. In order to check the moment condition (A2), consider

$$f_s \triangleq \mathbb{E}^*(||C_{N_1}||^s) = (n-l+1)^{-1}(||C_0||^s + \cdots + ||C_{n-l}||^s)$$

Assume that $E||(Z_j, Y_j)||^{qs} < \infty$ and let $f_s = l^{-1}\sum_{\nu=1}^{l} f_{s,\nu} + \mathbb{E}||C_1||^s$, where

$$f_{s,\nu} \triangleq b^{-1}\sum_{\mu=0}^{b-1}(||C_{\mu l+\nu}||^s - \mathbb{E}||C_1||^s).$$

By Theorem 2.11 in GH we have for any $x > 0$

$$\mathbb{P}(|f_{s,\nu} - \mathbb{E}f_{s,\nu}| > x) = O((l/n)^{(q-2)/2})$$

and therefore

$$\mathbb{P}(|f_s - \mathbb{E}f_s| > x) = O(l(l/n)^{(q-2)/2}) = o(1)$$

for $q \geq 3$.

The conditions (A3), (A4) and (A6) obviously hold true by independence when we choose $\mathcal{D}_j \triangleq \sigma(N_j), j = 1, 2, \ldots, b$. Hence it remains to check condition (A5) in this case. So we have to show that for some $0 < \zeta < 1/2$

$$\mathbb{P}(\sup_{d<|t|<b^{1/2}} |\mathbb{E}^* \exp[itB_{N_1}]| \leq 1 - \zeta) = 1 - o(n^{-1}). \tag{35}$$

By definition we have with $N \triangleq n - l + 1$

$$\begin{aligned} |\mathbb{E}^* \exp[itB_{N_1}]| &\leq |N^{-1}\sum_{j=1}^{N}(\exp[itB_j] - \mathbb{E}\exp[itB_j])| \\ &+ \sup_j |\mathbb{E}\exp[itB_j]| \\ &\triangleq I_1 + I_2, \quad \text{say.} \end{aligned} \tag{36}$$

By assumption (A4) and Lemma (3.33) GH with $\alpha = 0$ and n replaced by l we have for $|t| \leq l^{\varepsilon}$

$$I_2 \leq c_1 \exp[-c_2 t^2]$$

for some absolute constants $c_1, c_2 > 0$.

For large t we get from the proof of Lemma 3.43 GH with n replaced by l and $\alpha = 0$.

$$I_2 \leq \exp[-c_2 l^{\delta}]$$

for some absolute constants $c_2 > 0$ and $\delta > 0$ and every t with $l^\varepsilon < |t| < b^{\frac{1}{2}}$, provided that $l \gg m \triangleq K \log n$, K sufficiently large but independent of n. Hence we obtain

$$I_2 \leq 1 - 2\zeta \tag{37}$$

for every $d < |t| < b^{1/2}$ and some $0 < \zeta < 1/2$.

In order to estimate $I_1 = I_1(t)$ let $\xi_j \triangleq \exp[itB_j] - \mathbb{E}\exp[itB_j]$ denote a complex bounded random variable with mean zero. By condition (A3) we can approximate ξ_j by a mean zero random variable $\xi_j^\dagger$ which is $\mathcal{D}_{j-m,j+l+m}$ measurable such that

$$\mathbb{E}|\xi_j - \xi_j^\dagger| \leq |t| l^{1/2} d^{-1} \exp[-dm] \tag{38}$$

We then write

$$\begin{aligned} \mathbb{P}(|N^{-1}\sum_{j=1}^N \xi_j| > x) &\leq \mathbb{P}(|N^{-1}\sum_{j=1}^N \xi_j^\dagger| > \frac{x2}{)} + \mathbb{P}(|N^{-1}\sum_{j=1}^N (\xi_j^\dagger - \xi_j)| > \frac{x2}{)} \\ &\triangleq I_3 + I_4, \quad \text{say.} \end{aligned} \tag{39}$$

By (38) we have

$$\sup_{|t| \leq b^{1/2}} I_4 \leq N^{-1} \sum_{j=1}^N \mathbb{E}|\xi_j - \xi_j^\dagger|/x = O(n^{1/2}\exp[-dm]) = o(n^{-1})$$

for $m = K\log(n)$ and K sufficiently large. For I_3 we repeat the argument used to check (A2). If

$$(2l)^{-1} \sum_{\nu=1}^{2l} 2lN^{-1} \sum_{p=0}^{N/(2l)} \xi_{2pl+\nu}^\dagger > x/2,$$

then there exists a ν with

$$m_\nu \triangleq 2lN^{-1} \sum_{p=0}^{N/(2l)} \xi_{2pl+\nu}^\dagger > x/2.$$

By assumption (A4) we may assume that $\xi_{2pl+\nu}^\dagger, p = 0, 1, \ldots$ are independent up to an error of order $O(N\exp[-dm])$. Thus (39) implies

$$\mathbb{P}(I_1(t) > x) \leq (2l)\mathbb{P}(m_\nu > x/2) + O(N\exp[-dm]).$$

By standard exponential inequalities for bounded independent r.v. (in m_ν) we get

$$\mathbb{P}(I_1(t) > x) \leq O(l)\exp[-cx^2N/l] + O(N\exp[-dm]) \tag{40}$$

for some absolute constant $c > 0$.

We now have to prove (35) using (40). To this end split the intervals $\{t : d < |t| < b^{\frac{1}{2}}\}$ into n^C intervals of equal length with endpoints, say A_n. Then we have by standard arguments

$$\begin{aligned} \mathbb{P}(\sup_{d<|t|<b^{1/2}} I_1(t) > \zeta) &\leq \mathbb{P}(\sup_{t\in A_n} I_1(t) > \zeta/2, \\ &+ \mathbb{P}(\sup_{|t-s|\leq n^{-C}b^{1/2}} |I_1(t) - I_1(s)| > \zeta/2) \\ &= I_4 + I_5, \quad \text{say.} \end{aligned}$$

By (40) we have $I_4 = O(ln^C \exp[-c'N/l]) + o(n^{-1})$ for some absolute constant $c' > 0$ choosing $x \triangleq \zeta/2$.

In order to estimate I_5 note that we have for $\xi_j^{\dagger}(t)$

$$\mathbb{E} \sup_{|t-s|\leq\varepsilon} |\xi_j^{\dagger}(t) - \xi_j^{\dagger}(s)| \leq c\varepsilon$$

by (14) and therefore

$$I_5 \leq O(n^{-C}b^{1/2})$$

which finally yields the desired condition (35) by choosing $C > 0$ sufficiently large.

In order to complete the proof, we have to show (8). But this is straightforward because the behavior of the bootstrap moments follows from the argument to show (A2) used at the beginning of this proof.

4 A simulation example

In order to obtain an idea how our procedure works for a fixed sample size, we consider the following situation. The statistic of interest is the arithmetic mean of $X_1, \ldots, X_{120}$ where

$$X_j = (Y_{1,j}^2 + Y_{2,j}^2)/2$$

and $(Y_{1,j})$ and $(Y_{2,j})$ are two independent Gaussian processes with mean zero and

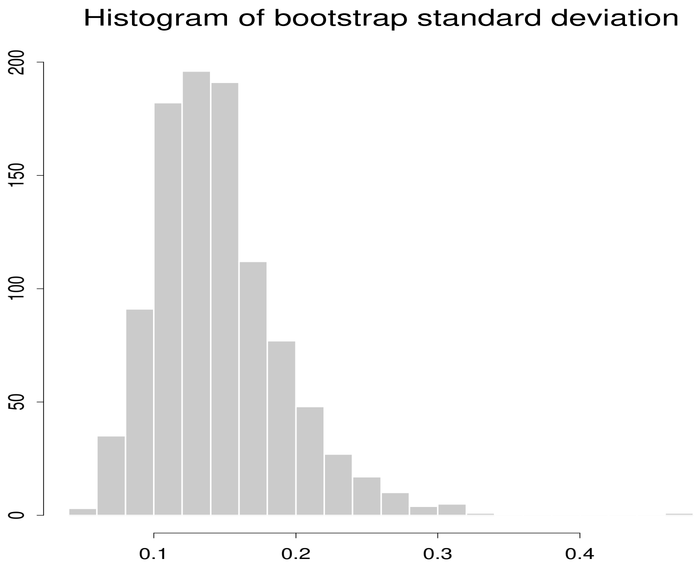
$$Cov(Y_{i,j}, Y_{i,j+h}) = 0.8^{|h|}.$$

Then X_j is $Exp(1)$-distributed and $Cov(X_j, X_{j+h}) = 0.8^{2|h|}$. It is easily seen that our Assumptions A(1)-A(6) are satisfied.

We study two statistics and two approximations for their distribution, namely the centered and the studentized mean and the normal and the bootstrap approximation respectively. According to the theory, the bootstrap should approximate the skewness of both statistics well. Of course the normal approximation cannot deal with skewness. On the other hand for the centered mean we expect a noticeable error in the variance for both approximations whereas with studentization all variances should be one.

The true distribution of both statistics was obtained from 1000 samples of the series $X_1, \ldots, X_{120}$. For the block size we chose $l = 1, 3, 6$ ($l = 1$ being the standard bootstrap of Efron, 1979) and as lag weights we chose $w_k = 2 (1 < k < l)$. As was to be expected, $l = 1$ gave totally wrong answers. Among the other two block sizes, $l = 6$ gave better results, so we only present this case. For each of the 1000 samples we calculated the bootstrap standard deviation of the mean exactly. Their mean is 0.146 whereas the true standard deviation of the mean is 0.191. So the bootstrap standard deviation is biased due to the independence of the blocks. The distribution of the bootstrap standard deviation is shown in Fig.1.

Figure 1: Histogram of bootstrap standard deviations of the mean in 1000 samples.



For each of the first 10 samples we also computed the bootstrap distribution of the two statistics by generating 1000 bootstrap samples. The values of the bootstrap standard deviation for these first 10 samples are given in Tab.1. Comparing with the histogram of Fig.1, we see that the first 10 samples are a reasonable cross section of all samples.

Table 1: Bootstrap standard deviations of the mean for the first 10 samples.

sample number	1	2	3	4	5	6	7	8	9	10
bootstrap s.d.	.145	.067	.153	.147	.213	.169	.161	.160	.128	.160

Figure 2: Quantile-quantile plots of true versus normal and bootstrap approximation to the distributions of arithmetic and studentized mean using the first sample.

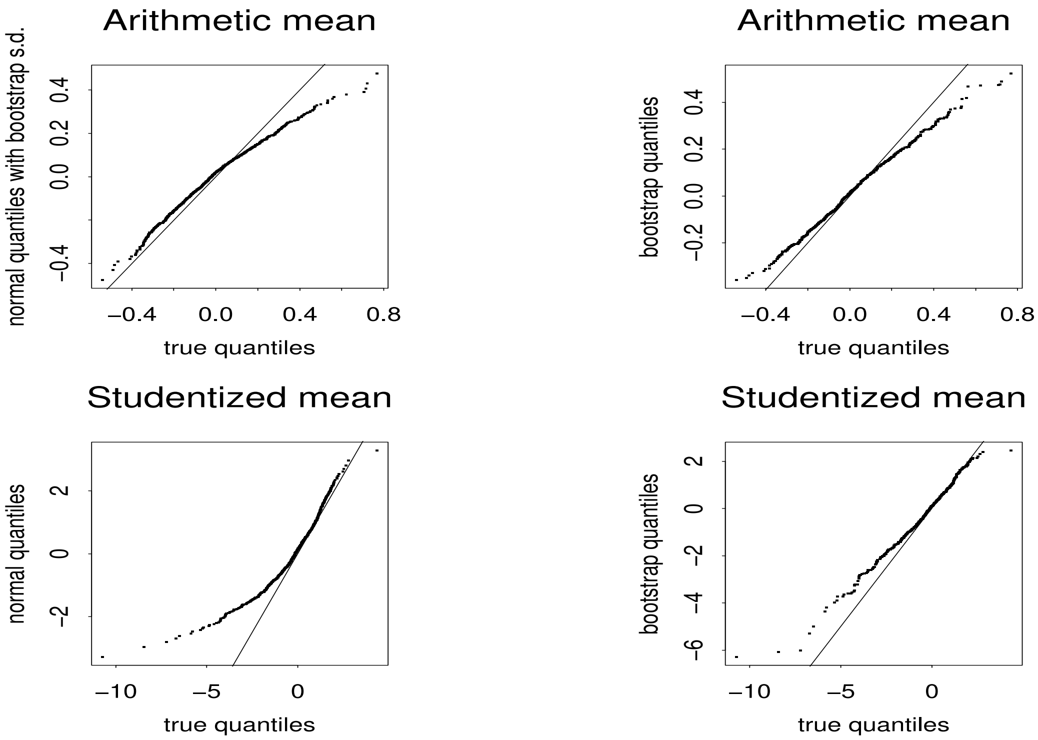


Fig.2 shows the quantile-quantile plot for all four combinations of statistic and approximation. We also show the line $y = x$ where all points would lie if the approximation were perfect. The results agree quite well with what we expect from theory except that the bootstrap approximation for the studentized mean is slightly in error at the lower tails. For other samples the normal approximation is unchanged for the studentized mean and for the centered mean only the slope changes. So we only show the bootstrap approximations for the other 9 samples. For the studentized mean the variability between samples is higher than for the centerd mean and the lower tail is often problematic whereas the approximation is quite good in the upper tails. In all cases the bootstrap approximation is better than the normal approximation for the same statistic. So at least in this example the bootstrap offers some improvement over the normal approximation.

References

Bhattacharya, R.N., Ranga Rao, R. (1976). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions.

Ann. Statist. **16**, 1709-1722.

Bulinskii, A.V. and Zhurbenko, I.G. (1976). The central limit theorem for random fields. Soviet Math. Dokl. **17**, 14-17.

Chibisov, D.M. (1972). An asymptotic expansion for the distribution of a statistic admitting an asymptotic expansion. Th. Probab. Appl. **17**, 620-630.

Davison, A.C. and Hall, P. (1993). On studentizing and blocking methods for implementing the bootstrap with dependent data. Austral. J. Statist. **35**, to appear.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. Ann. Statist. **7**, 1-26.

Götze, F. and Hipp, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors. Z. Wahrsch. verw. Geb. **64**, 211-239.

Götze, F. and Hipp, C. (1992). On the validity of Edgeworth-expansions. Preprint 92-045, SFB 343, Bielefeld.

Götze, F. and Künsch, H. R. (1990). Blockwise bootstrap for dependent observations: Higher order approximations for studentized statistics (Abstract). Bull. Inst. Math. Statist. **19**, 443.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals (with discussion). Ann. Statist. **16**, 927-985.

Jensen, J.L. (1991). A note on asymptotic expansions for sums over a weakly dependent random field. Research Report No, 232. Department of theoretical Statistics, Aarhus University.

Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. Ann. Statist. **17**, 1217-1241.

Lahiri, S.N. (1991). Second order optimality of stationary bootstrap. Statist. Probab. Let. **11**, 335-341.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. Ann. Statist. **9**, 1187-1195.

Tikhomirov, A.N. (1980). On the convergence rate of in the central limit theorem for weakly dependent random variables. Th. Probab. Appl. **25**, 790-809.

Zhurbenko, I.G. (1972). On strong estimates of mixed semiinvariants of random processes. Siberian Math. J. **13**, 202-213.

Quantile-quantile plots of true versus bootstrap approximation to the distributions of arithmetic and studentized mean using samples 2 - 10.

Figure 3: Samples 3 and 4

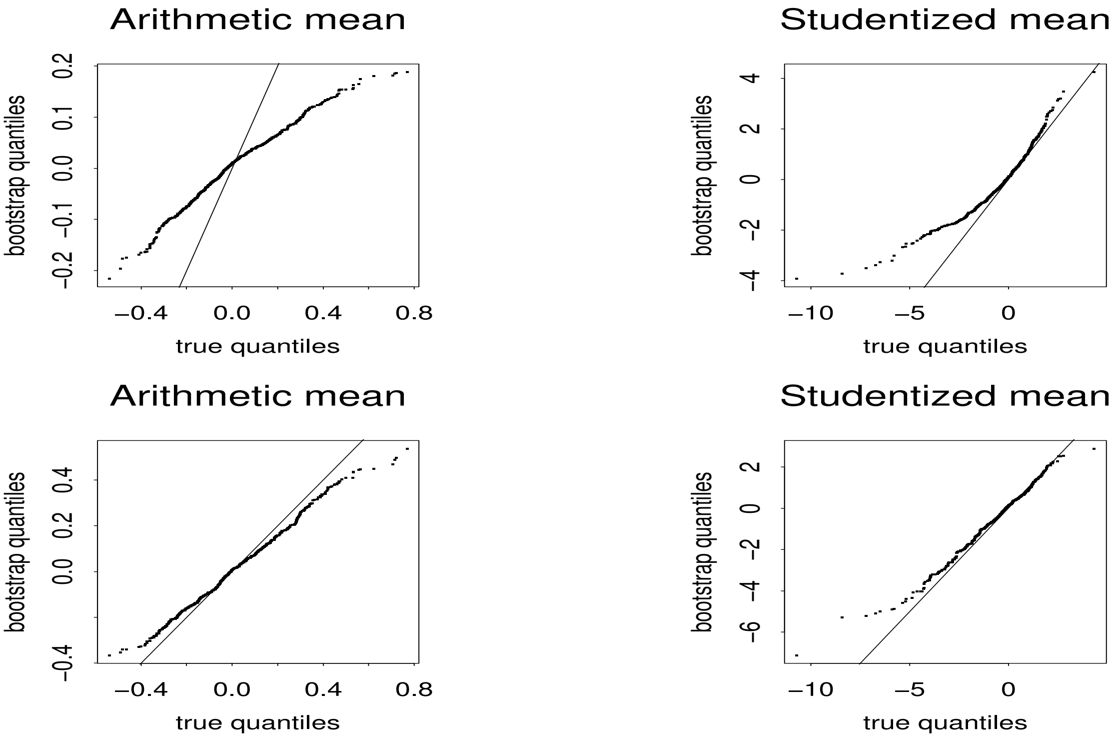


Figure 4: Samples 5 and 6

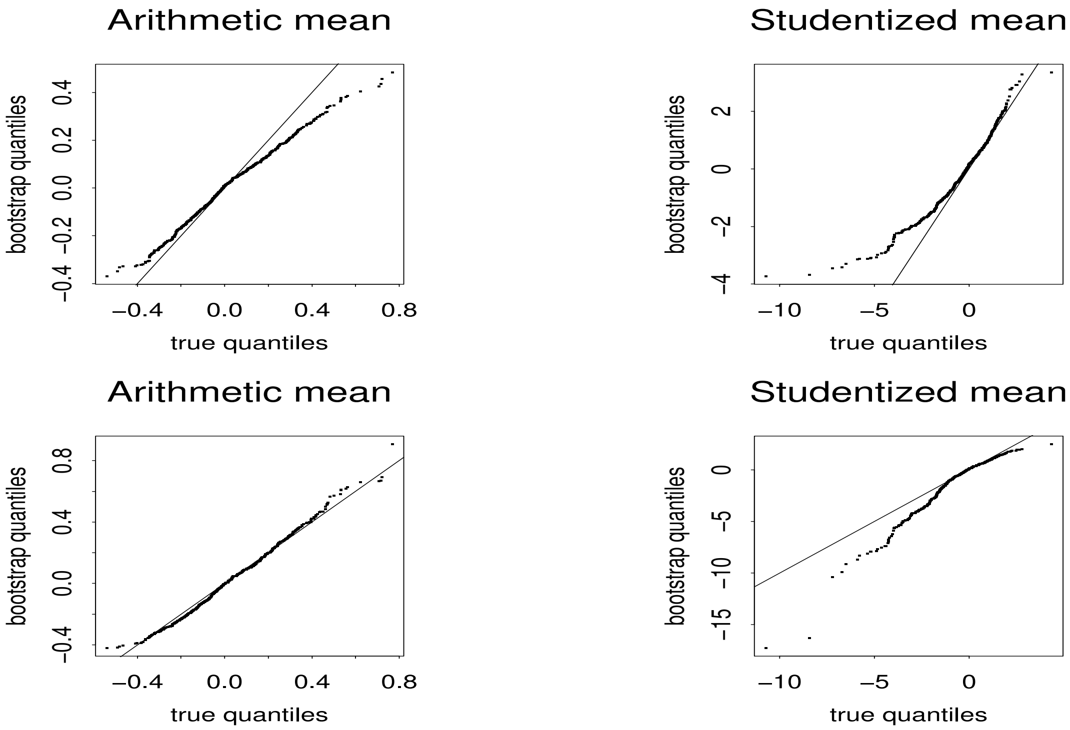


Figure 5: Samples 7 and 8

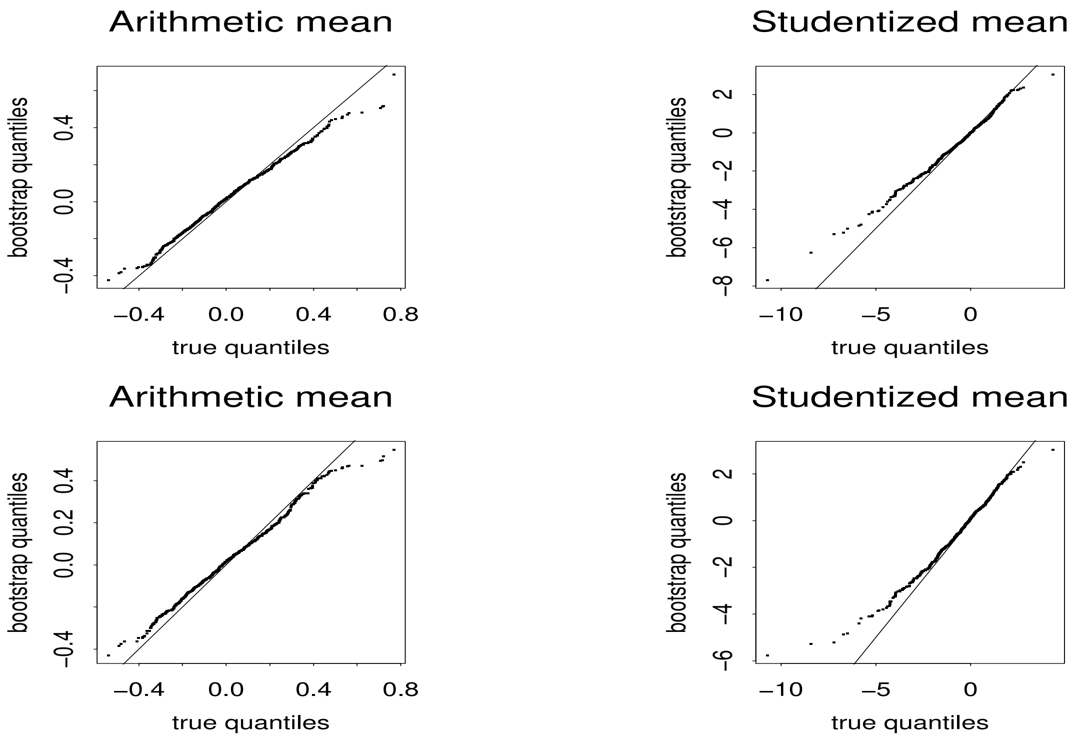


Figure 6: Samples 9 and 10

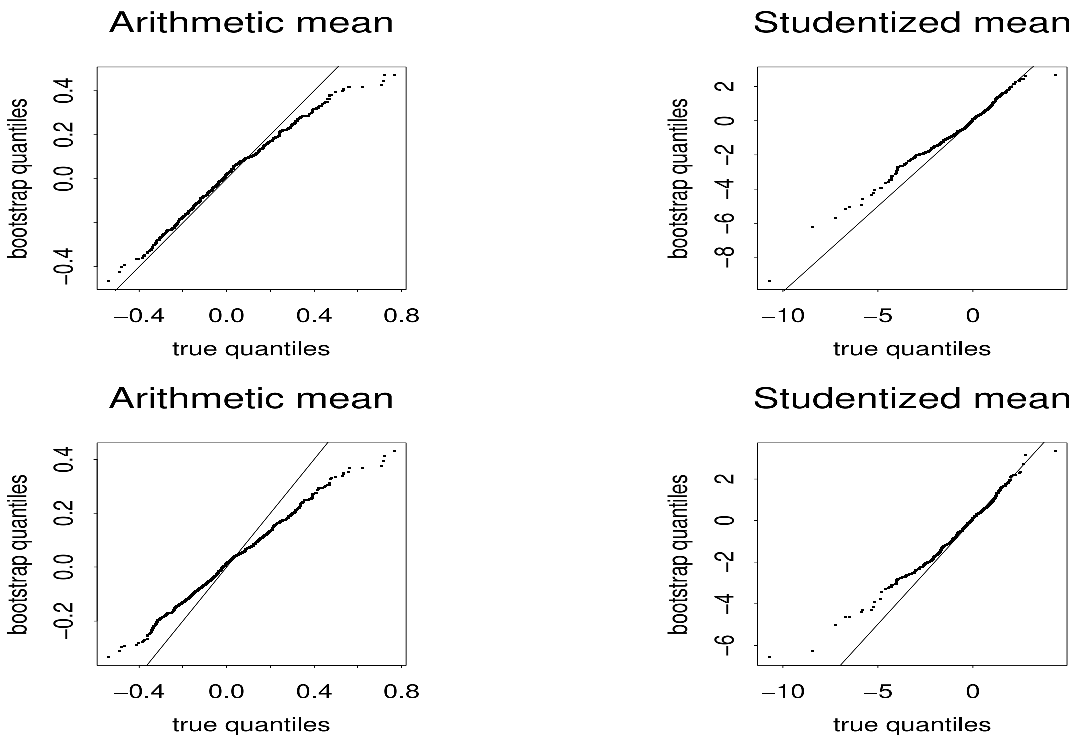


Figure 7: Sample 11