Nonparametric estimation of the distribution of the autoregressive coefficient from panel random-coefficient AR(1) data
Remigijus Leipus, Anne Philippe, Vytautė Pilipauskaitė, Donatas Surgailis

To cite this version:
Remigijus Leipus, Anne Philippe, Vytautė Pilipauskaitė, Donatas Surgailis. Nonparametric estimation of the distribution of the autoregressive coefficient from panel random-coefficient AR(1) data. Journal of Multivariate Analysis, Elsevier, 2017, 153, pp.121 - 135. 10.1016/j.jmva.2016.09.007. hal-01174833

HAL Id: hal-01174833
https://hal.archives-ouvertes.fr/hal-01174833
Submitted on 9 Jul 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright
Nonparametric estimation of the distribution of the autoregressive coefficient from panel random-coefficient AR(1) data∗

Remigijus Leipus1, Anne Philippe2, Vytautė Pilipauskaitė1,2, Donatas Surgailis1

July 9, 2015

1Vilnius University and 2Université de Nantes

Abstract

We discuss nonparametric estimation of the distribution function \( G(x) \) of the autoregressive coefficient from a panel of \( N \) random-coefficient AR(1) data, each of length \( n \), by the empirical distribution of lag 1 sample correlations of individual AR(1) processes. Consistency and asymptotic normality of the empirical distribution function and a class of kernel density estimators is established under some regularity conditions on \( G(x) \) as \( N \) and \( n \) increase to infinity. A simulation study for goodness-of-fit testing compares the finite-sample performance of our nonparametric estimator to the performance of its parametric analogue discussed in Beran et al. (2010).

Keywords: random-coefficient autoregression, empirical process, Kolmogorov-Smirnov statistic, kernel density estimator.

2010 MSC: 62G10, 62M10, 62G07.

1 Introduction

Panel data can describe a huge population of heterogeneous units/agents which evolve over the time, e.g. households, firms, industries, countries, stock market indices. In this paper we consider a panel where each individual unit evolves according to order-one random coefficient autoregressive model (RC AR(1)). It is well known that aggregation of specific RC AR(1) models can explain long memory phenomenon, which is often empirically observed in economic time series (see Granger (1980) for instance). More precisely, consider a panel \( \{X_i(t), t = 1, \ldots, n, i = 1, \ldots, N\} \), where \( X_i = \{X_i(t), t \in \mathbb{Z}\} \) is RC AR(1) process with the \((0, \sigma^2)\)-noise and the random coefficient \( a_i \in (-1, 1) \), whose autocovariance

\[
\text{EX}_i(0)X_i(t) = \sigma^2 \int_{-1}^{1} \frac{x^{|t|}}{1-x^2} \, dG(x)
\]

(1.1)
is determined by the distribution \( G(x) = P(a \leq x) \) of the autoregressive coefficient. Granger (1980) showed, for a specific Beta-type distribution \( G(x) \), that the contemporaneous aggregation of independent processes \( \{X_i(t), i = 1, \ldots, N\} \), results in a stationary Gaussian long memory process \( \{X(t)\} \), i.e.

\[
N^{-1/2} \sum_{i=1}^{N} X_i(t) \to_{fdd} \mathcal{X}(t) \quad \text{as } N \to \infty,
\]

∗ The first, third and fourth authors are supported by a grant (No. MIP-063/2013) from the Research Council of Lithuania.
where the autocovariance \( E \mathcal{X}(0) \mathcal{X}(t) = EX_1(0)X_1(t) \) decays slowly as \( t \to \infty \) so that \( \sum_{t \in \mathbb{Z}} |E \mathcal{X}(0) \mathcal{X}(t)| = \infty \).

A natural statistical problem is recovering the correlation function in (1.1) or the distribution \( G(x) \) (the frequency of \( a \) across the population of individual AR(1) ‘microagents’) from the aggregated sample \( \{ \mathcal{X}(t), t = 1, \ldots, n \} \). This problem was treated in Leipus et al. (2006), Celov et al. (2010), Chong (2006). Some related results were obtained in Celov et al. (2007), Horváth and Leipus (2009), and Jirak (2013). Albeit nonparametric, the estimators in Leipus et al. (2006) and Celov et al. (2010) involve an expansion of the density \( g = G' \) in an orthogonal polynomial basis and are sensitive to the choice of the tuning parameter (the number of polynomials), being limited in practice to very smooth densities \( g \). The last difficulty in estimation of \( G \) from aggregated data is not surprising due to the fact that aggregation per se inflicts a considerable loss of information about the evolution of individual ‘micro-agents’.

Clearly, if the available data comprises evolutions \( \{ X_i(t), t = 1, \ldots, n \}, i = 1, \ldots, N \), of all \( N \) individual ‘micro-agents’ (the panel data), we may expect a much more accurate estimate of \( G \). Robinson (1978) constructed an estimator for the moments of \( G \) using sample autocovariances of \( X_i \) and derived its asymptotic properties as \( N \to \infty \), whereas the length \( n \) of each sample remains fixed. Beran et al. (2010) discussed estimation of two-parameter Beta densities \( g \) from panel AR(1) data using maximum likelihood estimators with unobservable \( a_i \) replaced by sample lag 1 autocorrelation of \( X_i(1), \ldots, X_i(n) \) (see Sec. 5), and derived the asymptotic normality and some other properties of the estimators as \( N \) and \( n \) tend to infinity.

The present paper studies nonparametric estimation of \( G \) from panel random-coefficient AR(1) data using the empirical distribution function:

\[
\hat{G}_{N,n}(x) := \frac{1}{N} \sum_{i=1}^{N} 1(\hat{a}_{i,n} \leq x), \quad x \in \mathbb{R},
\]

(1.2)

where \( \hat{a}_{i,n} \) is the lag 1 sample autocorrelation of \( X_i, i = 1, \ldots, N \) (see (3.3)). We also discuss kernel estimation of the density \( g(x) = G'(x) \) based on smoothed version of (1.2). We assume that individual AR(1) processes \( X_i \) are driven by identically distributed shocks containing both common and idiosyncratic (independent) components. Consistency and asymptotic normality as \( N, n \to \infty \) of the above estimators are derived under some regularity conditions on \( G(x) \). Our results can be applied to test goodness-of-fit of the distribution \( G(x) \) to a given hypothesized distribution (e.g., a Beta distribution) using the Kolmogorov-Smirnov statistic, and to construct confidence intervals for \( G(x) \) or \( g(x) \).

The paper is organized as follows. Section 2 obtains the rate of convergence of the sample autocorrelation coefficient \( \hat{a}_{i,n} \) to \( a_i \), in probability, the result of independent interest. Section 3 discusses the weak convergence of the empirical process in (1.2) to a generalized Brownian bridge. In Section 4 we study kernel density estimators of \( g(x) \). We show that these estimates are asymptotically normally distributed and their mean integrated square error tends to zero. A simulation study of Section 5 compares the empirical performance of (1.2) and the parametric estimator of Beran et al. (2010) to the goodness-of-fit testing for \( G(x) \) under null Beta distribution. The proofs of auxiliary statements can be found in the Appendix.

In what follows, \( C \) stands for a positive constant whose precise value is unimportant and which may change from line to line. We write \( \to_p, \to_d, \to_{fdd} \) for the convergence in probability and the convergence of (finite-dimensional) distributions respectively, whereas \( \Rightarrow \) denotes the weak convergence in the space \( D[-1,1] \) with the supremum metric.
2 Estimation of random autoregressive coefficient

Consider a random-coefficient AR(1) process

\[ X(t) = aX(t-1) + \zeta(t), \quad t \in \mathbb{Z}, \]  

(2.1)

where innovations \{\zeta(t)\} admit the following decomposition:

\[ \zeta(t) = b\eta(t) + c\xi(t), \quad t \in \mathbb{Z}, \]  

(2.2)

where random sequences \{\eta(t)\}, \{\xi(t)\} and random coefficients \(a, b, c\) satisfy the following conditions:

**Assumption A_1** \{\eta(t)\} are independent identically distributed (i.i.d.) random variables (r.v.s) with \(E\eta(0) = 0, E\eta^2(0) = 1, E|\eta(0)|^p < \infty\) for some \(p > 1\).

**Assumption A_2** \{\xi(t)\} are i.i.d. r.v.s with \(E\xi(0) = 0, E\xi^2(0) = 1, E|\xi(0)|^p < \infty\) for the same \(p\) as in A_1.

**Assumption A_3** \(b, c\) are possibly dependent r.v.s such that \(P(b^2 + c^2 > 0) = 1\) and \(Eb^2 < \infty, Ec^2 < \infty\).

**Assumption A_4** \(a \in (-1, 1)\) is a r.v. with a distribution function (d.f.) \(G(x) := P(a \leq x)\) supported on \([-1, 1]\) and satisfying

\[ E\left[ \frac{1}{1 - |a|} \right] = \int_{-1}^{1} \frac{dG(x)}{1 - |x|} < \infty. \]  

(2.3)

**Assumption A_5** \(a, \{\eta(t)\}, \{\xi(t)\}\) and the vector \((b, c)\)' are mutually independent.

**Remark 2.1** In the context of panel observations (see (3.1) below), \{\eta(t)\} is the common component and \{\xi(t)\} is the idiosyncratic component of shocks. The innovation process \{\zeta(t)\} in (2.2) is i.i.d. if the coefficients \(b, c\) are nonrandom. In the general case \{\zeta(t)\} is a dependent and uncorrelated stationary process with \(E\zeta(0) = 0, E\zeta^2(0) = Eb^2 + Ec^2, E\zeta(0)\zeta(t) = 0, t \neq 0\).

Under conditions A_1–A_5, a unique strictly stationary solution of (2.1) with finite variance exists and is written as

\[ X(t) = \sum_{s \leq t} a^{t-s} \zeta(s), \quad t \in \mathbb{Z}. \]  

(2.4)

Clearly, \(EX(t) = 0\) and \(EX^2(t) = E\zeta^2(0)E(1 - a^2)^{-1} < \infty\). Note that (2.3) is equivalent to

\[ E\left[ \frac{1}{1 - |a|^p} \right] < \infty, \quad 1 < p \leq 2, \]

since \(1 - |a| \leq 1 - |a|^p \leq 2(1 - |a|)\) for \(a \in (-1, 1)\).

For an observed sample \(X(1), \ldots, X(n)\) from the stationary process in (2.4), define the sample lag 1 autocorrelation coefficient

\[ \hat{a}_n := \frac{\sum_{t=1}^{n-1} X(t)X(t+1)}{\sum_{t=1}^{n} X^2(t)}. \]  

(2.5)

Note \(|\hat{a}_n| \leq 1\) a.s. by the Cauchy inequality.
Proposition 2.1 \textit{Under assumptions $A_1$–$A_5$, for any $0 < \gamma < 1$ and $n \geq 1$, it holds}

$$
P(|\hat{a}_n - a| > \gamma) \leq C(n^{-(p/2)\wedge(p-1)}\gamma^{-p} + n^{-1}),
$$

\textit{with $C > 0$ independent of $n, \gamma$.}

\textbf{Proof.} See Appendix.

Assume now that the d.f. $G(x) = P(a \leq x)$ satisfies the following Hölder condition:

\textbf{Assumption A}_6 \textit{There exist constants $L_G > 0$ and $\rho \in (0, 1]$ such that}

$$
|G(x) - G(y)| \leq L_G|x - y|^{\rho}, \quad x,y \in [-1,1].
$$

Consider the d.f. of $\hat{a}_n$:

$$
G_n(x) := P(\hat{a}_n \leq x), \quad x \in \mathbb{R}.
$$

Proposition 2.2 \textit{Let assumptions $A_1$–$A_6$ hold. Then, for any $0 < \gamma < 1$ and $n \geq 1$,}

$$
\sup_{x \in [-1,1]} |G_n(x) - G(x)| \leq L_G\gamma^{\rho} + C(n^{-1} + n^{-(p/2)\wedge(p-1)}\gamma^{-p})
$$

\textit{with $C > 0$ independent of $n, \gamma$.}

\textbf{Proof.} Denote $\delta_n := \hat{a}_n - a$. For any (nonrandom) $\gamma > 0$ we have

$$
\sup_{x \in [-1,1]} |G_n(x) - G(x)| = \sup_{x \in [-1,1]} |P(a + \delta_n \leq x) - P(a \leq x)| \leq L_G\gamma^{\rho} + P(|\delta_n| > \gamma).
$$

To see this, note that for any $\gamma > 0$ and any r.v. $\delta$ (possibly dependent on $a$)

$$
\sup_{x \in \mathbb{R}} |P(a + \delta \leq x) - P(a \leq x)| \leq L_G\gamma^{\rho} + P(|\delta| > \gamma).
$$

Indeed, using (2.7), which actually holds for all $x, y \in \mathbb{R}$, we have for any $x \in \mathbb{R}$ that

$$
P(a + \delta \leq x) - P(a \leq x) = P(a + \delta \leq x, |\delta| \leq \gamma) - P(a \leq x) + P(a + \delta \leq x, |\delta| > \gamma)
\leq P(a \leq x + \gamma) - P(a \leq x) + P(|\delta| > \gamma)
\leq L_G\gamma^{\rho} + P(|\delta| > \gamma)
$$

by (2.7), and similarly $P(a \leq x) - P(a + \delta \leq x) \leq L_G\gamma^{\rho} + P(|\delta| > \gamma)$, which proves (2.10). It remains to apply Proposition 2.1. \qed

Corollary 2.3 \textit{Let assumptions $A_1$–$A_6$ hold. Then, as $n \to \infty$,}

$$
\sup_{x \in [-1,1]} |G_n(x) - G(x)| = O(n^{-\frac{\rho}{\rho+p}((\frac{p}{2}) \wedge (p-1))}).
$$

\textbf{Proof follows from Proposition 2.2 by taking $\gamma = \gamma_n = o(1)$ such that $\gamma_n^{\rho} \sim n^{-(p/2)\wedge(p-1)}\gamma_n^{-p}$ and noting that the exponent $\frac{\rho}{\rho+p}((\frac{p}{2}) \wedge (p-1)) < 1$.} \qed
3 Asymptotics of the empirical distribution function

Consider random-coefficient AR(1) processes \( \{X_i(t)\} \), \( i = 1, 2, \ldots \), which are stationary solutions to

\[
X_i(t) = a_i X_i(t-1) + \zeta_i(t), \quad t \in \mathbb{Z},
\]

(3.1)

with innovations \( \{\zeta_i(t)\} \) having the same structure as in (2.2):

\[
\zeta_i(t) = b_i \eta(t) + c_i \xi_i(t), \quad t \in \mathbb{Z}.
\]

(3.2)

More precisely, we make the following assumption:

**Assumption B** \( \{\eta(t)\} \) satisfies A1; \( \{\xi_i(t)\} \), \( (b_i, c_i)' \), \( a_i \), \( i = 1, 2, \ldots \), are independent copies of \( \{\xi(t)\} \), \( (b, c)' \), \( a \) which satisfy assumptions A2–A6. (Note that we assume A5 for any \( i = 1, 2, \ldots \))

Define the corresponding sample correlation coefficients

\[
\widehat{a}_{i,n} := \frac{\sum_{t=1}^{n} X_i(t) X_i(t+1)}{\sum_{t=1}^{n} X_i^2(t)}
\]

(3.3)

and the empirical d.f.

\[
\widehat{G}_{N,n}(x) := \frac{1}{N} \sum_{i=1}^{N} 1(\widehat{a}_{i,n} \leq x), \quad x \in \mathbb{R}.
\]

(3.4)

Recall that (3.4) is a nonparametric estimate of the d.f. \( G(x) = P(a_i \leq x) \) from observed panel data \( \{X_i(t), t = 1, \ldots, n, i = 1, \ldots, N\} \). In the following theorem we show that \( \widehat{G}_{N,n}(x) \) is an asymptotically unbiased estimator of \( G(x) \), as \( n \) and \( N \) both tend to infinity, and prove the weak convergence of the corresponding empirical process.

**Theorem 3.1** Assume the panel data model in (3.1)–(3.2). Let Assumption B hold and \( N, n \rightarrow \infty \). Then

\[
\sup_{x \in [-1,1]} |E[\widehat{G}_{N,n}(x) - G(x)]| = O(n^{-\frac{2}{p} + \frac{1}{2p^2}}).
\]

(3.5)

If, in addition, \( N = o(n^{\frac{2}{p} + \frac{1}{2p^2}}) \), then

\[
N^{1/2}(\widehat{G}_{N,n}(x) - G(x)) \Rightarrow W(x), \quad x \in [-1,1],
\]

(3.6)

where \( \{W(x), x \in [-1,1]\} \) is a continuous Gaussian process with zero mean and covariance

\[
\text{Cov}(W(x), W(y)) = G(x \wedge y) - G(x)G(y), \quad x, y \in [-1,1], \text{ and } W(-1) = W(1) = 0.
\]

**Proof.** Note \( \widehat{a}_{i,n}, i = 1, \ldots, N, \) are identically distributed, in particular, \( E\widehat{G}_{N,n}(x) = G_n(x) \) with \( G_n(x) \) defined in (2.8). Hence, (3.5) follows immediately from Corollary 2.3.

To prove the second statement of the theorem, we approximate \( \widehat{G}_{N,n}(x) \) by the empirical d.f.

\[
\widehat{G}_N(x) := \frac{1}{N} \sum_{i=1}^{N} 1(a_i \leq x), \quad x \in [-1,1]
\]

of i.i.d. r.v.s \( a_i, i = 1, \ldots, N \). We have \( N^{1/2}(\widehat{G}_{N,n}(x) - G(x)) = N^{1/2}(\widehat{G}_N(x) - G(x)) + D_{N,n}(x) \) with

\[
D_{N,n}(x) := N^{1/2}(\widehat{G}_{N,n}(x) - \widehat{G}_N(x)).
\]

Since A6 guarantees the continuity of \( G \), it holds

\[
N^{1/2}(\widehat{G}_N(x) - G(x)) \Rightarrow W(x), \quad x \in [-1,1]
\]
by the classical Donsker’s theorem, and (3.6) follows from \( \sup_{x \in [-1,1]} |D_{N,n}(x)| \to_p 0 \). By definition,

\[
D_{N,n}(x) = N^{-1/2} \sum_{i=1}^{N} \left( 1(a_i + \delta_{i,n} \leq x) - 1(a_i \leq x) \right) = D'_{N,n}(x) - D''_{N,n}(x),
\]

where \( \delta_{i,n} := a_{i,n} - a_i, i = 1, \ldots, N, \) and

\[
D'_{N,n}(x) := N^{-1/2} \sum_{i=1}^{N} 1(x < a_i \leq x - \delta_{i,n}, \delta_{i,n} \leq 0),
\]

\[
D''_{N,n}(x) := N^{-1/2} \sum_{i=1}^{N} 1(x - \delta_{i,n} < a_i \leq x, \delta_{i,n} > 0).
\]

For \( \gamma > 0 \) we have

\[
D'_{N,n}(x) \leq N^{-1/2} \sum_{i=1}^{N} 1(x < a_i \leq x + \gamma) + N^{-1/2} \sum_{i=1}^{N} 1(|\delta_{i,n}| > \gamma) =: V'_{N}(x) + V''_{N,n}.
\]

(Note that \( V''_{N,n} \) does not depend on \( x \).) By Proposition 2.1, we obtain

\[
EV''_{N,n} = N^{-1/2} \sum_{i=1}^{N} P(|\delta_{i,n}| > \gamma) \leq CN^{1/2}(n^{-(p/2)\wedge(p-1)}\gamma^{-p} + n^{-1}),
\]

which tends to 0 when \( \gamma \) is chosen as \( \gamma^{\theta+p} = n^{-(p/2)\wedge(p-1)} \to 0 \). Next,

\[
V'_{N}(x) = N^{1/2}(\hat{G}_{N}(x + \gamma) - \hat{G}_{N}(x)) = N^{1/2}(G(x + \gamma) - G(x)) + U_{N}(x, x + \gamma),
\]

\[
U_{N}(x, x + \gamma) := N^{1/2}(\hat{G}_{N}(x + \gamma) - G(x + \gamma)) - N^{1/2}(\hat{G}_{N}(x) - G(x)).
\]

The above choice of \( \gamma^{\theta+p} = n^{-(p/2)\wedge(p-1)} \) implies \( \sup_{x \in [-1,1]} N^{1/2}|G(x + \gamma) - G(x)| = O(N^{1/2}\gamma^\theta) = o(1) \), whereas \( U_{N}(x, x + \gamma) \) vanishes in the uniform metric in probability (see Lemma 6.2 in Appendix). Since \( D''_{N,n}(x) \) is analogous to \( D'_{N,n}(x) \), this proves the theorem. \( \square \)

**Remark 3.1** Theorem 3.1 can be used for testing goodness-of-fit, i.e. the null hypothesis \( H_{0} : G = G_{0} \) vs. \( H_{1} : G \neq G_{0} \) with \( G_{0} \) being a certain hypothetical distribution satisfying the Hölder condition in (2.7). Accordingly, the corresponding Kolmogorov-Smirnov test rejecting \( H_{0} \) whenever

\[
N^{1/2} \sup_{x \in [-1,1]} |\hat{G}_{N,n}(x) - G_{0}(x)| > c_{\alpha}
\]

has asymptotic size \( \alpha \in (0, 1) \) provided \( N, n, G_{0} \) satisfy the assumptions for (3.6) in Theorem 3.1. (Here, \( c_{\alpha} \) is the upper \( \alpha \)-quantile of the Kolmogorov distribution.) See also Section 5. One can also consider other statistics that are continuous functionals of the empirical process \( \{N^{1/2}(\hat{G}_{N,n}(x) - G_{0}(x)), x \in [1,1]\} \).

### 4 Kernel density estimation

In this section we assume \( G \) has a bounded probability density function \( g(x) = G'(x), x \in [-1,1] \), implying assumption \( A_{6} \) with Hölder exponent \( \varrho = 1 \) in (2.7). It is of our interest to estimate \( g(x) \) in a nonparametric way from \( \tilde{a}_{i,n}, \ldots, \tilde{a}_{N,n} \) (3.3).
Consider the kernel density estimator
\[
\hat{g}_{N,n}(x) := \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x - \tilde{a}_{i,n}}{h}\right), \quad x \in \mathbb{R},
\] (4.1)
where \(K\) is a kernel, satisfying Assumption A7 and \(h = h_{N,n}\) is a bandwidth which tends to zero as \(N\) and \(n\) tend to infinity.

**Assumption A7** \(K : [-1, 1] \to \mathbb{R}\) is a continuous function of bounded variation that satisfies \(\int_{-1}^{1} K(x)dx = 1\). Set \(\|K\|_{2}^2 := \int_{-1}^{1} K(y)^2dy\) and \(\mu_{2}(K) := \int_{-1}^{1} y^2K(y)dy\) and \(K(x) := 0, x \in \mathbb{R} \setminus [-1, 1]\).

We consider two cases separately.

**Case (i)** \(P(b_1 = 0) = 1\), meaning that the coefficient \(b_i = 0\) for the common shock in (3.2) is zero and that the individual processes \(\{X_i(t)\}, i = 1, 2, \ldots\), are independent and satisfy
\[
X_i(t) = a_iX_i(t-1) + c_i\xi_i(t), \quad t \in \mathbb{Z}.
\]

**Case (ii)** \(P(b_1 \neq 0) > 0\), meaning that \(\{X_i(t)\}, i = 1, 2, \ldots\), are mutually dependent processes.

**Proposition 4.1** Let assumptions B and A7 hold. If \(h^{1+p_{n}(p/2)^{(p-1)}} \to \infty\), then
\[
E\hat{g}_{N,n}(x) \to g(x)
\] (4.2)
at every continuity point \(x \in \mathbb{R}\) of \(g\). Moreover, if
\[
\begin{cases}
n^{(p/2)^{(p-1)}}h^{1+p} \to \infty & \text{in Case (i)}, \\
n^{(p/2)^{(p-1)}}(h/N)^{1+p} \to \infty & \text{in Case (ii)},
\end{cases}
\] (4.3)
then
\[
Nh\text{Cov}(\hat{g}_{N,n}(x_1), \hat{g}_{N,n}(x_2)) \to \begin{cases} g(x_1)\|K\|_{2}^2 & \text{if } x_1 = x_2, \\
0 & \text{if } x_1 \neq x_2 \end{cases}
\] (4.4)
at any continuity points \(x_1, x_2 \in \mathbb{R}\) of \(g\). If \(Nh \to \infty\) holds in addition to (4.3), then the estimator \(\hat{g}_{N,n}(x)\) is consistent at each continuity point \(x \in \mathbb{R}\):
\[
E|\hat{g}_{N,n}(x) - g(x)|^2 \to 0.
\] (4.5)

**Proof.** Throughout the proof, let \(K_h(x) := K(x/h), x \in \mathbb{R}\). Consider (4.2). Note \(E\hat{g}_{N,n}(x) = h^{-1}EK_h(x - \tilde{a}_n)\), because \(\tilde{a}_{i,n}, i = 1, \ldots, N\), are identically distributed. Let us approximate \(\hat{g}_{N,n}(x)\) by
\[
\tilde{g}_N(x) := \frac{1}{Nh} \sum_{i=1}^{N} K_h(x - a_i), \quad x \in \mathbb{R},
\] (4.6)
which satisfies \(E\tilde{g}_N(x) = h^{-1}EK_h(x - a) \to g(x)\) as \(h \to 0\) at a continuity point \(x\) of \(g\), see Parzen (1962). Integration by parts and Corollary 2.3 yield
\[
h|E\hat{g}_{N,n}(x) - E\tilde{g}_N(x)| = \left| \int_{\mathbb{R}} (G_n(y) - G(y))dK_h(x - y) \right| \leq V(K) \sup_{y \in [-1, 1]} |G_n(y) - G(y)| = O(n^{-(p/2)^{(p-1)}}/(1+p)),
\] (4.7)
uniformly in $x \in \mathbb{R}$, where $V(K)$ denotes the total variation of $K$ and $V(K) = V(K_h)$. This proves (4.2).

Next, let us prove (4.4). We have

$$Nh \text{Cov}(\tilde{g}_N(x_1), \tilde{g}_N(x_2)) = \frac{1}{h} EK_h(x_1 - a)K_h(x_2 - a) \to \begin{cases} g(x_1)\|K\|_2^2 & \text{if } x_1 = x_2, \\ 0 & \text{if } x_1 \neq x_2, \end{cases} \quad (4.8)$$

as $h \to 0$ at any points $x_1, x_2$ of continuity of $g$, see Parzen (1962). Split $Nh\{\text{Cov}(\tilde{g}_N(x_1), \tilde{g}_N(x_2)) - \text{Cov}(\tilde{g}_N(x_1), \tilde{g}_N(x_2))\} = \sum_{i=1}^3 Q_i(x_1, x_2)$, where

$$Q_1(x_1, x_2) := h^{-1}[EK_h(x_1 - \tilde{a}_n)K_h(x_2 - \tilde{a}_n) - EK_h(x_1 - a)K_h(x_2 - a)],$$

$$Q_2(x_1, x_2) := h^{-1}[EK_h(x_1 - \tilde{a}_n)EK_h(x_2 - \tilde{a}_n) - EK_h(x_1 - a)EK_h(x_2 - a)],$$

$$Q_3(x_1, x_2) := (N - 1)h^{-1} \text{Cov}(K_h(x_1 - \tilde{a}_1,n), K_h(x_2 - \tilde{a}_2,n)).$$

Note $Q_3(x_1, x_2) = 0$ in Case (i). Similarly to (4.7), $|Q_1(x_1, x_2)| = h^{-1}|\int \int \{G_n(y) - G(y)\}dK_h(x_1 - y)K_h(x_2 - y)| \leq C h^{-1}n^{-(p/2\wedge(p-1))/(1+p)} \to 0$ since $V(K_h(x_1 - y)K_h(x_2 - y)) \leq C$ and $|Q_2(x_1, x_2)| \leq C h^{-1}n^{-(p/2\wedge(p-1))/(1+p)} \to 0$ uniformly in $x_1, x_2$. Finally,

$$|Q_3(x_1, x_2)| = (N - 1)h^{-1}\left|\int \int \{P(\tilde{a}_1,n \leq y_1, \tilde{a}_2,n \leq y_2) - P(\tilde{a}_1,n \leq y_1)P(\tilde{a}_2,n \leq y_2)\}dK_h(x_1 - y_1)dK_h(x_2 - y_2)\right|$$

$$ \leq C N h^{-1} \sup_{y_1, y_2 \in [-1,1]} |P(\tilde{a}_1,n \leq y_1, \tilde{a}_2,n \leq y_2) - P(\tilde{a}_1,n \leq y_1)P(\tilde{a}_2,n \leq y_2)|$$

$$ = O(N h^{-1}n^{-(p/2\wedge(p-1))/(1+p)}) = o(1),$$

proving (4.4) and the proposition. \qed

**Remark 4.1** It follows from the proof of the above proposition that in the case of a (uniformly) continuous density $g(x), x \in [-1,1]$ relations (4.2), (4.5) and the first relation in (4.4) hold uniformly in $x \in \mathbb{R}$, implying the convergence of the mean integrated squared error:

$$\int_{-\infty}^{\infty} E|\tilde{g}_{N,n}(x) - g(x)|^2 dx \to 0.$$

**Proposition 4.2** (Asymptotic normality) *Let assumptions B and A7 hold and assume Nh $\to \infty$ in addition to (4.3). Moreover, let $K$ be a Lipschitz function in Case (ii). Then*

$$\frac{\tilde{g}_{N,n}(x) - E\tilde{g}_{N,n}(x)}{\sqrt{\text{Var}(\tilde{g}_{N,n}(x))}} \to_d N(0,1), \quad (4.9)$$

*at every continuity point $x \in (-1,1)$ of $g$.*

**Proof.** First, consider Case (i). Since $\tilde{g}_{N,n}(x) = (Nh)^{-1}\sum_{i=1}^N V_i,N$ is a (normalized) sum of i.i.d. r.v.s $V_i,N := K_h(x - \tilde{a}_i,n)$ with common distribution $V_N := V_{1,N}$, it suffices to verify Lyapunov’s condition

$$\frac{E|V_N - EV_N|^{2+\delta}}{N^{\delta/2} [\text{Var} (V_N)]^{(2+\delta)/2}} \to 0,$$

(4.10)

for some $\delta > 0$. This follows by the same arguments as in Parzen (1962). Analogously to Proposition 4.1, we have $E|V_N|^{2+\delta} = E|K_h(x - \tilde{a}_n)|^{2+\delta} \sim hg(x) \int_1^1 |K(y)|^{2+\delta} dy = O(h)$ while $\text{Var} (V_N) = Nh^2 \text{Var} (\tilde{g}_{N,n}(x)) \sim hg(x)\|K\|_2^2$ according to (4.4). Hence the l.h.s. of (4.10) is $O((Nh)^{-\delta/2}) = o(1)$, proving (4.9) in Case (i).
Let us turn to Case (ii). It suffices to prove that \( \sqrt{Nh}(\hat{g}_{N,n}(x) - \hat{g}_N(x)) \to_p 0 \), for \( \hat{g}_N(x) \) given in (4.6). By \(|K(x) - K(y)| \leq L_K|x - y|, x, y \in \mathbb{R} \), for \( \epsilon > 0 \)

\[
P\left( \frac{\sqrt{Nh}}{\sqrt{Nh}}|\hat{g}_{N,n}(x) - \hat{g}_N(x)| > \epsilon \right) \leq P\left( \frac{L_K}{Nh} \sum_{i=1}^{N} |\hat{a}_{i,n} - a_i| > \epsilon \right)
\]

\[
\leq NP\left( |\hat{a}_n - a| \right. > \sqrt{Nh}\left( \frac{h}{N} \right) \frac{\epsilon}{L_K}
\]

\[
\leq C(h Nh)^{-p/2}(Nh)^{1+p}n^{-(p/2)(p-1)} + (N/n) = o(1)
\]

follows from Proposition 2.1 and (4.3) with \( Nh \to \infty \). \qed

**Corollary 4.3** Let assumptions of Proposition 4.2 hold with \( h \sim cN^{-1/5} \) for some \( c > 0 \), i.e.

\[
N = \begin{cases} 
\begin{array}{ll}
o(n^{\frac{2}{5}} + (p-1)^2) & \text{in Case (i),} \\
o(n^{\frac{2}{5}} + (p-1)^2) & \text{in Case (ii).}
\end{array} 
\end{cases}
\]

Moreover, let \( g \in C^2[-1, 1] \) and \( \int_{-1}^{1} yK(y)dy = 0 \). Then

\[
N^{2/5}(\hat{g}_{N,n}(x) - g(x)) \to_d N(\mu(x), \sigma^2(x)),
\]

where \( \mu(x) := (e^2/2)g''(x)(x)1_2(K) \) and \( \sigma^2(x) := (1/e)g(x)\|K\|_2^2 \).

**Proof** follows from Proposition 4.2, by noting that \( E\hat{g}_N(x) - g(x) \sim h^2g''(x)\mu_2(K) \) as \( h \to 0 \) and \( E\hat{g}_{N,n}(x) - E\hat{g}_N(x) = O(h^{-1}n^{-(p/2)(p-1)}/(1+p)) \) by (4.7). \qed

## 5 Simulations for goodness-of-fit testing

In this section we compare our nonparametric goodness-of-fit test in (3.7) for testing the null hypothesis \( H = G_0 \) with its parametric analogue studied in Beran et al. (2010). In accordance with the last paper, we assume \( \{X_i(t)\} \) in (3.1) to be independent AR(1) processes with standard normal i.i.d. innovations \( \{\zeta_i(t)\}, \zeta(0) \sim N(0, 1) \) and the random autoregressive coefficient \( a_i \in (0, 1) \) having a Beta-type density \( g(x) \) with unknown parameters \( \theta := (\alpha, \beta)' \):

\[
g(x) = \frac{2}{B(\alpha, \beta)}x^{2\alpha-1}(1-x)^{\beta-1}, \quad x \in (0, 1), \quad \alpha > 1, \beta > 1,
\]

where \( B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta) \) is Beta function. Note that \( \beta \in (1, 2) \) implies the long memory property in \( \{X_i(t)\} \). Beran et al. (2010) discuss a maximum likelihood estimate \( \hat{\theta}_{N,n,\kappa} = (\hat{\alpha}, \hat{\beta})' \) of \( \theta = (\alpha, \beta)' \) when each unobservable coefficient \( a_i \) is replaced by its estimate \( \hat{a}_{i,n,\kappa} := \min\{\max\{\hat{a}_{i,n}, \kappa\}, 1 - \kappa\} \) with \( \hat{a}_{i,n} \) given in (3.3) and \( 0 < \kappa = \kappa(N, n) \to 0 \) is a truncation parameter. Under certain conditions on \( N, n \to \infty \) and \( \kappa \to 0 \), Beran et al. (2010, Theorem 2) showed that

\[
N^{1/2}(\hat{\theta}_{N,n,\kappa} - \theta_0) \to_d N(0, A^{-1}(\theta_0)),
\]

where \( \theta_0 \) is the true parameter vector,

\[
A(\theta) := \begin{pmatrix}
\psi_1(\alpha) - \psi_1(\alpha + \beta) & -\psi_1(\alpha + \beta) \\
-\psi_1(\alpha + \beta) & \psi_1(\beta) - \psi_1(\alpha + \beta)
\end{pmatrix},
\]
and $\psi_1(x) := d^2 \ln \Gamma(x)/dx^2$ is the Trigamma function. Based on (3.7) and (5.2), we consider testing both ways (nonparametrically and parametrically) the hypothesis that the unobserved autoregressive coefficients $\{a_1, \ldots, a_N\}$ are drawn from the reference distribution $G_0$ having density function in (5.1) with a specific $\theta_0$, i.e. the null $G = G_0$ vs. the alternative $G \neq G_0$. The respective test statistics are

$$T_1 := N^{1/2} \sup_x |\hat{G}_{N,n}(x) - G_0(x)| \quad \text{and} \quad T_2 := N(\hat{\theta}_{N,n,\kappa} - \theta_0)'A(\theta_0)(\hat{\theta}_{N,n,\kappa} - \theta_0).$$

(5.3)

Under the null hypothesis, the statistics $T_1$ and $T_2$ converge to the Kolmogorov distribution and the chi-square distribution with 2 degrees of freedom, respectively, see (3.7), (5.2).

To compare the performance of the above testing procedures, we compute the empirical distribution of the p-value of $T_1$ and $T_2$ under null and alternative hypotheses. The p-value of observed $T_i$ is defined as $p(T_i) = 1 - \mathcal{K}_i(T_i), i = 1, 2$, where $\mathcal{K}_i(x), i = 1, 2$ denote the limit distribution functions of (5.3). Recall that when the significance level of the test is correct, the (asymptotic) distribution of the p-value is uniform on [0,1]. The simulation procedure to compare the performance of $T_1$ and $T_2$ is the following:

Step $S_0$ We fix the parameter under the null hypothesis $H_0: \theta = \theta_0$ with $\theta_0 = (2, 1.4)'$.

Step $S_1$ We simulate 5000 panels with $N = 250$, $n = 817$ for five chosen values $\theta = (2, 1.2)', (2, 1.3)', (2, 1.4)', (2, 1.5)', (2, 1.6)'$ of Beta parameters.

Step $S_2$ For each simulated panel we compute the p-value of statistics $T_1$ and $T_2$.

Step $S_3$ The empirical c.d.f.’s of computed p-values of statistics $T_1$ and $T_2$ are graphed.

The values of Beta parameters $\theta_0 = (2, 1.4)'$, $N, n$ were chosen in accordance with the simulation study in Beran et al. (2010).

Fig. 1 presents the simulation results under the true hypothesis $\theta = \theta_0$ with zoom-in on small p-values. We see that both c.d.f.’s in the left graph are approximately linear. Somewhat surprisingly, it appears that the empirical size of $T_1$ (the nonparametric test) is better than the size of $T_2$ (the parametric test). Particularly, for significance levels 0.05 and 0.1 we provide the empirical size values in Table 1.

Fig. 2 gives the graphs of the empirical c.d.f.’s of p-values of $T_1$ and $T_2$ for several alternatives $\theta \neq \theta_0$. It appears that for $\beta > \beta_0 = 1.4$ the parametric test $T_2$ is more powerful than the nonparametric test $T_1$ but for $\beta < \beta_0$ the power differences are less significant. Table 1 illustrates the empirical power for the significance levels 0.05, 0.1.

| Signif. level | 5%          | 10%          |
|--------------|-------------|-------------|
| $\beta$      | 1.2  1.3  1.4  1.5  1.6 | 1.2  1.3  1.4  1.5  1.6 |
| $T_1$        | .532 .137 .049 .208 .576 | .653 .223 .103 .315 .702 |
| $T_2$        | .500 .104 .077 .313 .735 | .634 .184 .134 .421 .827 |

Table 1: Numerical results of the comparison for testing procedure $H_0: \theta = (2, 1.4)'$ at significance level 5% and 10% . The column for $\beta = 1.4$ provides the empirical size.

The above simulations (Fig. 1 and 2, Table 1) refer to the case of independent individual processes $\{X_i(t)\}$. There are no theoretical results for the parametric test $T_2$, when AR(1) series are dependent. Although the nonparametric test $T_1$ is valid for the latter case, one may expect that the presence of the common shock component in the panel data in (3.2) has a negative effect on the test performance for short series. To illustrate this effect, we simulate 5000 panels with AR(1) processes $\{X_i(t)\}$ driven by dependent shocks in
Figure 1: [left] Empirical c.d.f. of p-values of $T_1$ and $T_2$ under $H_0 : \theta_0 = (2, 1.4)'$; 5000 replications with $N = 250$, $n = 817$. [right] Zoom-in on the region of interest: p-values smaller than 0.1.

$$(3.2) \text{ with } b_i = b, \ c_i = (1 - b^2)^{1/2}. \text{ As previously, we choose } \theta_0 = (2, 1.4)'$, $N = 250$, $n = 817$ and we fix $\theta = (2, 1.4)'$ to evaluate the empirical size of $T_1$. Fig. 3 [left] presents the graphs of the empirical c.d.f.'s of the p-values of $T_1$ for $b = 1$, $b = 0.6$ and $b = 0$, the latter corresponding to independent individual processes as in Fig. 1. We see that the size of the test worsens when $b$ increases, particularly when $b = 1$ and the individual processes are all driven by the same common noise. To overcome the last effect, the sample length $n$ of each series in the panel may be increased as in Fig. 3 [right], where the choice of $n = 5500$ and $b = 1$ shows a much better performance of $T_1$ under the null hypothesis $\theta = \theta_0 = (2, 1.4)'$ and the alternative ($\theta = (2, 1.5)'$ and $\theta = (2, 1.6)'$) scenarios.

In conclusion,

1. We do not observe an important loss of the power for the nonparametric KS test $T_1$ compared to the parametric approach.

2. The KS test $T_1$ does not require to choose any tuning parameter contrary to the test $T_2$.

3. One can use the KS test $T_1$ under weaker assumptions on AR(1) innovations. We only impose moment conditions. The dependence between the series is allowed by $(3.2)$. 

11
Figure 2: Empirical c.d.f. of p-values of $T_1$ and $T_2$ for testing $H_0 : \theta_0 = (2, 1.4)'$ under several alternatives of the form $\theta = (2, \beta)'$; 5000 replications with $N = 250$, $n = 817$. 
Figure 3: [left] Empirical c.d.f. of p-values of $T_1$ under $H_0: \theta_0 = (2, 1.4)'$ for different dependence structure between AR(1) series: $b_i = b$ and $c_i = \sqrt{1 - b^2}$ and $N = 250, n = 817$. [right] Empirical c.d.f. of p-values of $T_1$ for testing $H_0: \theta_0 = (2, 1.4)'$. AR(1) series are driven by common innovations, i.e. $b_i = 1, c_i = 0$, for $\theta = (2, \beta)'$; 5000 replications with $N = 250, n = 5500$. 

References

Beran, J., Schützner, M. and Ghosh, S. (2010). From short to long memory: Aggregation and estimation. *Computational Statistics and Data Analysis, 54*, 2432–2442.

Burkholder, D. L. (1973). Distribution function inequalities for martingales. *Annals of Probability, 1*, 19–42.

Celov, D., Leipus, R. and Philippe, A. (2007). Time series aggregation, disaggregation and long memory. *Lithuanian Mathematical Journal, 47*, 379–393.

Celov, D., Leipus, R. and Philippe, A. (2010). Asymptotic normality of the mixture density estimator in a disaggregation scheme. *Journal of Nonparametric Statistics, 22*, 425–442.

Giraitis, L., Koul, H. L. and Surgailis, D. (2012). *Large Sample Inference for Long Memory Processes*. Imperial College Press, London.

Granger, C. W. J. (1980). Long memory relationship and the aggregation of dynamic models. *Journal of Econometrics, 14*, 227–238.

Horváth, L. and Leipus, R. (2009). Effect of aggregation on estimators in AR(1) sequence. *TEST, 18*, 546–567.

Jirak, M. (2013). Limit theorems for aggregated linear processes. *Advances in Applied Probability, 45*, 520–544.

Leipus, R., Oppenheim, G., Philippe, A. and Viano, M.-C. (2006). Orthogonal series density estimation in a disaggregation scheme. *Journal of Statistical Planning and Inference, 136*, 2547–2571.

Parzen, E. (1962). On estimation of probability density function and mode. *Annals of Mathematical Statistics, 33*, 1065–1076.

Robinson, P. M. (1978). Statistical inference for a random coefficient autoregressive model. *Scandinavian Journal of Statistics, 5*, 163–168.

Rosenthal, H. P. (1970). On the subspaces of $L^p$ ($p > 2$) spanned by the sequences of independent random variables. *Israel Journal of Mathematics, 8*, 273–303.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

von Bahr, B. and Esséén, C.-G. (1965). Inequalities for the $r$th absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Annals of Mathematical Statistics, 36*, 299–303.

6 Appendix: some proofs and auxiliary lemmas

We use the following martingale moment inequality.
Lemma 6.1 Let $p > 1$ and $\{\xi_j, j \geq 1\}$ be a martingale difference sequence: $E[\xi_j|\xi_1, \ldots, \xi_{j-1}] = 0$, $j = 2, 3, \ldots$ with $E|\xi|^p < \infty$. Then there exists a constant $C_p < \infty$ depending only on $p$ and such that

$$
E\left|\sum_{j=1}^{\infty} \xi_j^p\right| \leq C_p \left\{ \begin{array}{ll}
\sum_{j=1}^{\infty} E|\xi_j|^p, & 1 < p \leq 2, \\
\left( \sum_{j=1}^{\infty} (E|\xi_j|^p)^{2/p} \right)^{p/2}, & p > 2.
\end{array} \right. \tag{6.1}
$$

For $1 < p \leq 2$, inequality (6.1) is known as von Bahr and Essén inequality, see von Bahr and Essén (1965), and for $p > 2$, it is a consequence of the Burkholder and Rosenthal inequality (Burkholder (1973), Rosenthal (1970), see also Giraitis et al. (2012, Lemma 2.5.2)).

Proof of Proposition 2.1. Since $\hat{\alpha}_n$ in (2.5) is invariant w.r.t. a scale factor of innovations $\{\zeta(t)\}$, w.l.g. we can assume $b^2 + c^2 = 1$ and $E\zeta^2(0) = 1$, $E|\zeta(0)|^2p < \infty$. Then $\hat{\alpha}_n - a = \delta_{n1} + \delta_{n2}$, where

$$
\delta_{n1} := -\frac{aX^2(n)}{\sum_{t=1}^{n} X^2(t)}, \quad \delta_{n2} := \frac{\sum_{t=1}^{n-1} X(t)\zeta(t+1)}{\sum_{t=1}^{n} X^2(t)}. \tag{6.2}
$$

The statement of the proposition follows from

$$
P(|\delta_{ni}| > \gamma) \leq C(n^{-1} + n^{-(p/2)\wedge(p-1)}\gamma^{-p}) \quad (0 < \gamma < 1, i = 1, 2). \tag{6.3}
$$

To show (6.3) for $i = 1$, note that $P(|\delta_{n1}| > \gamma) \leq P(|S_{n1} > n/2) + P(S_{n2} > n\gamma/2)$, where $S_{n1} := \sum_{t=1}^{n}((1 - a^2)X^2(t) - 1)$, $S_{n2} := (1 - a^2)X^2(n)$. Thus, (6.3) for $i = 1$ follows from

$$
E|S_{n1}|^{p/2} \leq Cn \quad \text{and} \quad E|S_{n2}|^p \leq C. \tag{6.4}
$$

Consider the first relation in (6.4). Clearly, it suffices to prove it for $1 < p \leq 2$ only. We have $S_{n1} = 2S'_{n1} + S''_{n1}$, where

$$
S'_{n1} := (1 - a^2) \sum_{s_2<s_1 \leq n} \sum_{t=1 \vee s_1}^{n} a^{2(t-s_1)}a^{s_1-s_2}\zeta(s_1)\zeta(s_2),
$$

$$
S''_{n1} := (1 - a^2) \sum_{s \leq n} \sum_{t=1 \vee s}^{n} a^{2(t-s)}(\zeta^2(s) - 1).
$$

We will use the following elementary inequality: for any $-1 \leq a \leq 1$, $n \geq 1$, $s \leq n$

$$
\alpha_n(s) := (1 - a^2) \sum_{t=1 \vee s}^{n} a^{2(t-s)} = \begin{cases} 
    a^{2(1-s)}(1 - a^{2n}), & s \leq 0, \\
    1 - a^{2(n+1-s)}, & 1 \leq s \leq n \\
    a^{-2s} \min(1, 2n(1 - |a|)), & s \leq 0, \\
    1, & 1 \leq s \leq n.
\end{cases} \tag{6.5}
$$

Using the independence of $\{\zeta(s)\}$ and $a$ and inequality (6.1) (twice) for $1 < p \leq 2$ we obtain

$$
E|S'_{n1}|^p = E \left| \sum_{s_1 \leq n} \alpha_n(s_1)\zeta(s_1) \sum_{s_2<s_1} a^{s_1-s_2}\zeta(s_2) \right|^p 
\leq CE \sum_{s_1 \leq n} |\alpha_n(s_1)\zeta(s_1) \sum_{s_2<s_1} a^{s_1-s_2}\zeta(s_2)|^p 
\leq CE \sum_{s_1 \leq n} |\alpha_n(s_1)|^p \sum_{s_2<s_1} |a|^{p(s_1-s_2)} 
\leq CE(1 - |a|)^{-1} \sum_{s \leq n} |\alpha_n(s)|^p \leq Cn
$$
since $E(1 - |a|)^{-1} < \infty$ (see (2.3)) and $\sum_{s \leq n} |\alpha_n(s)|^p \leq Cn$ follows from (6.5). Similarly, since $\{\zeta^2(s) - 1, s \leq n\}$ form a martingale difference sequence,

$$E|S''_{n1}|^p \leq CE \sum_{s \leq n} |\alpha_n(s)|^p \leq Cn,$$

proving the first inequality (6.4).

Next, consider the second inequality in (6.4). We have $S_{n2} = 2S'_{n2} + S''_{n2}$, where

$$S'_{n2} := (1 - a^2) \sum_{s_2 < s_1 \leq n} a^{2(n-s_1)}a^{s_1-s_2}\zeta(s_1)\zeta(s_2), \quad S''_{n2} := (1 - a^2) \sum_{s \leq n} a^{2(n-s)}(\zeta^2(s) - 1).$$

Similarly as above, we obtain $E|S''_{n2}|^p \leq CE \sum_{s \leq n} |(1 - a^2)a^{-2(n-s)}|^p \leq C$ and $E|S'_{n2}|^p \leq CE \sum_{s_2 < s_1 \leq n} |(1 - a^2)|a^{2(n-s_1)}|a|^{s_1-s_2}|^p \leq CE(1 - |a|)^{p-2} < \infty$, proving (6.4) and hence (6.3) for $i = 1$.

Consider (6.3) for $i = 2$. We have $\delta_{n2} = R_n/(n + S_{n1})$, where $R_n := (1 - a^2) \sum_{t=1}^{n} X(t)\zeta(t+1)$ and $S_{n1} = \sum_{t=1}^{n} ((1 - a^2)X(t) - t - 1)$ is the same as in (6.4). Then $P(|\delta_{n2}| > \gamma) \leq P(|R_n| > n\gamma/2) + P(|S_{n1}| > n/2)$, where

$$P(|S_{n1}| > n/2) \leq (n/2)^{-p}\sum_{s \leq n} |a|^p \leq C \begin{cases} n^{-p-1}, & 1 < p \leq 2, \\ n^{-1}, & p > 2, \end{cases}$$

according to (6.4). Therefore (6.3) for $i = 2$ follows from

$$E|R_n|^p \leq C \begin{cases} n, & 1 < p \leq 2, \\ n^{p/2}, & p > 2, \end{cases} \tag{6.6}$$

Since $R_n = (1 - a^2) \sum_{s \leq n-1} \zeta(s) \sum_{t=1}^{n-1} a^{t-s}\zeta(t+1)$ is a sum of martingale differences, by inequality (6.1) with $1 < p \leq 2$ we obtain

$$E|R_n|^p \leq CE \sum_{s \leq n-1} |(1 - a^2)\zeta(s) \sum_{t=1}^{n-1} a^{t-s}\zeta(t+1)|^p \leq CE|1 - a^2|^p \sum_{s \leq n-1} \sum_{t=1}^{n-1} |a|^{p(t-s)} \leq CE|1 - a^2|^2 \{ (1 - |a|^p)^{-2} + n(1 - |a|^p)^{-1} \} \leq Cn,$$

proving (6.6) for $p \leq 2$. Similarly, using (6.1) with $p > 2$ we get

$$E|R_n|^p = E|1 - a^2|^p E \left[ \left| \sum_{s \leq n-1} \zeta(s) \sum_{t=1}^{n-1} a^{t-s}\zeta(t+1) \right|^p |a| \right] \leq CE|1 - a^2|^p \left[ \sum_{s \leq n-1} \left( E \left[ |\zeta(s) \sum_{t=1}^{n-1} a^{t-s}\zeta(t+1) |^p \right] |a| \right)^{2/p} \right]^{p/2} \leq CE|1 - a^2|^p \left[ \sum_{s \leq n-1} \sum_{t=1}^{n-1} a^{2(t-s)} \right]^{p/2} \leq CE|1 - a^2|^p \left[ \sum_{s \leq n-1} \sum_{t=1}^{n-1} a^{2(t-s)} \right]^{p/2} \leq CE|1 - a^2|^p \left[ \sum_{s \leq n-1} \sum_{t=1}^{n-1} a^{2(t-s)} \right]^{p/2} \leq CE|1 - a^2|^p \{ (1 - a^2)^{-2} + n(1 - a^2)^{-1} \}^{p/2} \leq Cn^{p/2},$$
proving (6.6), (6.3) and the proposition. □

Let \( a, a_1, \ldots, a_N \) be i.i.d. r.v.s with d.f. \( G(x) = P(a \leq x) \) supported by \([-1, 1]\). Define \( \hat{G}_N(x) := N^{-1} \sum_{i=1}^{N} 1(a_i \leq x) \), \( U_N(x) := N^{1/2}(\hat{G}_N(x) - G(x)) \), \( x \in \mathbb{R} \), and \( \omega_N(\delta) \) (= the modulus of continuity of \( U_N \)) by

\[
\omega_N(\delta) := \sup_{0 \leq y - x \leq \delta} |U_N(y) - U_N(x)|, \quad \delta > 0.
\]

**Lemma 6.2** Assume that \( G \) satisfies assumption \( A_6 \). Then for all \( \epsilon > 0 \),

\[
\epsilon^4 P(\omega_N(\delta) > 6\epsilon) \leq (3 + 3C)L_G(\delta) + N^{-1},
\]

where \( C \) is a constant independent of \( \epsilon, \delta, N \).

**Proof.** As in Shorack and Wellner (1986, p. 110, Inequality 1) we have that

\[
E[|U_N(y) - U_N(x)|^4] \leq 3P(a \in (x, y)]^2 + N^{-1}P(a \in (x, y]),
\]

for \(-1 \leq x \leq y \leq z \leq 1\). Now fix \( \delta > 0 \) and split \([-1, 1] = \bigcup_i \Delta_i \), where \( \Delta_i = [-1 + i\delta, -1 + (i + 1)\delta] \), \( i = 0, 1, \ldots, [2/\delta] - 1 \), \( \Delta_{[2/\delta]} = [-1 + [2/\delta]\delta, 1] \). According to Shorack and Wellner (1986, p. 49, Lemma 1), for all \( \epsilon > 0 \),

\[
\epsilon^4 P(\omega_N(\delta) > 6\epsilon) \leq (3 + 3C) \max_i P(a \in \Delta_i) + N^{-1},
\]

where \( C \) is a constant independent of \( \epsilon, \delta, N \). Lemma follows from Assumption \( A_6 \) on d.f. \( G \) of a r.v. \( a \). □

Note that if we take \( \delta = \delta_N = o(1) \), we then get \( P(\omega_N(\delta) > \epsilon) \to 0 \) as \( N \to \infty \).

**Lemma 6.3** Let \( \hat{a}_{1,n}, \hat{a}_{2,n} \) be given in (3.3) under assumptions \( A_1 - A_6 \) with \( p = 1 \). Then for all \( \gamma \in (0, 1) \) and \( n \geq 1 \), it holds

\[
\sup_{x, y \in [-1, 1]} |P(\hat{a}_{1,n} \leq x, \hat{a}_{2,n} \leq y) - P(\hat{a}_{1,n} \leq x)P(\hat{a}_{2,n} \leq y)| = O(n^{-(p/2)\wedge(p-1)}/(1+p)).
\]

**Proof.** Define \( \delta_{1,n} := \hat{a}_{i,n} - a_i, \ i = 1, 2 \). For \( \gamma \in (0, 1) \), we have

\[
P(|\delta_{1,n}| > \gamma \text{ or } |\delta_{2,n}| > \gamma) \leq P(|\delta_{1,n}| > \gamma) + P(|\delta_{2,n}| > \gamma) \leq C(n^{-(p/2)\wedge(p-1)}\gamma^{-p} + n^{-1}),
\]

by Proposition 2.1. Consider now

\[
P(\hat{a}_{1,n} \leq x, \hat{a}_{2,n} \leq y) = P(a_1 + \delta_{1,n} \leq x, a_2 + \delta_{2,n} \leq y)
\]

\[
\leq P(a_1 + \delta_{1,n} \leq x, a_2 + \delta_{2,n} \leq y, |\delta_{1,n}| \leq \gamma, |\delta_{2,n}| \leq \gamma) + P(|\delta_{1,n}| > \gamma \text{ or } |\delta_{2,n}| > \gamma).
\]

Then

\[
P(a_1 + \delta_{1,n} \leq x, a_2 + \delta_{2,n} \leq y, |\delta_{1,n}| \leq \gamma, |\delta_{2,n}| \leq \gamma) \leq P(a_1 \leq x + \gamma, a_2 \leq y + \gamma, |\delta_{1,n}| \leq \gamma, |\delta_{2,n}| \leq \gamma)
\]

\[
\leq G(x + \gamma)G(y + \gamma)
\]

and

\[
P(a_1 + \delta_{1,n} \leq x, a_2 + \delta_{2,n} \leq y, |\delta_{1,n}| \leq \gamma, |\delta_{2,n}| \leq \gamma) \geq P(a_1 \leq x - \gamma, a_2 \leq y - \gamma, |\delta_{1,n}| \leq \gamma, |\delta_{2,n}| \leq \gamma)
\]

\[
\geq G(x - \gamma)G(y - \gamma) - P(|\delta_{1,n}| > \gamma \text{ or } |\delta_{2,n}| > \gamma).
\]
From (2.7) we obtain

\[ |G(x \pm \gamma)G(y \pm \gamma) - G(x)G(y)| = |(G(x) + O(\gamma))(G(y) + O(\gamma)) - G(x)G(y)| \leq C\gamma. \]

Hence

\[ |P(a_1 \leq x, a_2 \leq y) - G(x)G(y)| \leq C(\gamma + n^{-1} + n^{-\left(p/2\right)\wedge(p-1)}\gamma^{-p}). \]  

(6.7)

In a similar way,

\[ |P(a_1 \leq x)P(a_2 \leq y) - G(x)G(y)| \leq C(\gamma + n^{-1} + n^{-\left(p/2\right)\wedge(p-1)}\gamma^{-p}). \]  

(6.8)

By (6.7), (6.8), the proof of the lemma is complete with \( \gamma = \gamma_n = o(1) \), which satisfies \( \gamma_n \sim n^{-\left(p/2\right)\wedge(p-1)}\gamma_n^{-p} \). 

\( \square \)