T E C H N I C A L
R E P O R T

0507

SERIAL AUTOREGRESSION AND REGRESSION RANK SCORES STATISTICS

HALLIN, M., JURECKOVA, J. and H. L. KOUL

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Serial Autoregression and Regression Rank Scores Statistics[1]

by

Marc Hallin [2]
Université Libre de Bruxelles, Brussels, Belgium

Jana Jurečková[3]
Charles University, Prague, Czech Republic

Hira L. Koul [4]
Michigan State University, East Lansing, Michigan, U.S.A.

Abstract

This paper establishes an asymptotic representation for regression and autoregression rank score statistics of the serial type. Applications include rank-based versions of the Durbin-Watson test, tests of AR(p) against AR($p+1$) dependence, or detection of the presence of random components in AR processes.

1 Introduction

1.1 Rank tests

Rank tests are known to be robust, distribution-free yet powerful alternative to Gaussian testing methods under a broad set of model assumptions. The *classical* theory of rank tests was mainly developed in the context of linear models with independent errors, but the domain of application of this theory naturally extends to the much broader class of semiparametric models under which the distribution $P_{n;\theta;f}$ of the observation vector $\boldsymbol{X}_n := (X_{n1}, \cdots, X_{nn})'$ belongs to a family $\mathcal{P}_n := \{P_{n;\theta;f};\ \theta \in \Theta,\ f \in \mathcal{F}\}$, where $\theta \in \Theta \subseteq \mathbb{R}^k$ is some parameter of interest, and $\mathcal{F}$ is a class of densities f on $\mathbb{R}$. More specifically, rank tests can be constructed whenever

(A) for all n, there exists a $(\theta, \boldsymbol{X}_n)$-measurable *residual function*

$$(\theta, \boldsymbol{X}_n) \mapsto \boldsymbol{\varepsilon}_n(\theta, \boldsymbol{X}_n) := (\varepsilon_{n,1}(\theta, \boldsymbol{X}_n),\ \cdots,\ \varepsilon_{n,n}(\theta, \boldsymbol{X}_n))'$$

---

[1] *Key words and phrases*: Time series; robustness; serial autoregression rank scores; rank tests.
Subject Classification: Primary 62M10, Secondary 62G10.

[2] Research supported by a P.A.I. contract of the Belgian Federal Government and an Action de Recherche Concertée of the Communauté française de Belgique.

[3] Research supported by the Czech Republic Grant 201/05/2340

[4] Research partly supported by the FNRS of the Communauté française de Belgique.

such that the distribution of $\boldsymbol{X}_n$ is $P_{n;\theta;f}$ iff the components of the vector $\boldsymbol{\varepsilon}_n(\theta, \boldsymbol{X}_n)$ are i.i.d. with common density f. Henceforth, we shall write $\varepsilon_{n,t}(\theta)$ instead of $\varepsilon_{n,t}(\theta, \boldsymbol{X}_n)$, $t = 1, \cdots, n$.

Let $R_{n,t}(\theta)$ denote the rank of $\varepsilon_{n,t}(\theta)$ and $\boldsymbol{R}_n(\theta) := (R_{n,1}(\theta), \cdots, R_{n,n}(\theta))'$. The rank tests for the simple null hypothesis $\mathcal{H}_0 : \theta = \theta_0$, where θ_0 is a given value of θ, are based on the vector of residual rank vector $\boldsymbol{R}_n \equiv \boldsymbol{R}_n(\theta_0)$.

The use of the rank tests can be justified by several main arguments.

(a) The vector of ranks is a maximal invariant with respect to the group of *order-preserving transformations* of residuals for a broad class of densities (typically, the class of all non-vanishing densities over the real line, possibly satisfying some regularity assumptions). In such invariant situations, every invariant statistic and test depend only on the maximal invariant, and hence are distribution-free under the null hypothesis.

(b) The rank tests are more robust with respect to some outliers than their parametric counterparts.

(c) Rank-based procedures are asymptotically powerful, as they achieve asymptotic semi-parametric information lower bounds, which is the best we can hope for when making inference about θ in the presence of an unknown f, at a prespecified value of f. Moreover, in linear or ARMA models where semiparametric and parametric efficiencies coincide, there are rank-based tests that yield asymptotically uniformly more powerful tests than their classical counterparts: for instance, the asymptotic Pitman relative efficiencies of normal score or van der Waerden rank tests, with respect to the corresponding normal-theory tests, are uniformly larger than one; cf. Chernoff and Savage (1958), Hallin (1994), Paindaveine (2004, 2005), among others.

A general result by Hallin and Werker (2003) shows that under LAN set up with *central sequences* $\Delta_{n;f}(\theta)$ and under some conditions on $\mathcal{F}$, a semiparametrically efficient inference about θ, at given (θ_0, f), can be based on the distribution-free rank-based efficient central sequence obtained by conditioning the central sequence $\Delta_{n;f}(\theta_0)$ on the vector of ranks $\boldsymbol{R}_n(\theta_0)$, under $\mathcal{H}_0$. In linear regression models where for some known non-random $p \times 1$ design vector $\{c_{n,t};\ 1 \le t \le n\}$, $\varepsilon_{n,t}(\theta) = X_{n,t} - c'_{n,t}\theta$ are i.i.d., rank-based efficient central sequences (at least, under "approximate score version"; see Hallin and Werker (2003) for details) take the form of *linear rank statistics* vectors $S_{n,\varphi}(\theta_0)$, where

$$S_{n,\varphi}(\theta) := \sum_{t=1}^{n} \varphi\Big(\frac{R_{n,t}(\theta)}{n+1}\Big) c_{n,t},$$

φ is a *score-generating function* from (0, 1) to $\mathbb{R}$. Classical asymptotic representation results show that, under some general conditions and under $P_{n,\theta,f}$,

$$S_{n,\varphi}(\theta) = \sum_{t=1}^{n} \varphi(F(\varepsilon_t(\theta)))c_{n,t} + o_P(n^{1/2}), \quad n \to \infty, \quad \forall\, \theta \in \Theta,$$

where F is the distribution function associated with f.

In ARMA models, rank-based central sequences can be expressed (Hallin, Ingenbleek and Puri 1985; Hallin and Puri 1988; Bentarzi and Hallin 1996) as linear combinations of *serial linear rank statistics* $S_{n,\varphi_1\varphi_2}(\theta_0)$, where

$$S_{n,\varphi_1\varphi_2}(\theta) := \sum_{t=i+1}^{n} \varphi_1\Big(\frac{R_{n,t}(\theta)}{n+1}\Big)\varphi_2\Big(\frac{R_{n,t-i}(\theta)}{n+1}\Big), \quad i = 1, 2, \cdots, \tag{1.2}$$

where φ_1 and φ_2 are adequately centered and scaled score generating functions. In more general problems– detection of random coefficients (Akharif and Hallin 2003), detection of nonlinearities (Benghabrit and Hallin 1992; Allal and El-Melhaoui 2005, among others)– rank-based efficient central sequences involve more complex serial linear rank statistics of the form

$$S_{n,\varphi_1\cdots\varphi_m}(\theta) := \sum_{t=i_{m-1}+1}^{n} \varphi_1\Big(\frac{R_{n,t}(\theta)}{n+1}\Big)\cdots\varphi_m\Big(\frac{R_{n,t-i_{m-1}}(\theta)}{n+1}\Big),\ 1 \leq i_1 \leq \cdots \leq i_{m-1}, \tag{1.3}$$

where φ_1, $\varphi_2 \cdots, \varphi_m$ are m score functions. Hallin et al. (1985) show under some general conditions that for every $\theta \in \Theta$,

$$S_{n,\varphi_1\cdots\varphi_m}(\theta) = \sum_{t=i_{m-1}+1}^{n} \varphi_1(F(\varepsilon_t(\theta)))\cdots\varphi_m(F(\varepsilon_{t-i_{m-1}}(\theta))) + o_P(n^{1/2}), \qquad n \to \infty.$$

In most problems of practical interest, however, one is interested in testing the composite null hypothesis $\tilde{\mathcal{H}}_0 : \theta \in \Theta_0$, where Θ_0 is a subset of Θ. In this case θ is not completely specified. It is then natural to first obtain an estimate $\hat{\theta}$ of θ under $\tilde{\mathcal{H}}_0$ and use the aligned ranks test statistics $S_{n,\varphi}(\hat{\theta})$ or $S_{n,\varphi_1\cdots\varphi_m}(\hat{\theta})$, cf., e.g., Koul (1970) and Jurečková (1971) for linear regression models; Hallin and Puri (1994) for the ARMA models. Unless adequate projections that would compensate asymptotically for the standardized differences $n^{1/2}(\hat{\theta}-\theta)$ are performed, these statistics typically are not asymptotically distribution-free, and thus are unsuitable for testing purposes. Once such projections are performed, aligned rank tests achieve the same asymptotic performances as those (likewise projected) based on exact ranks; but still, their robustness heavily depends on the robustness of the estimator $\hat{\theta}$ on which the alignment device is based.

1.2 Autoregression and Regression Rank Scores

The lack of robustness of aligned rank statistics motivated Gutenbrunner and Jurečkovà (1992) to introduce *regression rank scores* in the context of linear regression models with independent observations, as an alternative to the aligned ranks. The regression rank scores are n functions $\hat{a}_n(u) = (\hat{a}_{n,1}(u), \cdots, a_{n,n}(u))'$ with $\hat{a}_{n,t} : [0,1] \mapsto [0,1]$, $t = 1, \cdots, n$ obtained from the observations as the solution of a linear programming problem itself depending on $\tilde{\mathcal{H}}_0$; see Section 2.1 below for details. The regression rank score statistic (RSS) corresponding to a function φ is defined as

$$\tilde{S}_{n,\varphi} := -\sum_{t=1}^{n} \int_0^1 \varphi(u) d\hat{a}_{n,t}(u)\ c_{n,t}.$$

Note that this is like $S_{n,\varphi}(\theta)$ but where $\varphi(R_{n,t}(\theta)/(n+1))$ are replaced by $-\int_0^1 \varphi(u)d\hat{a}_{n,t}(u)$. These scores palliate the lack of invariance of aligned ranks. If not exactly (for fixed n) distribution-free, $\tilde{S}_{n,\varphi}$, indeed, contrary to $S_{n,\varphi}(\hat{\theta})$, is asymptotically equivalent to $S_{n,\varphi}(\theta)$ in probability under $P_{n,\theta,f}$, for each θ, hence asymptotically invariant with respect to the group of order-preserving transformations acting on residuals and, therefore, asymptotically distribution-free. Being moreover *regression-invariant* over Θ_0, it is robust against the influence of possible outliers—if not against the possible leverage effect of certain regression constants. And, the asymptotic performance of tests based on $\tilde{S}_{n,\varphi}$ is matching that of the tests based on $S_{n,\varphi}(\theta)$, for all $\theta \in \Theta_0$; see Gutenbrunner et al. (1993).

Koul and Saleh (1995) and Hallin and Jurečková (1999) developed similar ideas for linear autoregressive models where $\theta' = (\rho_0, \rho_1, \cdots, \rho_p)$, and $\varepsilon_t(\theta) = X_i - \rho_0 - \rho_1 X_{t-1} - \cdots - \rho_p X_{t-p}$ are i.i.d. innovations with mean zero. The autoregression rank score statistics these authors consider are of the form

$$\tilde{S}^*_{n,\varphi_1} := -\sum_{t=i+1}^{n} \int_0^1 \varphi_1(u) d\hat{a}_{n,t}(u) X_{t-i},$$

where $\hat{a}_{n,t}(\cdot)$ are the autoregression rank scores defined in Section 2.1 below and φ_1 is a function like φ.

Unlike the linear regression models, in autoregressive models the outliers in the errors affect the leverage points X_{t-i} also. This fact renders the statistics $\tilde{S}^*_{n,\varphi_1}$ non-robust against outliers in the errors. Genuine autoregression rank scores statistics are the serial autoregression rank score statistics obtained from $\tilde{S}^*_{n,\varphi_1}$ after replacing X_{t-i} by $-\int_0^1 \varphi_2(v) d\hat{a}_{n,t-i}(v)$, yielding

$$\tilde{S}_{n,\varphi_1\varphi_2} := \sum_{t=i+1}^{n} \int_0^1 \int_0^1 \varphi_1(u)\varphi_2(v) da_{n,t}(u)\ \mathrm{d}\hat{a}_{n,t-i}(v), \tag{1.4}$$

(when the lag is to be emphasized, we write $\tilde{S}_{n,\varphi_1\varphi_2;i}$) or, more generally,

$$(1.5)\quad \tilde{S}_{n,\varphi_1\cdots\varphi_m} := (-1)^m \sum_{t=i_{m-1}+1}^{n} \int_0^1 \cdots \int_0^1 \varphi_1(u_1)\ \cdots \varphi_m(u_m) d\hat{a}_{n,t}(u_1)\cdots\ d\hat{a}_{n,t-i_{m-1}}(u_m),$$

analogous to the serial rank statistics (1.2) and (1.3). Here $\varphi_j; 1 \leq j \leq m$ are m functions from $(0,1)$ to $\mathbb{R}$.

The main objective of this paper is to obtain an asymptotic representation of these *serial* regression or autoregression rank score statistics for possibly unbounded functions φ_j's, which so far have not been considered in the literature, but can be expected to enjoy the same nice properties (asymptotic invariance, distribution-freeness, and robustness) as their non-serial counterparts $\tilde{S}^*_{n,\varphi}(\theta)$'s.

1.3 Outline of the paper

Section 2 provides the precise conditions under which serial regression or autoregression rank score statistics can be used in hypothesis testing. Section 2.2 describes three potential applications: a version of the classical Durbin-Watson test based on regression rank scores, a test of AR(p) against AR($p+1$) dependence based on autoregression rank scores, and a test, based on serial autoregression rank scores, detecting the presence of a random component in the autoregressive coefficient of an AR(1) model. Technical assumptions are collected in Section 2.3. The main result of this paper is Proposition 3.1 giving an asymptotic representation result for a class of serial autoregression rank score statistics.

2 Notation and basic assumptions

2.1 Autoregression and regression quantiles and rank scores

We shall now recall the definition of autoregression and regression quantiles and rank scores. First consider the stationary linear autoregressive time series model, where starting with an observable p-vector $X_{1-p}, \cdots, X_0$, one observes the process

$$(2.1)\qquad X_t = \rho_0 + \sum_{j=1}^{p} \rho_j X_{t-j} + \varepsilon_t, \quad (\rho_0, \rho_1, \cdots, \rho_p)' \in \mathbb{R}^{1+p}.$$

The errors ε_t are assumed to be i.i.d. with zero mean and variance σ^2. The parameters $\rho^* := (\rho_1, \cdots, \rho_p)'$ are such that all solutions of the equation $1 - \sum_{t=1}^{p} \rho_i z^i = 0$ lie outside the unit sphere and for each t, ε_t is independent of the vector $y^*_{t-1} := (X_{t-1}, \cdots, X_{t-p})'$. Note that this model satisfies the assumption (A) with $k = 1 + p$, $\theta' = (\rho_0, \rho^{*\prime})$, $\boldsymbol{X}'_n =$

$(y_0^{*\prime}, X_1, X_2, \cdots, X_n)$, and $\varepsilon_{n,t}(\theta) = X_t - \rho_0 - \rho' y^*_{t-1}$. Now, let $y'_{t-1} := (1, y^{*\prime}_{t-1})$, and

$$h_\alpha(z) := |z|\Big(\alpha I[z > 0] + (1-\alpha) I[z \le 0]\Big), \quad z \in \mathbb{R}, \ \alpha \in (0,\, 1).$$

Then αth autoregression quantiles $\rho_n(\alpha)' = (\rho_{n0}(\alpha), \rho_n^*(\alpha)')$, for an $0 < \alpha < 1$, are defined as an

$$\text{argmin}_{r_0 \in \mathbb{R},\, r \in \mathbb{R}^p} \sum_{t=1}^n h_\alpha\Big(X_t - r_0 - y^{*\prime}_{t-1} r\Big).$$

The corresponding autoregression rank scores are defined to be an n-vector $\hat{a}_n(\alpha) := (\hat{a}_{n,1}(\alpha), \cdots, \hat{a}_{n,n}(\alpha))'$ in $[0,1]^n$ maximizing $\sum_{t=1}^n X_t a_t$ with respect to vectors $a \in [0,1]^n$, subject to the conditions

$$Y_n'(a - (1-\alpha)1_n) = 0, \tag{2.2}$$

where Y_n is the $n \times (1+p)$ matrix whose tth row is y'_{t-1}, $t = 1, \cdots, n$, $1_n := (1, \cdots, 1)'$, an $n \times 1$ vector of 1's, and 0 in the right hand side is the $(1+p) \times 1$ vector of zeros.

These autoregression quantiles and rank scores are the analogues of their counterparts in linear regression models of Koenker and Bassett (1978) and Gutenbrunner and Jurečková (1992), respectively, defined as follows. In linear regression models the observations $X_{n,t}$ and the $p \times 1$ non-random design vectors $c_{n,t}$ obey the relation

$$X_{n,t} = \beta_0 + c'_{n,t}\beta + \varepsilon_t, \qquad \beta_0 \in \mathbb{R},\, \beta \in \mathbb{R}^p. \tag{2.3}$$

Note that this model satisfies the assumption (A) with $k = 1 + p$, $\theta' = (\beta_0, \beta')$, $\boldsymbol{X}'_n = \{(c'_{n,t}, X_{n,t});\, 1 \le t \le n\}$, and $\varepsilon_{n,t}(\theta) = X_{n,t} - \beta_0 - \beta' c_{n,t}$. Now, let C_n denote the $n \times (1+p)$ matrix whose tth row consists of $(1, c'_{n,t})$, $1 \le t \le n$. An αth *regression quantile* vector $\hat{\theta}_n(\alpha) := (\hat{\beta}_{0n}(\alpha),\, \hat{\beta}_n(\alpha)')$, for an $\alpha \in (0,1)$, is defined as an

$$\text{argmin}_{b_0 \in \mathbb{R},\, b \in \mathbb{R}^p} \sum_{t=1}^n h_\alpha\Big(X_{n,t} - b_0 - c'_{n,t} b\Big).$$

The corresponding regression rank scores are defined to be an n-vector $\hat{a}_n(\alpha) := (\hat{a}_{n,1}(\alpha), \cdots,\, \hat{a}_{n,n}(\alpha))'$ in $[0,1]^n$ maximizing $\sum_{t=1}^n X_t a_t$ with respect to vectors $a \in [0,1]^n$, subject to the conditions $C_n'(a - (1-\alpha)1_n) = 0$.

2.2 Examples

2.2.1 The Durbin-Watson problem

The objective of the classical Durbin-Watson test is the detection of first-order autocorrelation in the noise of a traditional regression model; its extension to higher-order dependencies is straightforward.

The general overarching model is a linear regression with AR(1) errors

$$X_t = \beta_0 + c'_{n,t}\beta + e_t, \qquad e_t = \rho e_{t-1} + \varepsilon_t, \qquad t = 1, \cdots n,$$

where $\rho \in [0,1)$, $\beta_0 \in \mathbb{R}$, $\beta' := (\beta_1, \cdots, \beta_p) \in \mathbb{R}^p$, and $\varepsilon_1, \cdots, \varepsilon_n$ are i.i.d. with density f. The null hypothesis of interest here is $\mathcal{H}_0 : \rho = 0$, against the alternatives of the form $\mathcal{H}_1 : \rho > 0$. Thus, here $\Theta = \mathbb{R}^{1+p} \times [0,1)$ and $\Theta_0 = R^{1+p} \times \{0\}$. The regression parameters β_0, β play the role of nuisance parameters.

Let $\hat{\theta}' := (\hat{\beta}_0, \hat{\beta}')$ be the least square estimators of (β_0, β') under the above null hypothesis, and let $\hat{\varepsilon}_t := \varepsilon_t(\hat{\beta}_0, \hat{\beta}) = X_t - \hat{\beta}_0 - c'_{n,t}\hat{\beta}$. The traditional Durbin-Watson test is based on the first-order residual autocorrelation

$$\hat{r}_{n1} := \sum_{t=2}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1} \Big/ \sum_{t=1}^{n} \hat{\varepsilon}_t^2.$$

When the errors are Gaussian with mean zero and variance σ^2, i.e. when $F(x) \equiv \Phi(x/\sigma)$, $n\hat{r}_{n1}$ coincides with

$$\begin{aligned}\frac{\sigma^2 \sum_{t=2}^{n} \Phi^{-1}(F(\hat{\varepsilon}_t))\Phi^{-1}(F(\hat{\varepsilon}_{t-1}))}{n^{-1}\sum_{t=1}^{n} \hat{\varepsilon}_t^2} &= \sum_{t=2}^{n} \Phi^{-1}(F(\hat{\varepsilon}_t))\Phi^{-1}(F(\hat{\varepsilon}_{t-1})) + o_P(n^{1/2}) \\ &= \sum_{t=2}^{n} \varphi_1(F(\varepsilon_t))\varphi_2(F(\varepsilon_{t-1})) + o_P(n^{-1/2}).\end{aligned}$$

where the last claim readily follows from Le Cam's Third Lemma with $\varphi_1 = \varphi_2 = \Phi^{-1}$. The aligned rank based version of $n\hat{r}_{n1}$ is the serial statistic $S_{n,\varphi_1\varphi_2}(\hat{\theta})$ defined in (1.2), with $i = 1$ and the van der Waerden scores $\varphi_1 = \varphi_2 = \Phi^{-1}$; an asymptotic representation result of Hallin, Ingenbleek and Puri (1985) establishes the equivalence $S_{n,\varphi_1\varphi_2}(\hat{\theta}) = T_{n,\varphi_1\varphi_2} + o_P(n^{1/2})$, where

$$T_{n,\varphi_1\varphi_2} := \sum_{t=2}^{n} \varphi_1(F(\varepsilon_t))\varphi_2(F(\varepsilon_{t-1})).$$

By Proposition 3.1 below it follows that the autoregression rank score statistic $\tilde{S}_{n,\varphi_1\varphi_2}$ of (1.4) is also asymptotically equivalent in probability to $T_{n,\varphi_1\varphi_2}$, under the above $\mathcal{H}_0$. An advantage of using $\tilde{S}_{n,\varphi_1\varphi_2}$ is that one does not need any preliminary estimates of the nuisance parameters.

In the case of non-Gaussian errors one uses the above serial autoregression rank score statistics with $\varphi_2(v) = F^{-1}(v)$ and $\varphi_1(u) = -\dot{f}(F^{-1}(u))/f(F^{-1}(u))$, to perform an asymptotically optimal test of $\tilde{\mathcal{H}}_0$, see, e.g., Hallin and Werker (1998).

2.2.2 AR order identification

The objective here is to test AR(p) against AR($p+1$) dependence. The overarching model is thus the AR($p+1$) model, where

$$X_t = \rho_0 + \sum_{i=1}^{p+1} \rho_i X_{t-i} + \varepsilon_t,$$

with $\rho_1, \cdots, \rho_{p+1}$ being such that the corresponding characteristic polynomial has all its roots outside the unit disc, $\rho_p \neq 0$, and $\varepsilon_1, \cdots, \varepsilon_n$ are i.i.d. with density f. The null hypothesis of interest here is $\mathcal{H}_0 : \rho_{p+1} = 0$, against the alternatives of the form $\mathcal{H}_1 : \rho_{p+1} \neq 0$. The autoregressive parameters $\rho_1, \cdots, \rho_p$ play the role of nuisance parameters.

The classical Gaussian test for this problem is based on a Lagrange multiplier type statistic, which is a quadratic form in the residual autocorrelations

$$\hat{r}_{ni} := \sum_{t=i+1}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-i} \Big/ \sum_{t=1}^{n} \hat{\varepsilon}_t^2, \quad i = 1, 2, \cdots ,$$

where the estimated residuals $\hat{\varepsilon}_t$ are computed from fitting an AR(p) model to the data: see Garel and Hallin (1999) for details. Arguing as in the previous example, a rank-based version of this test statistic is obtained by substituting the aligned serial rank statistics $(n-i)^{-1} S_{n,\varphi_1\varphi_2;i}(\hat{\theta})$'s of (1.2) for the residual autocorrelations $\hat{r}_{ni}$ into the quadratic test statistic. But such tests are not asymptotically distribution-free, while by Proposition 3.1, the tests based on the analogous quadratic form using serial autoregression rank score statistic $\tilde{S}_{n,\varphi_1\varphi_2;i}$ will be asymptotically distribution-free. Here again, asymptotically optimal tests at non-Gaussian errors case can be handled by an adequate choice of the scores φ_1 and φ_2.

Contrary to the previous case, $S_{n,\varphi_1\varphi_2}(\hat{\theta})$ and $\tilde{S}_{n,\varphi_1\varphi_2}$ are no longer asymptotically equivalent: $S_{n,\varphi_1\varphi_2}(\hat{\theta})$ suffers from an *alignment effect* (which is not distribution-free), whereas $\tilde{S}_{n,\varphi_1\varphi_2}$ remains unaffected. Hallin and Jurečková (1999) constructed asymptotically distributionfree tests of $\mathcal{H}_0$ against $\mathcal{H}_1$ based on non-serial autoregression rank score statistics of the type $\tilde{S}^*_{n,\varphi}(\theta)$'s. A simulation study of these tests can be found in Hallin et al. (1997) and an application to meteorological data in Kalvová et al. (2000).

2.2.3 Detection of random coefficients in AR models

The general overarching model is the autoregressive model (for simplicity, a first-order one) with random coefficients, of the form

$$X_t = (\rho + \tau u_t) X_{t-1} + \varepsilon_t,$$

where $\rho \in (0, 1)$, $\tau \geq 0$, $u_1, \cdots, u_n$ are i.i.d. standardized r.v.'s with density g, and $\varepsilon_1, \cdots, \varepsilon_n$ are i.i.d. with density f, independent of the u_t's. The null hypothesis of interest here is

$\mathcal{H}_0 : \tau = 0$ (ordinary AR(1) dependence), against the alternatives of the form $\mathcal{H}_1 : \tau > 0$. The autoregression parameter ρ and the densities g and f are nuisance parameters. Here $\Theta = \{\theta = (\rho, \tau)' \in (0,1) \times [0,1); \rho^2 + \tau^2 < 1\}$, $\Theta_0 = \{\theta = (\rho, 0)'; 0 < \rho < 1\}$, and $\varepsilon(\theta) = X_t - \rho X_{t-1}$, for a $\theta \in \Theta_0$.

This problem has been studied (in the more general AR(p) case) by Akharif and Hallin (2003), from a pseudo-Gaussian point of view. The locally asymptotically optimal Gaussian test statistic for this problem is the combination

$$(2.4) \quad \sum_{k=1}^{n-1} \hat{\rho}^{2(k-1)} (n-k)^{-1/2} \sum_{t=k+1}^{n} \Big(1 - \frac{\hat{\varepsilon}_t^2}{\hat{\sigma}^2}\Big)\Big(\frac{\hat{\varepsilon}_{t-k}}{\hat{\sigma}}\Big)^2 + 2 \sum_{1 \leq k} \sum_{< \ell \leq n-1} \hat{\rho}^{k-1} \hat{\rho}^{\ell-1} (n-\ell)^{-1/2} \sum_{t=\ell+1}^{n} \Big(1 - \frac{\hat{\varepsilon}_t^2}{\hat{\sigma}^2}\Big)\Big(\frac{\hat{\varepsilon}_{t-k}}{\hat{\sigma}}\Big)\Big(\frac{\hat{\varepsilon}_{t-\ell}}{\hat{\sigma}}\Big)$$

of the statistics of the form

$$(2.5) \quad (n-k)^{-1/2} \sum_{t=k+1}^{n} \Big(1 - \frac{\hat{\varepsilon}_t^2}{\hat{\sigma}^2}\Big)\Big(\frac{\hat{\varepsilon}_{t-k}}{\hat{\sigma}}\Big)^2 \quad \text{and} \quad (n-\ell)^{-1/2} \sum_{t=\ell+1}^{n} \Big(1 - \frac{\hat{\varepsilon}_t^2}{\hat{\sigma}^2}\Big)\Big(\frac{\hat{\varepsilon}_{t-k}}{\hat{\sigma}}\Big)\Big(\frac{\hat{\varepsilon}_{t-\ell}}{\hat{\sigma}}\Big),$$

where $\hat{\rho}$ is an arbitrary root-n consistent (under $\mathcal{H}_0$) of ρ, $\hat{\varepsilon}_t := X_t - \hat{\rho} X_{t-1}$, and $\hat{\sigma}^2 := n^{-1} \sum_{t=1}^{n} \hat{\varepsilon}_t^2$. Just as in the Durbin-Watson case, the diagonality of the information matrix (relative to ρ and σ^2) implies, via Le Cam's Third Lemma, that the impact of the estimation of ρ in (2.4) and (2.5) is $o_P(1)$ under $\mathcal{H}_0$. These statistics, under Gaussian assumptions ($F(x) = \Phi(x/\sigma)$) thus coincide, up to $o_P(1)$ terms, with

$$(n-k)^{-1/2} T_{\varphi_1, \varphi_{k;1}} := (n-k)^{-1/2} \sum_{t=k+1}^{n} \Big(1 - (\Phi^{-1}(F(\varepsilon_t)))^2\Big)\Big(\Phi^{-1}(F(\varepsilon_{t-k}))\Big)^2$$

and

$$(n-\ell)^{-1/2} T_{n,\varphi_1,\varphi_{k;2},\varphi_\ell} := (n-\ell)^{-1/2} \sum_{t=\ell+1}^{n} \Big(1 - (\Phi^{-1}(F(\varepsilon_t)))^2\Big) \Phi^{-1}(F(\varepsilon_{t-k})) \Phi^{-1}(F(\varepsilon_{t-\ell}),$$

respectively, with $\varphi_1(u) := 1 - (\Phi^{-1}(u))^2$, $\varphi_{k;1}(u) := (\Phi^{-1}(u))^2$, and $\varphi_{k;2}(u) = \varphi_\ell(u) := \Phi^{-1}(u)$. Asymptotic representation results for serial aligned rank statistics again imply the asymptotic equivalence, up to $o_P(n^{1/2})$, of $T_{n,\varphi_1,\varphi_{k;1}}$ and $T_{n,\varphi_1,\varphi_{k;2},\varphi_\ell}$ with the serial rank statistics (of the van der Waerden type)

$$S_{n,\varphi_1,\varphi_{k;1}}(\rho) := \sum_{t=k+1}^{n} \Big(1 - \Big(\Phi^{-1}\Big(\frac{R_{n,t}(\rho)}{n+1}\Big)\Big)^2\Big)\Big(\Phi^{-1}\Big(\frac{R_{n,t-k}(\rho)}{n+1}\Big)\Big)^2,$$

$$S_{n,\varphi_1,\varphi_{k;2},\varphi_\ell}(\rho) := \sum_{t=\ell+1}^{n} \Big(1 - \Big(\Phi^{-1}\Big(\frac{R_{n,t}(\rho)}{n+1}\Big)\Big)^2\Big) \Phi^{-1}\Big(\frac{R_{n,t-k}(\rho)}{n+1}\Big) \Phi^{-1}\Big(\frac{R_{n,t-\ell}(\rho)}{n+1}\Big),$$

respectively, where $R_t(\rho)$ is the rank of $\varepsilon_t(\rho) = X_t - \rho X_{t-1}$. These statistics, based on exact residual ranks, cannot be computed from the observations. However, in view of Proposition 3.1 below, $S_{n\varphi_1,\varphi_{k;1}}$ and $S_{n,\varphi_1,\varphi_{k;2},\varphi_\ell}$ in turn are asymptotically equivalent to their autoregression rank score counterparts $\tilde{S}_{n,\varphi_1,\varphi_{k;1}}$ and $\tilde{S}_{n,\varphi_1,\varphi_{k;2},\varphi_\ell}$, which are measurable with respect to the observations.

Perhaps it should be emphasized that serial autoregression rank scores based tests for a given choice of φ_j's can be always implemented regardless of the knowledge of the error density.

2.3 Assumptions on f and the score functions

We shall now state additional assumptions needed for obtaining the asymptotic representation result for serial autoregression rank scores. Besides the structural assumption (A), we also need some technical assumptions on the density f and the score functions $\varphi_1, \cdots, \varphi_m$. As usual, these assumptions cannot be separated: stronger assumptions on φ's allow for weaker assumptions on the densities, and vice-versa. Therefore, we formulate two sets of assumptions, (F1)-(F4), (φ-1) and (F1), (F5) and (φ-2), that can be used equivalently.

We assume that all densities f in the class $\mathcal{F}$ are such that

(F1) $\int_{-\infty}^{\infty} x dF(x) = 0, \qquad 0 < \int_{-\infty}^{\infty} x^2 dF(x) = \sigma^2 < \infty;$

(F2) The density f is positive on $\mathbb{R}$ and absolutely continuous, with a.e. derivative $\dot{f}$, satisfying $\mathcal{I}_f := \int_{-\infty}^{\infty} (\dot{f}(x)/f(x))^2 f(x) dx < \infty$.

(F3) There exists a constant $K = K_f \geq 0$ such that, for $|x| \geq K$, f has two bounded derivatives, f' and f'', respectively.

(F4) As $x \longrightarrow \pm\infty$, $f(x)$ is monotonically decreasing to 0 and,

$$\lim_{x \longrightarrow -\infty} \frac{-\log F(x)}{b|x|^r} = 1 = \lim_{x \longrightarrow \infty} \frac{-\log(1 - F(x))}{b|x|^r}$$

for some $b = b_f > 0$ and $r = r_f \geq 1$.

As for the functions $\varphi_1, \cdots, \varphi_m$, we assume the following:

(φ-1) The functions $\varphi_1, \cdots, \varphi_m$ from $(0,1)$ to $\mathbb{R}$ are square integrable, nondecreasing, differentiable, with respective derivatives $\dot{\varphi}_1, \cdots, \dot{\varphi}_m$, and satisfy

$$\int_0^1 \varphi_j(u) du = 0, \quad \text{for at least one } j = 1, \cdots, m,$$
$$|\dot{\varphi}_j| \leq C(u(1-u))^{-1-\delta}, \quad \forall j = 1, \cdots, m, \text{ for some } 0 < C < \infty \text{ and } 0 < \delta < 1/4.$$

The second set of assumptions consists of (F1) and the following two assumptions.

(F5) f is uniformly continuous and a.e. positive on $\mathbb{R}$.

(φ-2) The functions $\varphi_1, \cdots, \varphi_m$ from $(0,1)$ to $\mathbb{R}$ are nondecreasing bounded and $\int_0^1 \varphi_j(u)du = 0$, for some $j = 1, \cdots, m$.

The assumption (φ-2) rules out, for instance, the Wilcoxon and van der Waerden scores, as well as the so-called rank-based f-autocorrelation coefficients (Hallin and Puri 1994).

3 Asymptotic representation

The following main result of the paper gives the asymptotic representation of the serial autoregression rank score statistics. It enables one to construct the asymptotic rejection regions of the pertaining tests and their asymptotic powers against the Pitman alternatives. A similar result holds for serial regression rank scores of linear regression models with bounded designs.

Proposition 3.1 *Suppose the linear AR(p) model (2.1) holds. Suppose additionally either (F1)-(F4) and (φ-1) or (F1), (F5) and (φ-2) hold. Then, under $P_{n,\theta;f}$, as $n \to \infty$,*

$$\tilde{S}_{n,\varphi_1\cdots\varphi_m} = T_{n,\varphi_1\cdots\varphi_m} + o_P(n^{1/2}) = S_{n,\varphi_1\cdots\varphi_m}(\theta) + o_P(n^{1/2}), \tag{3.1}$$

where $\tilde{S}_{n,\varphi_1\cdots\varphi_m}$ is the serial autoregression rank score statistic (1.5), $S_{n,\varphi_1\cdots\varphi_m}(\theta)$ the serial rank statistic (1.3), and

$$T_{n,\varphi_1\cdots\varphi_m} := \sum_{t=i_{m-1}+1}^{n} \varphi_1(F(\varepsilon_{n,t}))\cdots\varphi_m(F(\varepsilon_{n,t-i_{m-1}})).$$

Proof. Without loss of generality, we restrict the proof to the autoregressive case for $m = 2$ and $i_1 = 1$, hence to statistics of the form $\tilde{S}_{n,\varphi_1\varphi_2}$, $T_{n,\varphi_1\varphi_2}$ and $S_{n,\varphi_1\varphi_2}$. Additionally, we shall assume the first set of conditions (F1)-(F4) and (φ-1) with the proviso that $\int \varphi_j = 0$, for both $j = 1, 2$. See Remark 3.1 below for the case when one of the φ's is not centered. The proof is much simpler under the second set of assumptions. We systematically drop subscripts n in the proof.

Let $0 < \alpha_0 < 1/2$ be a fixed number, $\alpha_n := n^{-1}(\log n)^2(\log\log n)^2$, and take n large enough so that $\alpha_n < \alpha_0$. Define, for a $0 < u < 1$,

$$\rho(u) := \rho + F^{-1}(u)e_1, \quad \text{with} \quad e_1' := (1, 0, \cdots, 0),$$

where $\rho := (1, \rho_1, \cdots, \rho_p)'$ of the model (2.1). For all $0 < u < 1$, put

$$\begin{aligned} D_t(u) &:= I(X_t \le y_{t-1}'\hat{\rho}(u)) - I(X_t \le y_{t-1}'\rho(u)), \\ \tilde{a}_t(u) &:= I(\varepsilon_t > F^{-1}(u)) - (1-u), \\ \hat{a}_t(u) - (1-u) &:= \tilde{a}_t(u) - D_t(u) + \hat{a}_t(u)I(X_t = y_{t-1}'\hat{\rho}(u)). \end{aligned} \tag{3.2}$$

Also, let

$$\hat{b}_{j,t} := -\int_0^1 \varphi_j(u)d\hat{a}_t(u), \quad \hat{b}_{n;j,t} := \int_{\alpha_n}^{1-\alpha_n} [\hat{a}_t(u) - (1-u)]d\varphi_j(u), \quad j = 1, 2.$$

Note that $\int_0^1 \varphi_j(u)du = 0$ and integration by parts yield that, for all t,

$$\hat{b}_{j,t} = -\int_0^1 \varphi_j(u)d[\hat{a}_t(u) - (1-u)] = \int_0^1 [\hat{a}_t(u) - (1-u)]d\varphi_j(u).$$

Decomposing this further gives, with $\bar{a}_t(u) := \hat{a}_t(u) - (1-u)$,

$$\hat{b}_{j,t} = \int_0^{\alpha_n} \bar{a}_t(u)d\varphi_j(u) + \hat{b}_{n;j,t} + \int_{1-\alpha_n}^1 \bar{a}_t(u)d\varphi_j(u),$$

$$\begin{aligned}\hat{b}_{n;j,t} &= \int_{\alpha_n}^{\alpha_0} \bar{a}_t(u)d\varphi_j(u) + \int_{\alpha_0}^{1-\alpha_0} \bar{a}_t(u)d\varphi_j(u) + \int_{1-\alpha_0}^{1-\alpha_n} \bar{a}_t(u)d\varphi_j(u) \\ &= \hat{c}_{n1;j,t} + \hat{c}_{n2;j,t} + \hat{c}_{n3;j,t}, \qquad \text{say.}\end{aligned}$$

We start with analyzing the sum $n^{-1/2}\sum_{t=2}^n \hat{c}_{n1;1,t}\hat{c}_{n1;2,t-1}$. The analysis of the similar sum involving $\hat{c}_{n3;j,t}$'s is exactly similar, while the similar sum corresponding to the $\hat{c}_{n2;j,t}$ terms can be analyzed using the results for bounded scores. The analysis of the cross product sums is also similar and relatively less involved. For the ease of writing, let $\varphi_{jn}(u) := \varphi_j(u)I(\alpha_n < u \le \alpha_0)$. Using (3.2), rewrite

$$\hat{c}_{n1;j,t} = \int \tilde{a}_t d\varphi_{jn} - \int D_t d\varphi_{jn} + \int \hat{a}_t(u)I(X_t = y'_{t-1}\hat{\rho}(u))d\varphi_{jn}(u).$$

Letting $A_{j,t} := \int \tilde{a}_t d\varphi_{jn}$, we thus obtain

$$\begin{aligned}(3.3)\quad & n^{-1/2}\sum_{t=2}^n \hat{c}_{n1;j,t}\hat{c}_{n1;j,t-1} \\ = {}& n^{-1/2}\sum_{t=2}^n \Big[A_{1,t}A_{2,t} - \int D_t d\varphi_{1n} A_{2,t} + \int \hat{a}_t(u)I\Big(X_t = y'_{t-1}\hat{\rho}(u)\Big)d\varphi_{1n}(u)A_{2,t} \\ & -A_{1,t}\int D_{t-1}d\varphi_{2n} + \int D_t d\varphi_{1n}\int D_{t-1}d\varphi_{2n} \\ & -\int \hat{a}_t(u)I(X_t = y'_{t-1}\hat{\rho}(u))d\varphi_{1n}(u)\int D_{t-1}d\varphi_{2n} \\ & +A_{1,t}\int \hat{a}_{t-1}(u)I\Big(X_{t-1} = y'_{t-2}\hat{\rho}(u)\Big)d\varphi_{2n}(u) \\ & -\int D_t d\varphi_{1n}\int \hat{a}_{t-1}(u)I\Big(X_{t-1} = y'_{t-2}\hat{\rho}(u)\Big)d\varphi_{2n}(u) \\ & +\int \hat{a}_t(u)I\Big(X_t = y'_{t-1}\hat{\rho}(u)\Big)d\varphi_{1n}(u)\int \hat{a}_{t-1}(u)I\Big(X_{t-1} = y'_{t-2}\hat{\rho}(u)\Big)d\varphi_{2n}(u)\Big] \\ = {}& C_1 - C_2 + C_3 - C_4 + C_5 - C_6 + C_7 - C_8 + C_9, \qquad \text{say.}\end{aligned}$$

In order to show that the first term $C_1 := n^{-1/2}\sum_{t=2}^n A_{1,t}A_{2,t}$ provides the approximating terms to the left hand side above, we shall verify that all of the remaining terms tend to zero in probability. Let $d_{jn} := \varphi_j(\alpha_0) - \varphi_j(\alpha_n)$, $j = 1, 2$, and $d_n := \max(d_{1n}, d_{2n})$. From the linear programming definition of $\hat{a}_t(u)$'s, see e.g., Hallin and Jurčková (1999), we obtain that for all $0 < u < 1$,

$$\sum_{t=2}^{n} \hat{a}_t(u) I\Big(X_t = y'_{t-1}\hat{\rho}(u)\Big) \leq (p+1), \qquad \text{a.s.} \tag{3.4}$$

This in turn implies

$$\begin{aligned} n^{-1/2}\sum_{t=2}^{n}\int \hat{a}_t(u) I\Big(X_t = y'_{t-1}\hat{\rho}(u)\Big) d\varphi_{1n}(u) &\leq n^{-1/2}(p+1)d_n, \qquad \text{a.s} \\ n^{-1/2}\sum_{t=2}^{n}\int \hat{a}_{t-1}(u) I\Big(X_{t-1} = y'_{t-2}\hat{\rho}(u)\Big) d\varphi_{2n}(u) &\leq n^{-1/2}(p+1)d_n, \quad \text{a.s.} \end{aligned} \tag{3.5}$$

Now, consider the term C_9. The fact that $\hat{a}_t \leq 1$ and (3.5) imply

$$C_9 \leq d_n^2 n^{-1/2}(p+1), \quad \text{a.s.} \tag{3.6}$$

Similarly using the fact that $|D_t| \leq 1$, for all t and (3.5), we obtain

$$|C_8| \;\leq\; d_n^2 n^{-1/2}(p+1), \qquad \text{a.s.} \tag{3.7}$$

The same argument and the fact that $|A_{j,t}| \leq d_n$, for all t, imply that

$$\max\{|C_3|,\, |C_6|,\, |C_7|\} \leq d_n^2 n^{-1/2}(p+1), \qquad \text{a.s.} \tag{3.8}$$

The assumption (φ-1) on φ_1, φ_2 and the definition of α_n imply that

$$\begin{aligned} d_n &\leq \int_{\alpha_n}^{\alpha_0} (u(1-u))^{-1-\delta} du \leq \int_{\alpha_n}^{\alpha_0} u^{-1-\delta} du \\ &\leq \frac{C}{\delta}\,(\alpha_n^{-\delta} - \alpha_0^{-\delta}) = O(n^{\delta}(\log\, n)^{-2\delta}(\log\,\log\, n)^{-2\delta}), \end{aligned}$$

so that, because $0 < \delta < 1/4$, $d_n^2 n^{-1/2} = o(1)$. Hence,

$$\max\{|C_3|,\, |C_6|,\, |C_7|, |C_8|, |C_9|\} = o(1), \qquad \text{a.s.} \tag{3.9}$$

Note also that, because of the $n^{-1/2}$ factor, the same conclusions hold if φ_j, $j = 1, 2$, is replaced by $\tilde{\varphi}_{jn} := \varphi_j I\Big[[\alpha_n, \alpha_n(1+\epsilon)]\Big]$.

To deal with C_4, we need to center the factor involving D_{t-1} properly. For this the r.v.'s involved in the indicators need to be suitably standardized. This standardization is done differently for the u-quantiles in the tail and in the middle, because in the tail the consistency rate of $\hat\rho(u)$ is different from $n^{-1/2}$ and also depends on u, as was shown in Hallin and Jurečková (1999). We need to use these facts in the following analysis. Accordingly, let

$$q(u) \;:=\; f(F^{-1}(u)), \quad \sigma_u := (u(1-u))^{1/2}/q(u), \quad \Delta(u) := \sigma_u^{-1} n^{1/2}(\hat\rho(u) - \rho(u)).$$

Rewrite

$$D_t(u) = I(\varepsilon_t \le F^{-1}(u) + n^{-1/2}\sigma_u y'_{t-1}\Delta(u)) - I(\varepsilon_t \le F^{-1}(u)).$$

Its centering involves

$$\begin{aligned} \mu_t(u) &:= F\Big(F^{-1}(u) + \sigma_u n^{-1/2} y'_{t-1}\Delta(u)\Big) - F\Big(F^{-1}(u)\Big), \\ \nu_t(u) &:= \mu_t(u) - \sigma_u n^{-1/2} y'_{t-1}\Delta(u)\, q(u). \end{aligned}$$

Then,

$$\begin{aligned} C_4 &:= n^{-1/2}\sum_{t=2}^n A_{1,t}\int D_{t-1}d\varphi_{2n} \\ &= n^{-1/2}\sum_{t=2}^n A_{1,t}\int [D_{t-1} - \mu_{t-1}]d\varphi_{2n} + n^{-1/2}\sum_{t=2}^n A_{1,t}\int \nu_{t-1}d\varphi_{2n} \\ &\qquad + n^{-1}\sum_{t=2}^n A_{1,t}y'_{t-2}\int \Delta(u)\,\sigma_u q(u) d\varphi_{2n}(u) \\ &= C_{41} + C_{42} + C_{43}, \qquad \text{say.} \end{aligned} \tag{3.10}$$

But, because $\sigma_u q(u) = (u(1-u))^{1/2}$,

$$\begin{aligned} |C_{43}| &\le \|n^{-1}\sum_{t=2}^n A_{1,t}y_{t-2}\|\,\|\int \Delta(u)\sigma_u q(u)d\varphi_{2n}(u)\| \\ &= O_P(n^{-1/2}) \sup_{\alpha_n < u \le \alpha_0}\|\Delta(u)\|\int (u(1-u))^{1/2}d\varphi_{2n}(u). \end{aligned}$$

The first factor of $O_P(n^{-1/2})$ comes from the fact that $\sum_{t=2}^n A_{1,t}y_{t-2}$ is a vector of zero mean martingales, and hence

$$E\|n^{-1}\sum_{t=2}^n A_{1,t}y_{t-2}\|^2 \;=\; n^{-2}\sum_{t=2}^n Ey'_{t-2}y_{t-2}\,EA_{1,t}^2 = O(n^{-1}). \tag{3.11}$$

Also, by (φ-1), $\int (u(1-u))^{1/2}d\varphi_{jn}(u) \le \int_0^{\alpha_0} u^{-1/2-\delta}du < \infty$, $j = 1, 2$. Moreover, recall from Hallin and Jurečková (1999) that under (F1)-(F4),

$$\sup_{\alpha_n \le u \le 1-\alpha_n}\|\Delta(u)\| = O_P(\log\log n)^{1/2}. \tag{3.12}$$

Upon combining these observations we obtain

$$|C_{43}| \quad = \quad O_P(n^{-1/2}(\log\log n)^{1/2}) = o_P(1). \tag{3.13}$$

Next consider C_{42}. Let $\delta_{n,t,u} := n^{-1/2}\sigma_u\, y'_{t-2}\Delta(u)$, $\epsilon_n := C\,(\log n)^{2/r-2}\,(\log\log n)^{-1/4}$, $r \geq 1$, and $K_n := C\,(\log\log n)^{1/2}$. We need the following results from Hallin and Jurečková (1999) obtained under (F1)-(F4). By (A.5) and (A.9) in there, for any $r \geq 1$,

$$\max_{1\leq t\leq n, \alpha_n\leq u\leq 1-\alpha_n} |\delta_{n,t,u}| = O_P(\epsilon_n), \tag{3.14}$$

$$\frac{|\dot{f}(x)|}{f(x)}|x|^{1-r} = O(1), \quad \text{as } x \to \pm\infty. \tag{3.15}$$

Let

$$\tau_y \quad = \quad \sigma_{F(y)}, \quad \tilde{\Delta}_y = \Delta(F(y)), \quad \tilde{\delta}_{n,t,y} = \delta_{n,t,F(y)}, \quad x_n = F^{-1}(\alpha_n), \quad x_0 = F^{-1}(\alpha_0).$$

Since $\alpha_n < \alpha_0 < 1/2$, we have $x_n < x_0 < 0$. Also, in the left tail $|d\varphi_j(F)| \leq C\,F^{-1-\delta}dF$. Let $\mathcal{A}_n := \{\sup_{\alpha_n\leq u\leq 1-\alpha_n} \|\Delta(u)\| \leq K_n\}$. For $x_n \leq y \leq x_0 < 0$, on the event $\mathcal{A}_n$,

$$|\tilde{\delta}_{n,t,y}| \leq n^{-1/2}\frac{F^{1/2}(y)}{f(y)}\|y_{t-2}\|K_n,$$

and

$$\begin{aligned}
|C_{42}| \quad &= \quad |n^{-1/2}\sum_{t=2}^{n} A_{1,t}\int \nu_{t-1}(u)d\varphi_{2n}(u)| \\
&\leq \quad n^{-1/2}\sum_{t=2}^{n}|A_{1,t}|\Big|\int [F(y+\frac{\tau_y y'_{t-2}\tilde{\Delta}_y}{\sqrt{n}}) - F(y) - \frac{\tau_y y'_{t-2}\tilde{\Delta}_y}{\sqrt{n}}f(y)]d\varphi_{2n}(F(y))\Big| \\
&\leq \quad n^{-1}\sum_{t=2}^{n}|A_{1,t}|\|y_{t-2}\|K_n\int_{x_n}^{x_0}\int_0^1 \frac{|f(y+v\,\tilde{\delta}_{n,t,y}) - f(y)|}{f(y)}\,dvF^{-\frac{1}{2}-\delta}(y)dF(y).
\end{aligned}$$

But, on the event $\max_{1\leq t\leq n, x_n\leq y\leq x_0}|\tilde{\delta}_{n,t,y}| \leq C\epsilon_n$, see (3.14), the double integral in this bound is further bounded above by $C\epsilon_n$ times the integral

$$\begin{aligned}
&\max_{1\leq t\leq n}\int_{x_n}^{x_0}\int_0^1\int_0^1 v\frac{|\dot{f}(y+\tilde{\delta}_{n,t,y}vz)|}{f(y)}\,dvdzF^{-\frac{1}{2}-\delta}(y)dF(y) \\
&\leq \max_{1\leq t\leq n}\int_0^1 v\int_0^1\int_{x_n}^{x_0}\frac{|\dot{f}(y+\tilde{\delta}_{n,t,y}vz)|}{f(y)}\,F^{-\frac{1}{2}-\delta}(y)dF(y)\,dzdv \\
&\longrightarrow_P (1/2)\int_{-\infty}^{x_0}\frac{|\dot{f}(y)|}{f(y)}\,F^{-\frac{1}{2}-\delta}(y)dF(y) \\
&\leq C\int_{-\infty}^{x_0}|y|^{r-1}F^{-\frac{1}{2}-\delta}(y)dF(y) \qquad\qquad \text{(by (3.15)} \\
&\leq C\Big(\int |y|^{p(r-1)}dF(y)\Big)^{1/p}\Big(\int F^{-q(\frac{1}{2}+\delta)}(y)dF(y)\Big)^{1/q},
\end{aligned} \tag{3.16}$$

where p, q are positive integers such that $1/p + 1/q = 1$, and such that $1 > q(1/2+\delta)$ so that the second integral in the above bound is finite. Such a q always exists since $\delta < 1$. Also note that because $r > 1$, we have $(2/r) - 2 < 0$, and hence

$$K_n \epsilon_n = C\,(\log\, n)^{\frac{2}{r}-2}(\log\log\, n)^{1/4} = o(1), \qquad n^{-1}\sum_{t=2}^{n} |A_{1,t}|\|y_{t-2}\| = O_P(1),$$

and, in view of (3.14) and (3.15),

$$|C_{42}| = O_P(K_n \epsilon_n) = o_P(1). \tag{3.17}$$

Next, we treat the C_{41} term. Let $a_n = n^{-1/2}K_n$. Define, for $y,\, a \in \mathbb{R}$ and $s \in \mathbb{R}^{m+1}$ such that $\|s\| \leq 1$,

$$\begin{aligned} m_{t-1}(y,s,a) &:= [F(y + a_n\tau_y(y'_{t-2}s + \|y_{t-2}\|a)) - F(y)], \\ C^{\pm}_{41}(y,s,a) &:= n^{-1/2}\sum_{t=2}^{n} A^{\pm}_{1,t}[I(\varepsilon_{t-1} \leq y + a_n\tau_y(y'_{t-2}s + \|y_{t-2}\|a)) - I(\varepsilon_{t-1} \leq y) \\ &\qquad\qquad - m_{t-1}(y,s,a)], \end{aligned}$$

where $A_t\pm$ stand for the positive and negative parts of A_t. Write $C^{\pm}_{41}(y,s)$ for $C^{\pm}_{41}(y,s,0)$ and let

$$\begin{aligned} C_{41}(y,s) &:= C^{+}_{41}(y,s) - C^{-}_{41}(y,s) \\ &= n^{-1/2}\sum_{t=2}^{n} A_{1,t}[I(\varepsilon_{t-1} \leq y + a_n\tau_y y'_{t-2}s) - I(\varepsilon_{t-1} \leq y) - \mu_{t-1}(F(y),s)]. \end{aligned}$$

Note that on the event $\mathcal{A}_n$,

$$\begin{aligned} |C_{41}| &\leq \int \sup_{\|s\|\leq 1} |C_{41}(y,s)|d\psi_{2n}(y) \\ &\leq \int \sup_{\|s\|\leq 1} |C^{+}_{41}(y,s)|d\psi_{2n}(y) + \int \sup_{\|s\|\leq 1} |C^{-}_{41}(y,s)|d\psi_{2n}(y), \end{aligned}$$

with $\psi_{jn}(y) := \varphi_{jn}(F(y))$. We shall show that

$$\int \sup_{\|s\|\leq 1} |C^{\pm}_{41}(y,s)|d\psi_{2n}(y) = o_P(1), \tag{3.18}$$

which obviously will imply

$$|C_{41}| = o_P(1). \tag{3.19}$$

Let

$$\begin{aligned} \ell_n(y,s,a) &:= \int_0^1 \frac{E\{\|y_0\|f(y + z\,\tau_y a_n(y'_0 s + \|y_0\|a))\}}{f(y)}dz, \\ \gamma_n(y,s,a) &:= \int_0^1 \frac{E\{\|y_0\|f(y + \tau_y a_n(y'_0 s - 2\|y_0\|a\,z))\}}{f(y)}dz, \quad s \in \mathbb{R}^{m+1},\ a,\, y \in \mathbb{R}. \end{aligned}$$

Arguing as above and conditionally, we have for all $y \in \mathbb{R}$, with $b = EA_{1,t}^2$,

$$
\begin{aligned}
(3.20)\quad E|C_{41}^{\pm}(y,s,a)|^2 &= E(A_{1,t}^{\pm})^2 E\Big[I(\varepsilon_1 \le y + a_n\tau_y(y_0's + \|y_0\|a)) - I(\varepsilon_1 \le y) \\
&\qquad -F(y + a_n\tau_y(y_0's + \|y_0\|a)) + F(y)\Big]^2 \\
&\le b\, E|F(y + a_n\tau_y(y_0's + \|y_0\|a)) - F(y)| \\
&\le b\, a_n\tau_y(\|s\| + a) \int_0^1 E\{\|y_0\| f(y + z\,\tau_y a_n(y_0's + \|y_0\|a))\}dz \\
&= b\, n^{-1/2}K_n\,(\|s\| + |a|)\,[F(y)(1-F(y))]^{1/2}\,\ell_n(y,s,a).
\end{aligned}
$$

Similarly, for any $s,\, t \in \mathbb{R}^p$, and $y,\, a \in \mathbb{R}$,

$$
\begin{aligned}
& E|C_{41}^{\pm}(y,t,a) - C_{41}^{\pm}(y,s,a)|^2 \\
&\quad \le\; b\,\|t-s\|\, n^{-1/2}K_n\,[F(y)(1-F(y))]^{1/2}\,\ell_n(y,s,a), \\
& E|C_{41}^{\pm}(y,t,a) - C_{41}^{\pm}(y,t,0)|^2 \\
(3.21)\quad &\quad \le\; b\,|a|\, n^{-1/2}K_n\,[F(y)(1-F(y))]^{1/2}\,\gamma_n(y,t,a).
\end{aligned}
$$

Since the unit ball is compact, there is an $\eta > 0$ and a finite number k of points $s_1, \cdots, s_k$ in the unit ball such that for any $\|s\| \le 1$, there is an s_j in the unit ball with $\|s - s_j\| \le \eta$. We will need to choose η to depend on n and hence so also k. Now,

$$
(3.22)\quad \sup_{\|s\|\le 1} |C_{41}^{\pm}(y,s)| \;\le\; \max_{1\le j\le k}\sup_{\|s-s_j\|\le\eta} |C_{41}^{\pm}(y,s) - C_{41}^{\pm}(y,s_j)| + \max_{1\le j\le k}|C_{41}^{\pm}(y,s_j)|.
$$

Now, $\|s - s_j\| \le \eta$ implies that for all $y \in \mathbb{R}, 1 \le j \le k,\; n \ge 1,\; 1 \le t \le n$,

$$
a_n\tau_y(y_{t-2}'s_j - \|y_{t-2}\|\eta) \le a_n\tau_y y_{t-2}'s \le a_n\tau_y(y_{t-2}'s_j + \|y_{t-2}\|\eta).
$$

This and the monotonicity of the indicators and nonnegativity of $A_{1,t}^{\pm}$'s, imply

$$
\begin{aligned}
& |C_{41}^{\pm}(y,s) - C_{41}^{\pm}(y,s_j)| \\
&\quad \le\; |C_{41}^{\pm}(y,s_j,\eta) - C_{41}^{\pm}(y,s_j,0)| + |C_{41}^{\pm}(y,s_j,-\eta) - C_{41}^{\pm}(y,s_j,0)| \\
(3.23)\quad &\qquad +2n^{-1/2}\sum_{t=2}^{n} A_{1,t}^{\pm}[m_{t-1}(y,s_j,\eta) - m_{t-1}(y,s_j,-\eta)]
\end{aligned}
$$

Moreover, by (3.21), and the Cauchy-Schwarz inequality,

$$
\begin{aligned}
& E\Big(\max_{1\le j\le k}|C_{41}^{\pm}(y,s_j,\pm\eta) - C_{41}^{\pm}(y,s_j,0)|\Big) \\
&\quad \le\; k\,\{2b\,\eta\, n^{-1/2}K_n\,[F(y)(1-F(y))]^{1/2}\,\max_{1\le j\le k}\gamma_n(y,s_j,\eta)\}^{1/2}.
\end{aligned}
$$

Let $g_n(y,\eta) := \max_{1\le j\le k}\gamma_n(y,s_j,\eta)$. Consider the measure $\nu(y) := \int_{-\infty}^{y} F^{-3/4-\delta}dF$. Note that for $\delta < 1/4$ this is a finite measure. Also note that $g_n(y,\eta)\to E\|y_0\| < \infty$. Arguing as for (3.16), we thus obtain for all η,

$$\begin{aligned} B_n := \int [F(y)(1-F(y))]^{1/4}\, g_n(y,\eta)d\psi_n(y) \;&\le\; \int_{x_n}^{x_0} F(y)^{-3/4-\delta} g_n^{1/2}(y,\eta)dF(y) \\ &\le\; \int g_n^{1/2}(y,\eta)d\nu(y) = O(1). \end{aligned}$$

Thus,

$$\int E\Big(\max_{1\le j\le k}|C_{41}^{\pm}(y,s_j,\pm\eta) - C_{41}^{\pm}(y,s_j,0)|\Big)d\psi_{2n}(y) \;\le\; B_n\, k\,\eta^{1/2}\,(n^{-1/2}K_n)^{1/2}. \tag{3.24}$$

Next, let $d_{n,t,y,s} := a_n\tau_y y_{t-1}' s$. The third term in the upper bound of (3.23) is bounded above by

$$\begin{aligned} &\frac{1}{\sqrt{n}}\sum_{t=2}^{n} A_{1,t}^{\pm}\Big\{F(y+d_{n,t-1,y,s_j}+a_n\tau_y\|y_{t-2}\|\eta) - F(y) - [d_{n,t-1,y,s_j}+a_n\tau_y\|y_{t-2}\|\eta]f(y)\Big\} \\ &-\frac{1}{\sqrt{n}}\sum_{t=2}^{n} A_{1,t}^{\pm}\Big\{F(y+d_{n,t-1,y,s_j}-a_n\tau_y\|y_{t-2}\|\eta) - F(y) - [d_{n,t-1,y,s_j}-a_n\tau_y\|y_{t-2}\|\eta]f(y)\Big\} \\ &+2\eta\,\frac{1}{\sqrt{n}}a_n\tau_y\sum_{t=2}^{n} A_{1,t}^{\pm}\|y_{t-2}\|f(y) \\ &= M_{1,j}(y) - M_{2,j}(y) + 2\eta K_n[F(y(1-F(y))]^{1/2}n^{-1}\sum_{t=2}^{n} A_{1,t}^{\pm}\|y_{t-2}\|. \end{aligned}$$

Note that for all $\|s_j\|\le 1$, and all t, and all y with $F(y)$ close to α_n, on the event $\{\|y_{t-2}\|\le C(\log\, n)^{1/r}(\log\log\, n)^{1/4}\}$,

$$[d_{n,t-1,y,s_j}+a_n\tau_y\|y_{t-2}\|\eta] \;\le\; a_n\tau_y\|y_{t-2}\|(1+\eta) \le \epsilon_n(1+\eta).$$

Hence, by arguing as for (3.16),

$$\begin{aligned} &\int \max_{1\le j\le k}|M_{1,j}(y)|d\psi_{2n}(y) \\ &\le\; \int n^{-1/2}\sum_{t=2}^{n}|A_{1,t}|[d_{n,t-1,y,s_j}+a_n\tau_y\|y_{t-2}\|\eta] \\ &\qquad\qquad \times\max_{1\le j\le k}\int_0^1 |f(y+[d_{n,t-1,y,s_j}+a_n\tau_y\|y_{t-2}\|\eta]z) - f(y)|dz\, d\psi_{2n}(y) \\ &\le\; K_n n^{-1}\sum_{t=2}^{n}|A_{1,t}|\|y_{t-2}\|(1+\eta) \\ &\quad\;\times\int \max_{1\le j\le k}[d_{n,t-1,y,s_j}+a_n\tau_y\|y_{t-2}\|\eta] \end{aligned}$$

$$
\begin{aligned}
&\times \int_0^1 z \int_0^1 \frac{|\dot{f}(y + [d_{n,t-1,y,s_j} + a_n \tau_y \|y_{t-2}\| \eta] z\, v)| dv dz}{f(y)} F^{-1/2-\delta} dF(y) \\
\leq\ & K_n \epsilon_n \,(1+\eta)^2\, n^{-1} \sum_{t=2}^{n} |A_{1,t}| \|y_{t-2}\|^2 \\
&\times \int \frac{\max_{1\leq j\leq k} \int_0^1 z \int_0^1 |\dot{f}(y + [d_{n,t-1,y,s_j} + a_n \tau_y \|y_{t-2}\| \eta] z\, v)| dv dz}{f(y)} F^{-1/2-\delta} dF(y) \\
=\ & O_P(K_n \epsilon_n) = o_P(1), \quad \forall\, \eta > 0.
\end{aligned}
$$

Similarly,

$$
\int \max_{1\leq j\leq k} |M_{2,j}(y)| d\psi_{2n}(y) = O_P(K_n \epsilon_n) = o_P(1), \quad \forall\, \eta > 0.
$$

Thus we obtain that the integral of the maximum over $1 \leq j \leq k$ of the third term in the bound (3.23) is bounded above by

$$
\begin{aligned}
(3.25) \qquad & O_P(K_n \epsilon_n (1+\eta)^2) + 2 K_n \eta n^{-1} \sum_{t=2}^{n} |A_{1,t}| \|y_{t-2}\| \int F(y)^{-1/2-\delta} dF(y) \\
& = O_P(\eta\, K_n) + O_P(K_n \epsilon_n).
\end{aligned}
$$

Next, let $L_n(y) := \max_{1\leq j\leq k} \ell_n(y, s_j, 0)$. By (3.20) applied with $a = 0$, and by the Cauchy-Schwarz inequality

$$
\begin{aligned}
E\Big(\max_{1\leq j\leq k} |C_{41}^{\pm}(y, s_j)| \Big) &\leq \sum_{j=1}^{k} E \Big| C_{41}^{\pm}(y, s_j) \Big| \\
&\leq k \Big\{ b\, n^{-1/2} K_n \, [F(y)(1-F(y))]^{1/2} \sum_{j=1}^{k} \|s_j\| \ell_n(y, s_j, 0) \Big\}^{1/2} \\
&\leq k \Big\{ b\, n^{-1/2} K_n \, [F(y)(1-F(y))]^{1/2} L_n(y) \Big\}^{1/2},
\end{aligned}
$$

so that

$$
\int E\Big(\max_{1\leq j\leq k} |C_{41}^{\pm}(y, s_j)| \Big) d\psi_{2n}(y) \quad \leq \quad C\, k\, n^{-1/4} C_n^{1/2} \int L_n(y) F^{-3/4-\delta}(y) dF(y).
$$

Putting all these bounds together with (3.22), (3.23) and (3.24), we obtain that

$$
\begin{aligned}
\int \sup_{\|s\|\leq 1} |C_{41}^{\pm}(y, s)| d\psi_{2n}(y) &= O_P(k\, \eta^{1/2}\, n^{-1/4} K_n^{1/2}) + O_P(\eta\, K_n) + O_P(\epsilon_n\, K_n), \\
&= O_P(\eta^{1/2-p} n^{-1/4} K_n^{1/2}) + O_P(\eta\, K_n).
\end{aligned}
$$

This in turn implies (3.18), by choosing η suitably. For example $\eta = K_n^{-a}$, $a > 1$, will suffice. The results (3.19), (3.17) and (3.13) together imply

$$
(3.26) \qquad\qquad C_4 = o_P(1).
$$

Similarly one can prove

$$C_2 = o_P(1). \tag{3.27}$$

Next, consider

$$\begin{aligned}
|C_5| &:= n^{-1/2}\Big|\sum_{t=2}^{n}\int D_t d\varphi_{1n}\int D_{t-1}d\varphi_{2n}\Big| \\
&\le \int\int \sup_{\|s\|\le 1} n^{-1/2}\sum_{t=2}^{n}\Big\{|I(\varepsilon_t\le x+\tau_x a_n y_{t-1}'s)-I(\varepsilon_t\le x)| \\
&\qquad\qquad \times|I(\varepsilon_{t-1}\le y+\tau_y a_n y_{t-2}'s)-I(\varepsilon_{t-1}\le y)|\Big\}d\psi_{1n}(x)d\psi_{2n}(y) \\
&\le \int\int n^{-1/2}\sum_{t=2}^{n}\Big\{I(x-\tau_x a_n\|y_{t-1}\|<\varepsilon_t\le x+\tau_x a_n\|y_{t-1}\|) \\
&\qquad\qquad \times I(y-\tau_y a_n\|y_{t-2}\|<\varepsilon_{t-1}\le y+\tau_y a_n\|y_{t-2}\|)\Big\}d\psi_{1n}(x)d\psi_{2n}(y)
\end{aligned}$$

so that, by a conditioning argument,

$$\begin{aligned}
E|C_5| &\le \int\int n^{1/2}E\{I(y-\tau_y a_n\|y_0\|<\varepsilon_1\le y+\tau_y a_n\|y_0\|) \\
&\qquad \times|F(x+\tau_x a_n\|Y_1\|)-F(x-\tau_x a_n\|Y_1\|)|\}d\psi_{1n}(x)d\psi_{2n}(y) \\
&\le K_n\int\int_{y=x_n}^{y=x_0} E\Big\{I(y-\tau_y a_n\|y_0\|<\varepsilon_1\le y+\tau_y a_n\|y_0\|) \\
&\qquad \times\|Y_1\|\frac{\int_{-1}^{1}f(x+\tau_x a_n\|Y_1\|\,v)dv}{f(x)}\Big\}F^{-1/2-\delta}(x)dF(x)d\psi_{2n}(y) \\
&\le K_n\int\int_{y=x_n}^{y=x_0} E^{1/2}|F(y+\tau_y a_n\|y_0\|)-F(y-\tau_y a_n\|y_0\|)| \\
&\qquad \times E^{1/2}\Big\{\|Y_1\|\frac{\int_{-1}^{1}f(x+\tau_x a_n\|Y_1\|\,v)dv}{f(x)}\Big\}^2F^{-1/2-\delta}(x)dF(x)d\psi_{2n}(y) \\
&\le K_n\int\int_{y=x_n}^{y=x_0} E^{1/2}\tau_y a_n\|y_0\|\int_{-1}^{1}f(y+\tau_y a_n\|y_0\|v)dv \\
&\qquad \times E^{1/2}\Big\{\|Y_1\|\frac{\int_{-1}^{1}f(x+\tau_x a_n\|Y_1\|\,v)dv}{f(x)}\Big\}^2F^{-1/2-\delta}(x)dF(x)d\psi_{2n}(y) \\
&\le n^{-1/4}C_n^2\int_{y=x_n}^{y=x_0} E^{1/2}\Big\{\|y_0\|\frac{\int_{-1}^{1}f(y+\tau_y a_n\|y_0\|v)dv}{f(y)}\Big\}F^{-1/2-\delta}(y)dF(y) \\
&\qquad \times\int_{x_n}^{x_0}E^{1/2}\Big\{\|Y_1\|\frac{\int_{-1}^{1}f(x+\tau_x a_n\|Y_1\|\,v)dv}{f(x)}\Big\}^2F^{-1/2-\delta}(x)dF(x) \\
&= o(1).
\end{aligned}$$

The above bounds clearly prove the following

Lemma 3.1 *Under the conditions of Proposition 3.1, we have*

$$\begin{aligned} n^{-1/2}&\sum_{t=2}^{n}\hat{c}_{n1;1,t}\hat{c}_{n1;2,t-1} \\ &= n^{-1/2}\sum_{t=2}^{n}\int\left[I\left(\varepsilon_t>F^{-1}(u)\right)-(1-u)\right]d\varphi_{n1}(u) \\ &\qquad\qquad\times\int\left[I\left(\varepsilon_{t-1}>F^{-1}(v)\right)-(1-v)\right]d\varphi_{n2}(v)+o_P(1). \end{aligned}$$

Next, consider the sum $n^{-1/2}\sum_{t=2}^{n}\hat{c}_{n2;1,t}\hat{c}_{n2;2,t-1}$. Note that this is similar to the above sum with φ_{jn} replaced with $\varphi_j^0 := \varphi_j(u)I(\alpha_0 \le u \le 1-\alpha_0)$. Thus several calculations are similar to those in the case considered in the proof of Lemma 3.1. To begin with, the decomposition (3.3) remains valid with φ_{jn} replaced by φ_j^0. Denote these terms by C_i^0, $i = 1, \cdots, 9$. That is, C_i^0 is the C_i of (3.3) with φ_j^0 substituted for φ_{jn}. The bounds (3.6), (3.7), (3.8) now hold with d_{jn} replaced by $d_j^0 := \varphi_j(1-\alpha_0)-\varphi_j(\alpha_0)$, so that the analog of (3.9) clearly holds here. The places where one uses a different argument is in the handling of the remaining terms C_1^0, C_2^0, C_4^0 and C_5^0.

To begin with consider C_4^0. As mentioned earlier, here one needs to standardize the random variables involved in the indicators of D_t differently. Accordingly, now let $\gamma(u) := n^{1/2}(\hat{\rho}(u)-\rho(u))$, and rewrite

$$\begin{aligned} \mu_t(u) &:= F\Big(F^{-1}(u)+n^{-1/2}y_{t-1}'\gamma(u)\Big)-F\Big(F^{-1}(u)\Big), \\ \nu_t(u) &:= \mu_t(u)-n^{-1/2}y_{t-1}'\gamma(u)\,q(u). \end{aligned}$$

Then,

$$\begin{aligned} C_4^0 &:= n^{-1/2}\sum_{t=2}^{n}A_{1,t}\int D_{t-1}d\varphi_2^0 \\ &= n^{-1/2}\sum_{t=2}^{n}A_{1,t}\Big[\int[D_{t-1}-\mu_{t-1}]d\varphi_2^0+\int\nu_{t-1}d\varphi_2^0+\int n^{-1/2}y_{t-2}'\gamma(u)\,d\varphi_2^0(u)\Big] \\ &= C_{41}^0+C_{42}^0+C_{43}^0, \qquad \text{say.} \end{aligned}$$

Now recall from Koul and Saleh (1995) that

$$\sup_{\alpha_0\le u\le 1-\alpha_0}\|\gamma(u)\| = O_P(1). \tag{3.28}$$

Using this, the fact φ_2^0 is bounded and arguing as for (3.13), we have

$$|C_{43}^0| \;\le\; \|n^{-1}\sum_{t=2}^{n}A_{1,t}y_{t-2}\|\sup_{\alpha_0\le u\le 1-\alpha_0}\|\gamma(u)\| = O_P(n^{-1/2}) = o_P(1).$$

Next, consider C_{42}^0. Let

$$\psi_j^0(y) := \varphi_j^0(F(y)),\ \ a_0 = F^{-1}(\alpha_0),\ \ a_1 = F^{-1}(1-\alpha_0),\ \ \tilde\gamma_y = \gamma(F(y)), \quad y \in \mathbb{R},$$
$$\zeta_n := \sup_{1\le t\le n, a_0\le y\le a_1} n^{-1/2}|y'_{t-2}\tilde\gamma_y|.$$

The stationarity of the time series, $E\|y_0\|^2 < \infty$ and (3.28) imply that $\zeta_n = o_P(1)$, and $n^{-1}\sum_{t=2}^n |A_{1,t}|\|y_{t-2}\| = O_P(1)$. Hence,

$$\begin{aligned}
|C_{42}^0| &= |n^{-1/2}\sum_{t=2}^n A_{1,t}\int \nu_{t-1}(u)d\varphi_2^0(u)| \\
&\le n^{-1/2}\sum_{t=2}^n |A_{1,t}|\int \Big|F(y+n^{-1/2}y'_{t-2}\tilde\gamma_y) - F(y) - n^{-1/2}y'_{t-2}\tilde\gamma_y f(y)\Big|d\psi_2^0(y) \\
&\le n^{-1}\sum_{t=2}^n |A_{1,t}|\|y_{t-2}\| \sup_{|y-x|\le\zeta_n}|f(y)-f(x)| = o_P(1).
\end{aligned}$$

Next, we treat the C_{41}^0 term. Define, for $s\in\mathbb{R}^{p+1}$, y, $a\in\mathbb{R}$,

$$\begin{aligned}
m_{t-1}(y,s,a) &:= [F(y+n^{-1/2}(y'_{t-2}s+\|y_{t-2}\|a)) - F(y)] \\
C_{41}^{0\pm}(y,s,a) &:= n^{-1/2}\sum_{t=2}^n A_{1,t}^{\pm}[I(\varepsilon_{t-1}\le y+n^{-1/2}(y'_{t-2}s+\|y_{t-2}\|a)) - I(\varepsilon_{t-1}\le y) \\
&\qquad\qquad -m_{t-1}(y,s,a)], \\
C_{41}^0(y,s) &:= C_{41}^{0,+}(y,s) - C_{41}^{0,-}(y,s) \\
&:= n^{-1/2}\sum_{t=2}^n A_{1,t}[I(\varepsilon_{t-1}\le y+n^{-1/2}y'_{t-2}s) - I(\varepsilon_{t-1}\le y) - \mu_{t-1}(F(y),s)],
\end{aligned}$$

where $C_{41}^{0,+}(y,s) = C_{41}^{0,\pm}(y,s,0)$. Note that on the event $\sup_{\alpha_0\le u\le 1-\alpha_0}\|\gamma(u)\|\le b$,

$$\begin{aligned}
|C_{41}^0| &\le \int \sup_{\|s\|\le b}|C_{41}(y,s)|d\psi_2^0(y) \\
&\le \int \sup_{\|s\|\le b}|C_{41}^{0,+}(y,s)|d\psi_2^0(y) + \int \sup_{\|s\|\le b}|C_{41}^{0,-}(y,s)|d\psi_2^0(y).
\end{aligned}$$

We shall show that for every $0<b<\infty$,

$$\int \sup_{\|s\|\le b}|C_{41}^{\pm}(y,s)|d\psi_2^0(y) = o_P(1) \tag{3.29}$$

which together with (3.28) will imply $|C_{41}^0| = o_P(1)$.

Let $c = EA_{1,t}^2$. By a conditioning argument and the boundedness of f, we obtain that for all $y\in\mathbb{R}$, $0\le b<\infty$, $s\in\mathbb{R}^{p+1}$ with $\|s\|\le b$ and for all $a\in\mathbb{R}$,

$$E|C_{41}^{0\pm}(y,s,a)|^2 = E(A_{1,2}^{\pm})^2 E\Big[I(\varepsilon_1\le y+n^{-1/2}(y'_0 s+\|y_0\|a)) - I(\varepsilon_1\le y) \tag{3.30}$$

$$-F(y+n^{-1/2}(y_0's+\|y_0\|a))+F(y)\Big]^2$$
$$\le \; c\,E|F(y+n^{-1/2}(y_0's+\|y_0\|a))-F(y)|$$
$$\le \; C\,n^{-1/2}(b+a)E\|y_0\|.$$

Similarly, and for any $s,\ t \in \mathbb{R}^{p+1}$, and $y,\ a \in \mathbb{R}$,

$$\begin{aligned} E|C_{41}^{0\pm}(y,t,a)-C_{41}^{0\pm}(y,s,a)|^2 &\le c\,\|t-s\|\,n^{-1/2}\,E\|y_0\|, \\ E|C_{41}^{0\pm}(y,t,a)-C_{41}^{0\pm}(y,t,0)|^2 &\le c\,|a|\,n^{-1/2}\,E\|y_0\|. \end{aligned} \tag{3.31}$$

Since the ball $\{\|s\|\le b\|\}$ is compact, there is an $\eta>0$ and a finite number of points $s_1,\cdots,s_k$ in the unit ball such that for any $\|s\|\le b$, there is an s_j with $\|s_j\|\le b$, $\|s-s_j\|\le\eta$. Now,

$$\sup_{\|s\|\le b}|C_{41}^{0\pm}(y,s)| \;\le\; \max_{1\le j\le k}\sup_{\|s-s_j\|\le\eta}|C_{41}^{0\pm}(y,s)-C_{41}^{0\pm}(y,s_j)|+\max_{1\le j\le k}|C_{41}^{0\pm}(y,s_j)|; \tag{3.32}$$

$\|s-s_j\|\le\eta$ implies that for all $y\in\mathbb{R}, 1\le j\le k,\ n\ge1,\ 1\le t\le n$,

$$y_{t-2}'s_j-\|y_{t-2}\|\eta\le y_{t-2}'s\le y_{t-2}'s_j+\|y_{t-2}\|\eta.$$

This, the monotonicity of the indicators, and nonnegativity of $A_{1,t}^{\pm}$'s, imply

$$\begin{aligned} &|C_{41}^{0\pm}(y,s)-C_{41}^{0\pm}(y,s_j)| \\ &\le\; |C_{41}^{0\pm}(y,s_j,\eta)-C_{41}^{0\pm}(y,s_j,0)|+|C_{41}^{0\pm}(y,s_j,-\eta)-C_{41}^{0\pm}(y,s_j,0)| \\ &\quad+2n^{-1/2}\{sum_{t=2}^{n}A_{1,t}^{\pm}[m_{t-1}(y,s_j,\eta)-m_{t-1}(y,s_j,-\eta)]. \end{aligned} \tag{3.33}$$

Moreover, by (3.31), and the Cauchy-Schwarz inequality,

$$\mathbb{E}\Big(\max_{1\le j\le k}|C_{41}^{0\pm}(y,s_j,\pm\eta)-C_{41}^{0\pm}(y,s_j,0)|\Big) \;\le\; C\,k\,\eta^{1/2}\,n^{-1/4}.$$

Because f is bounded, the third term in the upper bound of (3.33) is bounded above by

$$\begin{aligned} &\max_{1\le j\le k}n^{-1/2}\sum_{t=2}^{n}|A_{1,t}|\Big|F(y+n^{-1/2}y_{t-2}'s_j+n^{-1/2}\|y_{t-2}\|\eta) \\ &\qquad\qquad -F(y+n^{-1/2}y_{t-2}'s_j-n^{-1/2}\|y_{t-2}\|\eta)\Big| \\ &\le C\eta\,n^{-1}\sum_{t=2}^{n}|A_{1,t}|\|y_{t-2}\|=O_P(\eta). \end{aligned}$$

Finally, by (3.30) applied with $a=0$, and by the Cauchy-Schwarz inequality,

$$\begin{aligned} &\int E\Big(\max_{1\le j\le k}|C_{41}^{0\pm}(y,s_j)|\Big)d\psi_2^0(y) \\ &\le\; d_2^0\sup_y E\Big(\max_{1\le j\le k}|C_{41}^{0\pm}(y,s_j)|\Big)\le d_2^0\sum_{j=1}^{k}E\Big|C_{41}^{0\pm}(y,s_j)\Big|\le C\,k\,n^{-1/4}, \end{aligned}$$

Putting all these bounds together with (3.30) (3.32) and (3.33, we obtain that

$$\begin{aligned}\int \sup_{\|s\|\leq b} |C_{41}^{0\pm}(y,s)| d\psi_2^0(y) &= O_P(k\,\eta^{1/2}\,n^{-1/4}) + O_P(k\,n^{-1/4}) + O_P(\eta) \\ &= O_P(\eta^{-p}\,n^{-1/4}) \quad \forall\,\eta > 0.\end{aligned}$$

Letting $\eta \to 0$ at a suitable rate (such as, for instance, $\eta = n^{-a}$, with $ap < 1/4$), this in turn implies (3.29), and completes the proof of $C_4^0 = o_P(1)$. Similarly one can prove $C_2^0 = o_P(1)$.

Next, note that for $\|s\| \leq b$, $-\|y_{t-2}\|b \leq y'_{t-2}s \leq \|y_{t-2}\|b$. Therefore,

$$\begin{aligned}|C_5^0| &:= \left| n^{-1/2} \sum_{t=2}^n \int D_t d\varphi_1^0 \int D_{t-1} d\varphi_2^0 \right| \\ &\leq \int\int \sup_{\|s\|\leq b} n^{-1/2} \sum_{t=2}^n \Big\{ |I(\varepsilon_t \leq x + n^{-1/2} y'_{t-1}s) - I(\varepsilon_t \leq x)| \\ &\qquad \times |I(\varepsilon_{t-1} \leq y + n^{-1/2} y'_{t-2}s) - I(\varepsilon_{t-1} \leq y)| \Big\} d\psi_1^0(x) d\psi_2^0(y) \\ &\leq \int\int n^{-1/2} \sum_{t=2}^n \Big\{ I(x - n^{-1/2}\|y_{t-1}\|b < \varepsilon_t \leq x + n^{-1/2}\|y_{t-1}\|b) \\ &\qquad \times I(y - n^{-1/2}\|y_{t-2}\|b < \varepsilon_{t-1} \leq y + n^{-1/2}\|y_{t-2}\|b) \Big\} d\psi_1^0(x) d\psi_2^0(y)\end{aligned}$$

so that, by a conditioning argument, using $\|f\|_\infty < \infty$ and the fact that ψ_j^0, $j = 1, 2$ are finite measures on $\mathbb{R}$, we obtain

$$\begin{aligned}E|C_5^0| &\leq \int\int n^{1/2} E\{|F(x + n^{-1/2}\|Y_1\|b) - F(x - n^{-1/2}\|Y_1\|b)| \\ &\qquad \times I(y - n^{-1/2}\|y_0\|b < \varepsilon_1 \leq y + n^{-1/2}\|y_0\|b)\} d\psi_1^0(x) d\psi_2^0(y) \\ &\leq C \int E\Big\{ I(y - n^{-1/2}\|y_0\|b < \varepsilon_1 \leq y + n^{-1/2}\|y_0\|b) \|Y_1\| \Big\} d\psi_2^0(y) \\ &\leq C \int E^{1/2} |F(y + n^{-1/2}\|y_0\|b) - F(y - n^{-1/2}\|y_0\|b)| d\psi_2^0(y) \\ &\leq C\,n^{-1/4} = o(1).\end{aligned}$$

To summarize, we have proved the following

$$\begin{aligned}& n^{-1/2} \sum_{t=2}^n \hat{c}_{n2;1,t} \hat{c}_{n2;2,t-1} \\ &= n^{-1/2} \sum_{t=2}^n \int [I(\varepsilon_t > F^{-1}(u) - (1-u)] d\varphi_1^0(u) \int [I(\varepsilon_{t-1} > F^{-1}(v) - (1-v)] d\varphi_2^0(v) \\ &\qquad + o_P(1).\end{aligned}$$

Along the same lines as on page 1412 of Hallin and Jurečková (1999), one can show that the remaining cross-product terms are negligible. For example consider

$$T_n := n^{-1/2} \sum_{t=2}^{n} \int_0^{\alpha_n} \bar{a}_t d\varphi_1 \int_0^{\alpha_n} \bar{a}_{t-1} d\varphi_2.$$

The Cauchy-Schwarz inequality, the facts that $|u - (1 - \hat{a}_t(u)|^2 \le u + (1 - \hat{a}_t(u))$, and $\sum_{t=1}^n (1 - \hat{a}_t(u)) = nu$, $\forall\, 0 \le u \le 1$, imply

$$\begin{aligned} & n^{-1/2} \sum_{t=2}^{n} \Big| [\hat{a}_t(u) - (1-u)][\hat{a}_{t-1}(v) - (1-v)] \Big| \\ \le\ & n^{-1/2} \Big\{ \sum_{t=2}^{n} [u - (1 - \hat{a}_t(u)]^2 \Big\}^{1/2} \Big\{ \sum_{t=2}^{n} [v - (1 - \hat{a}_{t-1}(v)]^2 \Big\}^{1/2} \\ \le\ & n^{-1/2} \{2n\, u\}^{1/2} \{2n\, v\}^{1/2} = 2n^{1/2} u^{1/2} v^{1/2}. \end{aligned}$$

Thus,

$$\begin{aligned} |T_n| \ \le\ & \int_0^{\alpha_n} \int_0^{\alpha_n} n^{-1/2} \sum_{t=2}^{n} \Big| [\hat{a}_t(u) - (1-u)][\hat{a}_{t-1}(v) - (1-v)] \Big| d\varphi_1(u) d\varphi_2(v) \\ \le\ & 2n^{1/2} \Big(\int_0^{\alpha_n} u^{-1/2-\delta} du \Big)^2 = O(n^{-1/2+\delta}) = o(1). \end{aligned}$$

The next one is even easier because $| \int \bar{a}_t d\varphi_j^0 | \le \varphi_j(1 - \alpha_0)) - \varphi_j(\alpha_0)$, $j = 1, 2$, so that

$$\begin{aligned} \Big| n^{-1/2} \sum_{t=2}^{n} \int \bar{a}_t d\varphi_1^0 \int_0^{\alpha_n} \bar{a}_{t-1} d\varphi_2^0 \Big| \ \le\ & \int_0^{\alpha_n} n^{-1/2} \sum_{t=2}^{n} |u - (1 - \hat{a}_{t-1}(u)| d\varphi_2 \\ \le\ & 2n^{1/2} \int_0^{\alpha_n} u^{-\delta} du = O(n^{-1/2+\delta}). \end{aligned}$$

Exactly similar arguments can be used to show that the cross-product terms involving the right tail integrals are also negligible. Putting all these conclusions together implies

$$\begin{aligned} & n^{-1/2} \sum_{t=2}^{n} \hat{b}_t \hat{b}_{t-1} \\ =\ & n^{-1/2} \sum_{t=2}^{n} \int [I(\varepsilon_t > F^{-1}(u) - (1-u)] d\varphi_1(u) \int [I(\varepsilon_{t-1} > F^{-1}(v) - (1-v)] d\varphi_2(v) \\ & \qquad + o_P(1) \\ =\ & n^{-1/2} \sum_{t=2}^{n} \varphi_1(F(\varepsilon_t))\, \varphi_2(F(\varepsilon_{t-1})) + o_P(1). \end{aligned}$$

This proves the left hand side of the claim (3.1) while the right hand side of the claim follows from Hallin and Puri (1988) and Hallin and Werker (1998). □

Remark 3.1 Here we indicate the proof when not all φ's are centered at the origin. For example consider the case $m = 2$. Assume that $\mu_1 := \int \varphi_1(u)du = 0$ and $\int \varphi_2(u)du =: \mu_2 \neq 0$. Put $\varphi_2^0 = \varphi_2 - \mu_2$. Let $\bar{a}_t(u) := \hat{a}_t(u) - (1-u)$ and note that integration by parts and $\mu_1 = 0$ yields that

$$\begin{aligned}\int \varphi_1(u)d\hat{a}_t(u) \int \varphi_2^0(v)d\hat{a}_{t-1}(v) &= \int \varphi_1(u)d\bar{a}_t(u) \int \varphi_2^0(v)d\bar{a}_{t-1}(v) \\ &= \int \bar{a}_t(u)d\varphi_1(u) \int \bar{a}_{t-1}(v)d\varphi_2(v), \qquad \forall\ t.\end{aligned}$$

Hence $\tilde{S}_{n,\varphi_1\varphi_2^0} = \tilde{S}_{n,\varphi_1\varphi_2}$, and $\tilde{S}_{n,\varphi_1\varphi_2} = \tilde{S}_{n,\varphi_1\varphi_2^0} = S_{n,\varphi_1\varphi_2^0}(\theta) + o_P(n^{1/2})$, by Proposition 3.1. But,

$$\begin{aligned}n^{-1/2}S_{n,\varphi_1\varphi_2^0}(\theta) &= n^{-1/2}\sum_{t=1}^{n}\varphi_1\Big(\frac{R_t(\theta)}{n+1}\Big)\Big[\varphi_2\Big(\frac{R_{t-1}(\theta)}{n+1}\Big) - m_2\Big] \\ &= n^{-1/2}S_{\varphi_1\varphi_2}(\theta) - n^{1/2}m_2[n^{-1}\sum_{i=1}^{n}\varphi_1(i/(n+1)],\end{aligned}$$

where $n^{-1}\sum_{i=1}^{n}\varphi_1(i/(n+1)$ is a Rieman sum for the $\int \varphi_1(u)du = 0$. Moreover, since φ_1 is square integrable, the difference between this Rieman sum and its limit is $o(n^{-1/2})$. so that centering φ_2 has asymptotically negligible influence on $n^{-1/2}S_{n\varphi_1\varphi_2}$, and the conclusion of Proposition 3.1 continues to hold.

References

Akharif, A. and M. Hallin (2003). Efficient detection of random coefficients in AR(p) models. *Ann. of Statist.* **31**, 675-704.

Allal, J. and S. El Melhaoui (2005). Optimal detection of exponential components in autoregressive models, submitted.

Benghabrit, Y. and M. Hallin (1992). Optimal rank-based tests against first-order super diagonal bilinear dependence. *Journal of Statistical Planning and Inference* **32**, 45-61.

Bentarzi, M. and M. Hallin (1996). Locally optimal tests against periodic autoregression : parametric and nonparametric approaches. *Econometric Theory* **12**, 88-112.

Chernoff, H. and Savage, I.R. (1958). Asymptotic normality and efficiency of certain nonparametric tests. *Ann. Math. Statist.* **29**, 972-994.

Garel, B. and M. Hallin (1999). Rank-based AR order identification. *Journal of the American Statistical Association* **94**,1357-1371.

Gutenbrunner, C. and J. Jurečková (1992). Regression rank scores and regression quantiles. *Ann. Statist.* **20**,305-330.

Gutenbrunner, C., J. Jurečková, R. Koenker, and S. Portnoy (1993). Tests of linear hypotheses based on regression rank scores. *J. Nonparametric Statist.* **2**, 307-331.

Hallin, M. (1994). On the Pitman non-admissibility of correlogram based methods. *J. Time Ser. Anal.* **15**, 607-612.

Hallin, M., J.-Fr. Ingenbleek, and M.L. Puri. (1985). Linear serial rank tests for randomness against *ARMA* alternatives. *Ann. Statist.* **13**, 1156-1181.

Hallin, M. and J. Jurečková (1999). Optimal tests for autoregressive models based on autoregression rank scores. *Ann. of Statist.* **27**, 1385-1414.

Hallin, M., J. Jurečková, J. Kalvová, J. Picek and T. Zahaf (1997). Non-parametric tests in AR-models, with applications to climatic data. *Environmetrics* **8**, 651–660.

Hallin, M. and M.L. Puri (1988). Optimal rank-based procedures for time-series analysis: testing an *ARMA* model against other *ARMA* models. *Ann. Statist.* **16**, 402-432.

Hallin, M. and M.L. Puri (1994). Aligned rank tests for linear models with autocorrelated errors. *J. Multiv. Anal.* **50**, 175-237.

Hallin, M. and B. Werker. (1998). Optimal testing for semiparametric *AR* models: from Gaussian Lagrange multipliers to autoregression rank scores and adaptive tests. In S. Ghosh, Ed.: *Asymptotics, Nonparametrics and Time Series*, M. Dekker, New York, 295-350.

Hallin, M. and B. Werker. (2003). Semiparametric efficiency, distribution-freeness, and invariance. *Bernoulli* **9**, 137-165.

Jurečková, J. (1971). Asymptotic independence of rank test statistic for testing symmetry on regression. *Sankhyā Ser. A* **33**, 1–18.

Kalvová, J., J. Jurečková, I. Nemešová and J. Picek. (2000). On the order of autoregression (AR) models in temperature series. *Meteorologický časopis* **3**, 19–23.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica* **46**, 33-50.

Koul, Hira L. (1970). A class of ADF tests for sub-hypothesis in the multiple linear regression. *Ann. Math. Statist.* **41**, 1273–1281.

Koul, Hira L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models, Second Edition*, Springer Lecture Notes in Statistics, **166**, New York.

Koul, H.L. and A.K.Md.E. Saleh (1995). Autoregression quantiles and related rank scores processes. *Ann. Statist.* **23**, 670-689.

Paindaveine, D. (2004). A unified and elementary proof of serial and nonserial, univariate and multivariate, Chernoff-Savage results. *Statistical Methodology* **1**, 81-91.

Paindaveine, D. (2005). A Chernoff-Savage result for shape: on the non-admissibility of pseudo-Gaussian methods, submitted.