The Annals of Statistics
1990, Vol. 18, No. 3, 1470–1482

TESTING LINEAR HYPOTHESES IN AUTOREGRESSIONS

BY JENS-PETER KREISS

University of Hamburg

The problem of testing linear hypotheses about the parameter vector of an autoregressive process with finite order is considered. Based on the property of local asymptotic normality, we derive asymptotically optimal statistical tests. Additionally, we define and investigate so-called residual rank tests. For these tests we obtain under the null hypothesis an asymptotic distribution which does not depend on the distribution of the innovation.

1. Introduction. In this paper we deal with the problem of testing linear hypotheses for the parameters of an autoregressive process with finite order. For such models, the property of local asymptotic normality (LAN) holds [cf. Kreiss (1987)]. Because of this, it is possible to construct optimal estimators and optimal tests for linear hypotheses. Here optimal is meant in a local and asymptotical sense. This methodology goes back to Le Cam (1960, 1986) and is also fully exploited in Strasser (1985). Since we dealt with estimation problems in the above-mentioned paper [Kreiss (1987)], the aim in the following is to treat testing problems. Of course, the techniques for both kinds of problems are somewhat similar. So we omit certain technicalities here. Nevertheless we would like to mention that not all results of this paper are standard and easy to obtain. This is especially true for the tests with ranked residuals to be defined later.

Before introducing the underlying model we will describe the content of the paper. Section 2 contains some notation and basic results used throughout the whole paper. In Section 3 we restate some results concerning statistical tests for linear hypotheses in LAN models. These results are presented in a version adapted to the situation considered here. Section 4 introduces general score and ranked residual score tests for autoregressive models. This part also contains the main results of the paper. As a matter of fact we rediscover Quenouille's (1947) goodness-of-fit test for autoregressions as a score test with score function adapted to the normally distributed case. To make the paper more inviting to read, some proofs and auxiliary results are deferred to the final section.

Now let us introduce the stochastic model which we deal with. Consider a family of real-valued random variables $(X_t: t \in \mathbb{Z} = \{0, \pm 1, \pm 2, \cdots\})$ on an arbitrary probability space $(\Omega, \mathscr{A}, P)$, such that

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + \varepsilon_t, \qquad t \in \mathbb{Z}, \tag{1.1}$$

Received November 1985; revised August 1989.

AMS 1980 *subject classifications*. Primary 62F03, 62F05, 62M10; secondary 62F07, 62F35.

Key words and phrases. Autoregressive process, testing linear hypotheses, local asymptotic normality, ranked residuals, asymptotic distribution.

where $(\varepsilon_t\colon t \in \mathbb{Z})$ is a sequence of independent and identically distributed real-valued random variables. Further, we assume $E\varepsilon_1 = 0$ and $E\varepsilon_1^2 = \sigma^2 \in (0, \infty)$. To ensure stationarity of the considered stochastic process $(X_t\colon t \in \mathbb{Z})$, we assume that the parameter vector $\vartheta = (a_1, \dots, a_p)^{\mathrm{T}} \in \mathbb{R}^p$ is restricted to the following parameter space:

$$\Theta = \left\{ (a_1, \dots, a_p)^{\mathrm{T}} \in \mathbb{R}^p \colon 1 - \sum_{\nu=1}^{p} a_\nu z^\nu \neq 0 \text{ for all complex } z \text{ with } |z| \leq 1 \right\}.$$

Such autoregressive processes are well known in the literature on time series models, cf. Anderson (1971) and many others.

2. Basic notation and results. We assume throughout the whole paper that the distribution of ε_1 possesses an absolutely continuous Lebesgue density f with finite Fisher information $I(f)$ and with $f(x) > 0$ for all $x \in \mathbb{R}$. Then the autoregressive process defined above is asymptotically normal. Denoting by $P_{n,\vartheta}$ the distribution of $(X_t\colon t \geq 1 - p)$ on $\mathbb{R}^\infty$ restricted to $\sigma(X_{1-p}, \dots, X_n)$, this means that (2.1) and (2.2) hold. [Recall that $P_{n,\vartheta}$ is nothing else then the common distribution of $(X_{1-p}, \dots, X_n)$ written as a probability measure on $\mathbb{R}^\infty$.]

For all sequences $\{\vartheta_n\}_{n \in \mathbb{N}} \subset \Theta$ for which $\sqrt{n}(\vartheta_n - \vartheta)$ is bounded

$$\log \frac{dP_{n,\vartheta_n}}{dP_{n,\vartheta}} - \sqrt{n}(\vartheta_n - \vartheta)^{\mathrm{T}} \Delta_n(\vartheta) + \frac{1}{2} n (\vartheta_n - \vartheta)^{\mathrm{T}} \Gamma(\vartheta) I(f) (\vartheta_n - \vartheta) = o_p(1), \tag{2.1}$$

$$\mathscr{L}\big(\Delta_n(\vartheta) | P_{n,\vartheta}\big) \Rightarrow \mathscr{N}(0, I(f)\Gamma(\vartheta)). \tag{2.2}$$

Here we make use of the following notation.

$$\Delta_n(\vartheta) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \dot{\varphi}(\varepsilon_j)(X_{j-1}, \dots, X_{j-p})^{\mathrm{T}}, \qquad \dot{\varphi} = -f'/f, \tag{2.3}$$

$$\Gamma(\vartheta) = (EX_s X_t)_{s,t=1,\dots,p}. \tag{2.4}$$

Furthermore, $\mathscr{L}(X)$ denotes the distribution of a random variable, $\mathscr{N}(a, \Sigma)$ denotes the multivariate normal distribution with mean vector a and covariance matrix Σ and $o_p(1)$ stands for convergence to 0 in $P_{n,\vartheta}$-probability.

A detailed proof of this LAN result is given in Kreiss (1987). There the more general ARMA situation is considered. Note that the assumptions made here guarantee that the $p \times p$-covariance matrix $\Gamma(\vartheta)$ is positive definite.

The above results together with standard techniques imply contiguity of $\{P_{n,\vartheta}\}_{n \in \mathbb{N}}$ and $\{P_{n,\vartheta_n} = P_{n,\vartheta + n^{-1/2} h_n}\}_{n \in \mathbb{N}}$, and

$$\mathscr{L}\big(\Delta_n(\vartheta) - \Gamma(\vartheta) I(f) h_n | P_{n,\vartheta_n}\big) \Rightarrow \mathscr{N}(0, \Gamma(\vartheta) I(f)), \tag{2.5}$$

for each bounded sequence $\{h_n\}_{n \in \mathbb{N}} \subset \mathbb{R}^p$. Using the terminology of Strasser

(1985), we obtain that the sequence of experiments $[\mathscr{A}_n = \sigma(X_{1-p}, \dots, X_n)]$

$$\left(E_n = \left(\mathbb{R}^\infty, \mathscr{A}_n, \{P_{n,\vartheta+n^{-1/2}W(\vartheta)I(f)^{-1/2}h}: h \in H_n\}\right)\right)_{n\in N}$$

is asymptotically normal with central sequence $W(\vartheta)^{\mathrm{T}}I(f)^{-1/2}\Delta(\vartheta)$ [for every matrix $W(\vartheta)$ with $W(\vartheta)W(\vartheta)^{\mathrm{T}} = \Gamma(\vartheta)^{-1}$] [cf. Strasser (1985), Corollary 80.6]. H_n is defined to be the following set of local parameters $\{h \in \mathbb{R}^p: \vartheta + n^{-1/2}W(\vartheta)I(f)^{-1/2}h \in \Theta\}$.

For such experiments it is possible to treat linear testing problems in an asymptotic manner. Some useful results are stated in the next section and are later on used for the model considered here.

3. Testing linear hypotheses. Given the (localized) sequence of experiments (E_n), suppose we are interested in testing

$$H: h \in L \quad \text{versus} \quad K: h \in \mathbb{R}^p \setminus L, \tag{3.1}$$

where L denotes a linear subspace of $\mathbb{R}^p$ of dimension $s, 0 < s < p$.

Since Theorem 80.13 of Strasser (1985) yields equicontinuity for the sequence of experiments $(E_n: n \in \mathbb{N})$, we define the following sequence of tests $(\varphi_n^*: n \in \mathbb{N})$ for the above hypothesis.

$$\varphi_n^* = \begin{cases} 1, & \left\|(\mathrm{id} - \pi_L)\circ\left(W(\vartheta)^{\mathrm{T}}\Delta(\vartheta)\right)\right\|^2 > c_\alpha, \\ 0, & \left\|(\mathrm{id} - \pi_L)\circ\left(W(\vartheta)^{\mathrm{T}}\Delta(\vartheta)\right)\right\|^2 \le c_\alpha, \end{cases} \tag{3.2}$$

here π_L denotes the orthogonal projection onto L and id denotes the $p \times p$ identity matrix. c_α should be chosen in such a way that the tests asymptotically achieve level α, $0 < \alpha < 1$. Later on (cf. Theorem 4.1) we show how the critical values c_α can be chosen.

The sequence of tests defined above is asymptotically *maximin* in the following sense.

THEOREM 3.1 [Strasser (1985), Theorem 82.21]. *$\{\varphi_n^*\}_{n\in\mathbb{N}}$ is asymptotically unbiased, that is ($E_{n,h}$ denotes the expectation according to $P_{n,\vartheta+n^{-1/2}W(\vartheta)I(f)^{-1/2}h}$)*

$$\limsup_{n\to\infty} E_{n,h}\varphi_n^* \le \alpha \quad \textit{for all } h \in L$$

and

$$\liminf_{n\to\infty} E_{n,h}\varphi_n^* \ge \alpha \quad \textit{for all } h \in \mathbb{R}^p \setminus L.$$

If $\{\varphi_n\}_{n\in\mathbb{N}}$ denotes another arbitrary but asymptotically unbiased sequence of tests for H versus K, then

$$\limsup_{n\to\infty} \inf_{h\in B_c} E_{n,h}\varphi_n \le \lim_{n\to\infty} \inf_{h\in B_c} E_{n,h}\varphi_n^*,$$

where $B_c := \{h \in \mathbb{R}^p: \|h - \pi_L(h)\|^2 = c\}$, $c > 0$.

In order to make it possible to apply this test φ_n^*, we have to construct a version which does not depend on the underlying parameter ϑ. For simplicity we only treat the special hypothesis $(0 < s < p)$

$$L := \{x \in \mathbb{R}^p : x_{s+1} = \cdots = x_p = 0\}. \tag{3.3}$$

In case the center of localization ϑ lies in the linear space L, the corresponding linear hypothesis just describes the problem of testing $a_{s+1} = \cdots = a_p = 0$. That is, we have a test for the order of the underlying autoregressive process. The asymptotic properties (e.g., asymptotic power) of the test are judged under local alternatives of the form $\vartheta + n^{-1/2}h$, $h \in \mathbb{R}^p$.

In order to construct an applicable version of the sequence of maximin tests defined in (3.2) we make use of an arbitrary but $\sqrt{n}$-consistent sequence of estimators $\hat{\vartheta}_n$ for ϑ and a consistent sequence of estimates for the covariance matrix $\Gamma(\vartheta)$. If we assume, that we have observations $X_{1-p}, \dots, X_n$, $n \in \mathbb{N}$, of the underlying stochastic process both estimates exist and, for example, may be chosen according to ($\bar{X}_n := [1/(n+p)]\Sigma_{j=1-p}^n X_j$)

$$\hat{\Gamma}_n = \left(\frac{1}{n}\sum_{j=1}^{n}\left(X_{j-|r-s|} - \bar{X}_n\right)\left(X_j - \bar{X}_n\right)\right)_{r,s=1,\dots,p} \tag{3.4}$$

and

$$\hat{\vartheta}_n = \hat{\Gamma}_n^{-1} \cdot \left(\frac{1}{n}\sum_{j=1}^{n}\left(X_{j-r} - \bar{X}_n\right)\left(X_j - \bar{X}_n\right)\right)_{r=1,\dots,p}. \tag{3.5}$$

The properties of the empirical autocovariance matrix $\hat{\Gamma}_n$ and of the Yule–Walker estimate $\hat{\vartheta}_n$ are well known [cf. Brockwell and Davis (1986), Chapters 7.2 and 8.1].

Next define matrices $J(s \times s)$, $B(p-s \times s)$ and $M(p-s \times p-s)$ by

$$\Gamma(\vartheta) = \begin{pmatrix} J(\vartheta) & B^{\mathrm{T}}(\vartheta) \\ B(\vartheta) & M(\vartheta) \end{pmatrix}$$

and analogously $\hat{J}_n$, $\hat{B}_n$ and $\hat{M}_n$. Then $W(\vartheta)$ equal to

$$\begin{pmatrix} J(\vartheta)^{-1/2} & -J(\vartheta)^{-1}B(\vartheta)^{\mathrm{T}}C(\vartheta)^{-1/2} \\ 0 & C(\vartheta)^{-1/2} \end{pmatrix},$$

$C(\vartheta) = M(\vartheta) - B(\vartheta)J(\vartheta)^{-1}B(\vartheta)^{\mathrm{T}}$, is a suitable choice for the matrix $W(\vartheta)$ occurring in (3.2), i.e., $W(\vartheta)W(\vartheta)^{\mathrm{T}} = \Gamma(\vartheta)^{-1}$. Recall that all required inversions are possible since $\Gamma(\vartheta)$ is positive definite.

Thus we should use

$$\hat{W}_n = \begin{pmatrix} \hat{J}_n^{-1/2} & -\hat{J}_n^{-1}\hat{B}_n^{\mathrm{T}}\hat{C}_n^{-1/2} \\ 0 & \hat{C}_n^{-1/2} \end{pmatrix}$$

as a consistent estimate for $W(\vartheta)$. Now we are able to state one of the main results of the paper.

THEOREM 3.2. *Assume that $\dot{\varphi} = -f'/f$ is twice continuously differentiable with bounded derivatives and that $E\varepsilon_1^4 < \infty$. Define*

$$\varphi_n^+ = \begin{cases} 1, & \left\| \pi_2 \circ \left(\hat{W}_n^{\mathrm{T}} \Delta_n(\pi_1 \hat{\vartheta}_n) \right) \right\|^2 > c_\alpha, \\ 0, & \left\| \pi_2 \circ \left(\hat{W}_n^{\mathrm{T}} \Delta_n(\pi_1 \hat{\vartheta}_n) \right) \right\|^2 \le c_\alpha, \end{cases} \tag{3.6}$$

where c_α is defined as given in Theorem 4.1, below. Here π_1 and π_2 denote the projection on the first s, respectively last $p - s$ components. Then the sequence of tests $\{\varphi_n^+\}_{n \in \mathbb{N}}$ is maximin in the local sense of Theorem 3.1.

PROOF. Since, with $\vartheta \in L$,

$$\begin{aligned} &\pi_2 \circ \left(\hat{W}_n^{\mathrm{T}} \Delta_n(\pi_1 \hat{\vartheta}_n) \right) \\ &\quad = \pi_2 \circ \left(\hat{W}_n^{\mathrm{T}} \left[\Delta_n(\pi_1 \hat{\vartheta}_n) - \Delta_n(\vartheta) + \Gamma(\vartheta) I(f) \sqrt{n}\, \pi_1(\hat{\vartheta}_n - \vartheta) \right] \right) \\ &\qquad + \pi_2 \circ \left(W(\vartheta)^{\mathrm{T}} \Delta_n(\vartheta) \right) - \pi_2 \circ \left(W(\vartheta)^{-1} I(f) \sqrt{n}\, \pi_1(\hat{\vartheta}_n - \vartheta) \right) + o_p(1) \\ &\quad = \pi_2 \circ \left(W(\vartheta)^{\mathrm{T}} \Delta_n(\vartheta) \right) + o_p(1), \end{aligned}$$

the assertion follows from Theorem 3.1, the consistency of $\hat{W}_n$, the property that $W(\vartheta)^{-1}$ is upper triangular, the definition of φ_n^* and Lemma 4.2. Moreover, we have that if $\vartheta \in L$ then $\hat{W}_n^{\mathrm{T}} I(f)^{-1/2} \Delta_n(\pi_1 \hat{\vartheta}_n)$ is also a central sequence for $(E_n\colon n \in \mathbb{N})$. $\square$

REMARKS. Dzhaparidze (1977) considered likelihood ratio procedures for testing linear hypotheses. The associated test statistic is asymptotically equivalent to the maximin test defined above [cf. Kreiss (1985b)]. A test similar to φ_n^+ has been defined in Basawa and Scott (1983). They consider nonergodic models and they calculate asymptotic power of their proposal. In Basawa and Koul (1983) there are also some asymptotic optimality results for score tests like φ_n^+ in nonergodic models. Our Theorem 3.1 is a completely different optimality result for the ergodic autoregressive models considered herein. All results given in this part extend directly to more general ARMA models, since the key result (the LAN condition) holds for stationary ARMA processes as well [cf. Kreiss (1987)].

4. General score tests. In the previous part we derived a sequence of tests which is locally optimal. In real application of this test there occurs the problem, that the so-called *score function* $\dot{\varphi} = -f'/f$ is rarely known. If we replace this function $\dot{\varphi}$ by a suitably chosen function $\psi\colon \mathbb{R} \to \mathbb{R}$ we obtain general score tests based on the following kind of score statistic

$$\Psi_n(\vartheta) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \psi(\varepsilon_j(\vartheta)) \mathbf{X}(j-1), \tag{4.1}$$

where $\boldsymbol{X}(j-1) = (X_{j-1}, \dots, X_{j-p})$ and $\varepsilon_j(\vartheta) = X_j - \vartheta^T \boldsymbol{X}(j-1)$, $j = 1, \dots, n$. ψ is assumed to be square integrable according to f, and has to fulfill

$$\int_{\mathbb{R}} \psi f \, d\lambda = 0. \tag{4.2}$$

One possible choice is a bounded score function like $\psi = \arctan$ in order to robustify the procedure. Another possibility is to choose ψ data-dependent, for example $\psi = b \circ F_{\hat{\vartheta}_n}$ for a square integrable function b: $[0,1] \to \mathbb{R}$ and the empirical distribution function of $\varepsilon_1(\hat{\vartheta}_n), \dots, \varepsilon_n(\hat{\vartheta}_n)$. In this case we will assume that $\int_0^1 b(t)\, dt = 0$. Moreover, (4.1) equals

$$B_n(\vartheta) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} b\left(\frac{R_j(\vartheta)}{n}\right) \mathbf{X}(j-1), \tag{4.3}$$

where $R_1(\vartheta), \dots, R_n(\vartheta)$ denote the ranks of the residuals $\varepsilon_1(\vartheta), \dots, \varepsilon_n(\vartheta)$. One advantage of the ranked residual test, i.e. (again suppose that the hypothesis of interest is $a_{s+1} = \cdots = a_p = 0$),

$$\varphi_n^b = \begin{cases} 1, & \left\| \pi_2 \circ \left(\hat{W}_n^{\mathrm{T}} B_n(\pi_1 \hat{\vartheta}_n) \right) \right\|^2 > c_{\alpha, b}, \\ 0, & \left\| \pi_2 \circ \left(\hat{W}_n^{\mathrm{T}} B_n(\pi_1 \hat{\vartheta}_n) \right) \right\|^2 \le c_{\alpha, b}, \end{cases} \tag{4.4}$$

is that the critical value $c_{\alpha, b}$ can be chosen *independently* of the distribution $\mathscr{L}(\varepsilon_1)$. Such a result is not true for general score tests, i.e., for

$$\varphi_n^\psi = \begin{cases} 1, & \left\| \pi_2 \circ \left(\hat{W}_n^{\mathrm{T}} \Psi_n(\pi_1 \hat{\vartheta}_n) \right) \right\|^2 > c_{\alpha, \psi, F}, \\ 0, & \left\| \pi_2 \circ \left(\hat{W}_n^{\mathrm{T}} \Psi_n(\pi_1 \hat{\vartheta}_n) \right) \right\|^2 \le c_{\alpha, \psi, F}, \end{cases} \tag{4.5}$$

nor for the maximin test from the previous section. All critical values should be chosen in such a way that the associate test achieves level α, asymptotically, $0 < \alpha < 1$.

To be more precise, we obtain the following essential result for the tests (4.4) and (4.5).

THEOREM 4.1. *Under the assumptions that ψ, b and F are twice continuously differentiable with bounded derivatives, that $E\varepsilon_1^4 < \infty$, that $\int_{\mathbb{R}} \psi^2 \, dF$, $\int_0^1 b^2 \, d\lambda < \infty$ and that $\int_{\mathbb{R}} \psi \, dF = \int_0^1 b \, d\lambda = 0$, we obtain*

(i) $$\mathscr{L}\left(\left(\int \psi^2 \, dF \right)^{-1} \left\| \pi_2 \circ \left(\hat{W}_n^{\mathrm{T}} \Psi_n(\pi_1 \hat{\vartheta}_n) \right) \right\|^2 \Big| P_{n, \vartheta_n} \right) \Rightarrow \chi^2_{p-s}(\delta_\psi^2)$$

for the alternative $\vartheta_n = \vartheta + n^{-1/2} h$, $h \in \mathbb{R}^p$ and $\vartheta \in L$, where the noncentrality parameter δ_ψ^2 of the χ^2 distribution is equal to $h^{\mathrm{T}} C(\vartheta) h (\int \psi f'/f \, dF)^2 \cdot (\int \psi^2 \, dF)^{-1}$.

A similar result holds for the test statistic of the ranked residual test, namely

$$\text{(ii)}\quad \mathscr{L}\left(\left(\int_0^1 b^2(u)\,du\right)^{-1}\left\|\pi_2\circ\left(\hat{W}_n^{\mathrm{T}}B_n(\pi_1\hat{\vartheta}_n)\right)\right\|^2\middle|P_{n,\vartheta_n}\right)\Rightarrow\chi^2_{p-s}(\delta_b^2),$$

where

$$\delta_b^2=h^{\mathrm{T}}C(\vartheta)h\left(\int b\frac{f'\circ F^{-1}}{f\circ F^{-1}}dF\right)^2\left(\int_0^1 b^2\,d\lambda\right)^{-1}.$$

Because of this we may choose the critical values according to $(0<\alpha<1)$

$$c_{\alpha,\psi,F}=\chi^2_{p-s;\alpha}\int\psi^2\,dF\quad\text{and}\quad c_{\alpha,b}=\chi^2_{p-s;\alpha}\int_0^1 b^2\,d\lambda.$$

The asymptotic power of these tests according to the alternative $\{\vartheta_n\}_{n\in\mathbb{N}}$ *can be calculated easily from the noncentral* χ^2*-distribution.*

The proof of Theorem 4.1 relies mainly on the following lemma, which establishes a kind of asymptotic linearity of the statistics Ψ_n and B_n.

LEMMA 4.2. *Assume* ψ, b *and* F *to be twice continuously differentiable with bounded derivatives,* $E\varepsilon_1^4<\infty$, $\int_{\mathbb{R}}\psi^2\,dF$, $\int_0^1 b^2\,d\lambda<\infty$ *and* $\int_{\mathbb{R}}\psi\,dF=\int_0^1 b\,d\lambda=0$. *Then for all* $\{\hat{\vartheta}_n\}_{n\in\mathbb{N}}\subset\Theta$ *for which* $\{\sqrt{n}(\hat{\vartheta}_n-\vartheta)\}_{n\in\mathbb{N}}$ *stays bounded*

$$\text{(i)}\qquad \Psi_n(\hat{\vartheta}_n)-\Psi_n(\vartheta)-\Gamma(\vartheta)\sqrt{n}(\hat{\vartheta}_n-\vartheta)\int\psi\frac{f'}{f}dF=o_p(1)$$

and

$$\text{(ii)}\qquad \begin{aligned}&B_n(\hat{\vartheta}_n)-\frac{1}{\sqrt{n}}\sum_{j=1}^n b\circ F(\varepsilon_j)\mathbf{X}(j-1)\\&\quad-\Gamma(\vartheta)\sqrt{n}(\hat{\vartheta}_n-\vartheta)\int_0^1 b\frac{f'\circ F^{-1}}{f\circ F^{-1}}d\lambda=o_p(1).\end{aligned}$$

In Section 5 we just give the proof of (ii): This is the much more difficult case. In fact, (i) is an easy consequence of the assumption of differentiability of ψ.

PROOF OF THEOREM 4.1. Exactly as in the proof of Theorem 3.2 we obtain from Lemma 4.2(i) that

$$\pi_2\circ\left(\hat{W}_n^{\mathrm{T}}\Psi_n(\pi_1\hat{\vartheta}_n)\right)=\pi_2\circ\left(W(\vartheta)^{\mathrm{T}}\Psi_n(\vartheta)\right)+o_p(1).$$

Lemma 5.2(i) yields

$$\mathscr{L}\left(\pi_2\circ\left(W(\vartheta)^{\mathrm{T}}\Psi_n(\vartheta)\right)\middle|P_{n,\vartheta_n}\right)\Rightarrow\mathscr{N}\left(-C^{1/2}(\vartheta)h\int\psi\frac{f'}{f}dF,\,I_{p-s}\int\psi^2\,dF\right)$$

(where I_{p-s} denotes the $p-s$-dimensional identity matrix) from which (i) can be concluded. Exactly the same idea yields (ii) [use Lemmas 4.2(ii) and 5.2(ii)]. All other results are easy consequences of (i) and (ii). □

Let us finish this part with an example.

EXAMPLE 4.3 (Quenouille's goodness-of-fit test). Using the score test $\varphi_n^{\psi_0}$ for $\psi_0(x) = x/\sigma^2$ together with the least squares estimator $\hat{\vartheta}_n^{\mathrm{LS}}$ we obtain (by tedious but direct computation) that this test coincides with the one proposed by Quenouille (1947) [see also Anderson (1971), Section 5.6.3] for testing the order of autoregressions.

The asymptotic power for this case was discovered by Walker (1952). For normally distributed innovations ε_t, this correspondence is also obtained in Dzhaparidze (1977).

REMARKS. A possible way to construct adaptive tests for linear hypotheses (adaptation implies independence from the distribution F of the innovations and asymptotical equivalence to φ_n^*) is opened up by the ranked residual test: The optimal score function b is of course $-f' \circ F^{-1}/f \circ F^{-1}$. Along the lines of Beran (1974), this function can be estimated consistently from estimated innovations $\varepsilon_1(\hat{\vartheta}_n), \dots, \varepsilon_n(\hat{\vartheta}_n)$. For a proof of this nontrivial result the reader is referred to Kreiss (1988), where the more general case of autoregressive processes with infinite order is treated.

5. Some proofs and auxiliary results. In this final part of the paper we will give proofs of some results stated in Section 4. Let us begin with an evaluation of Lemma 4.2. Since part (ii) of this lemma is much more complicated and part (i) is contained in Kreiss (1985a), we restrict our further consideration to the second statement of Lemma 4.2.

PROOF OF LEMMA 4.2(ii). We have with $\mathbf{X}(j-1) = (X_{j-1}, \dots, X_{j-p})^{\mathrm{T}}$,

$$\begin{aligned}
B_n(\hat{\vartheta}_n) &- \frac{1}{\sqrt{n}} \sum_{j=1}^n b \circ F(\varepsilon_j)\mathbf{X}(j-1) \\
&= \frac{1}{\sqrt{n}} \sum_{j=1}^n \left\{ b\left(\frac{R_j(\hat{\vartheta}_n)}{n} \right) - b\big(F(\varepsilon_j)\big) \right\} \mathbf{X}(j-1) \\
&= \frac{1}{\sqrt{n}} \sum_{j=1}^n \Big\{ b\big(\hat{F}_n(\varepsilon_j(\hat{\vartheta}_n))\big) - b\big(\hat{F}_n(\varepsilon_j)\big) \\
&\qquad\qquad - b\big(F(\varepsilon_j(\hat{\vartheta}_n))\big) + b\big(F(\varepsilon_j)\big) \Big\} \mathbf{X}(j-1) \qquad (5.1) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{j=1}^n \big\{ b\big(\hat{F}_n(\varepsilon_j)\big) - b\big(F(\varepsilon_j)\big) \big\} \mathbf{X}(j-1) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{j=1}^n \Big\{ b\big(F(\varepsilon_j(\hat{\vartheta}_n))\big) - b\big(F(\varepsilon_j)\big) \Big\} \mathbf{X}(j-1),
\end{aligned}$$

where we make use of the following notation:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}_{(-\infty, x]}\big(\varepsilon_j(\hat{\vartheta}_n)\big), \qquad x \in \mathbb{R}, \tag{5.2}$$

$$F_n(x) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}_{(-\infty, x]}\big(\varepsilon_j(\vartheta)\big), \qquad x \in \mathbb{R}, \tag{5.3}$$

and

$$F(x) = P\{\varepsilon_1 \le x\}, \qquad x \in \mathbb{R}. \tag{5.4}$$

We will deal with all summands of (5.1) separately. With the help of

$$\sup_{x \in \mathbb{R}} \sqrt{n}\, \big|\hat{F}_n(x) - F_n(x)\big| = o_P(1), \tag{5.5}$$

[for proof of (5.5) see Boldin (1983)] together with the fact that

$$\frac{1}{n} \sum_{j=1}^{n} \|\mathbf{X}(j-1)\|$$

stays bounded in probability, we obtain from Taylor's formula that the first summand is equal to

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Big\{b\Big(F_n\big(\varepsilon_j(\hat{\vartheta}_n)\big)\Big) - b\big(F_n(\varepsilon_j)\big) - b\Big(F\big(\varepsilon_j(\hat{\vartheta}_n)\big)\Big) + b\big(F(\varepsilon_j)\big)\Big\}\mathbf{X}(j-1) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Big\{F_n\big(\varepsilon_j(\hat{\vartheta}_n)\big) - F\big(\varepsilon_j(\hat{\vartheta})\big)\Big\} b'\Big(F\big(\varepsilon_j(\hat{\vartheta}_n)\big)\Big)\mathbf{X}(j-1) \\
&\qquad - \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \big\{F_n(\varepsilon_j) - F(\varepsilon_j)\big\} b'\big(F(\varepsilon_j)\big)\mathbf{X}(j-1) + o_P(1).
\end{aligned}$$

This approximation is valid because of differentiability of b and the well-known fact that $\sqrt{n}\,\sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$ stays bounded in probability. Since

$$\varepsilon_j(\hat{\vartheta}) - \varepsilon_j(\vartheta) = -\big(\hat{\vartheta}_n - \vartheta\big)^{\mathrm{T}} \mathbf{X}(j-1) \tag{5.6}$$

and since $\sqrt{n}\,(\hat{\vartheta}_n - \vartheta)$ is bounded in probability, a Taylor expansion for $b'(F(\cdot))$ yields that the above expression equals

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Big\{F_n\big(\varepsilon_j(\hat{\vartheta}_n)\big) - F\big(\varepsilon_j(\hat{\vartheta}_n)\big) - F_n(\varepsilon_j) + F(\varepsilon_j)\Big\} \\
&\qquad \times b'\big(F(\varepsilon_j)\big)\mathbf{X}(j-1) + o_P(1).
\end{aligned} \tag{5.7}$$

If we restrict our attention to the set $S_n := \{\max_{j=1,\dots,n}|\varepsilon_j(\hat{\vartheta}_n) - \varepsilon_j| \le n^{-1/4}\}$ we can conclude that (5.7) is bounded by

$$\sqrt{n} \sup_{|x-y|\le n^{-1/4}} |F_n(x) - F(x) - F_n(y) + F(y)| \frac{1}{n}\sum_{j=1}^{n} \|\mathbf{X}(j-1)\|$$
$$= o_P(1)\frac{1}{n}\sum_{j=1}^{n} \|\mathbf{X}(j-1)\| = o_P(1).$$

To see the next to the last equality, use a sequence of routine arguments, which may be found in Billingsley (1968), proof of Theorem 13.1.

Eventually, we obtain that

$$P(S_n^c) \le P\Big\{\sqrt{n}\,\|\hat{\vartheta}_n - \vartheta\| \max_{j=1,\dots,n} \|\mathbf{X}(j-1)\| > n^{1/4}\Big\} \to_{n\to\infty} 0$$

[recall that $n^{-1/4}\max_{1\le t\le n}|X_t| = o_P(1)$ for every stationary process with existing fourth moments]. These results imply that the first summand of (5.1) vanishes in probability.

A Taylor expansion for the third summand leads to $[\varepsilon_j = \varepsilon_j(\vartheta)]$

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\{\varepsilon_j(\hat{\vartheta}_n) - \varepsilon_j(\vartheta)\}(b\circ F)'(\varepsilon_j(\vartheta))\mathbf{X}(j-1)\\ &\quad+\frac{1}{2\sqrt{n}}\sum_{j=1}^{n}\{\varepsilon_j(\hat{\vartheta}_n) - \varepsilon_j(\vartheta)\}^2(b\circ F)''(\delta_{n,j})\mathbf{X}(j-1),\end{aligned}\tag{5.8}$$

where $\delta_{n,j}$ lies between $\varepsilon_j(\hat{\vartheta}_n)$ and $\varepsilon_j(\vartheta)$.

Because of (5.6) and the fact that $\varepsilon_j(\vartheta) = \varepsilon_j$ is independent of $\mathbf{X}(j-1)$ we obtain by direct computation that (5.8) equals [recall that $\sqrt{n}(\hat{\vartheta}_n - \vartheta)$ is bounded in probability]

$$-\frac{1}{n}\sum_{j=1}^{n}(b\circ F)'(\varepsilon_j)\mathbf{X}(j-1)\mathbf{X}(j-1)^{\mathrm{T}}\sqrt{n}\,(\hat{\vartheta}_n - \vartheta) + o_P(1)$$
$$= -E(b\circ F)'(\varepsilon_1)\Gamma(\vartheta)\sqrt{n}\,(\hat{\vartheta}_n - \vartheta) + o_P(1).$$

To see the last equality, use the ergodicity of the underlying process.

Further observe that we obtain from integration by parts

$$E(b\circ F)'(\varepsilon_1) = \int_{-\infty}^{\infty}(b\circ F)'(x)\,f(x)\,dx = -\int_0^1 b\frac{f'\circ F^{-1}}{f\circ F^{-1}}\,d\lambda.$$

To conclude the proof of (ii), it now remains to show that the second summand of (5.1) vanishes in probability. To this end, consider the following Taylor expansion:

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\{b(F_n(\varepsilon_j)) - b(F(\varepsilon_j))\}\mathbf{X}(j-1)\\
(5.9)\qquad &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\{F_n(\varepsilon_j) - F(\varepsilon_j)\}b'(F(\varepsilon_j))\mathbf{X}(j-1)\\
&\quad+\frac{1}{2\sqrt{n}}\sum_{j=1}^{n}\{F_n(\varepsilon_j) - F(\varepsilon_j)\}^2 b''(\delta_{n,j})\mathbf{X}(j-1),
\end{aligned}
$$

where $\delta_{n,j}$ lies between $F_n(\varepsilon_j)$ and $F(\varepsilon_j)$. Because of the fact that $\sqrt{n}\sup_{x\in\mathbb{R}}|F_n(x) - F(x)|$ is bounded in probability, we obtain that the second summand of the above expression vanishes in probability. Finally, a tedious but straightforward computation of

$$E\left\|\sum_{j=1}^{n}\{F_n(\varepsilon_j) - F(\varepsilon_j)\}b'(F(\varepsilon_j))\mathbf{X}(j-1)\right\|^2$$

yields that this expression is bounded in probability. In the computation of the above expectation we use the independence of ε_j and $\mathbf{X}(j-1)$ in a fundamental way. From these facts we obtain the asymptotic negligibility of (5.9), which implies that the second summand of (5.1) vanishes in probability. This concludes the proof of Lemma 4.2(ii). □

The following lemma is needed in the proof of Theorem 4.1.

LEMMA 5.2. (i) *Assume that ψ and f are continuously differentiable, that $\int\psi^2\,dF < \infty$, and that $\int\psi\,dF = 0$. Then for each sequence $\vartheta_n = \vartheta + n^{-1/2}h$, $h \in \mathbb{R}^p$,*

$$(5.10)\qquad \mathscr{L}\left(\Psi_n(\vartheta) + \Gamma(\vartheta)h\int\psi\frac{f'}{f}\,dF\,\middle|\,dP_{n,\vartheta_n}\right) \Rightarrow \mathscr{N}\left(0, \Gamma(\vartheta)\int\psi^2\,dF\right)$$

(ii) *If b, f and F are continuously differentiable with $\int_0^1 b^2\,d\lambda < \infty$ and $\int_0^1 b\,d\lambda = 0$, then*

$$
\begin{aligned}
&\mathscr{L}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} b\circ F(\varepsilon_j)X(j-1) + \Gamma(\vartheta)h\int_0^1 b\frac{f'\circ F^{-1}}{f\circ F^{-1}}\,d\lambda\,\middle|\,P_{n,\vartheta_n}\right)\\
&\qquad\Rightarrow \mathscr{N}\left(0, \Gamma(\vartheta)\int_0^1 b^2\,d\lambda\right).
\end{aligned}
$$

PROOF. (ii) is a direct consequence of (i). Because of Theorem 7.1, Roussas (1972), page 33, in order to obtain (5.10), it suffices to show

$$
\begin{aligned}
&\mathscr{L}\left(\left[\log\frac{dP_{n,\vartheta_n}}{dP_{n,\vartheta}}, \Psi_n(\vartheta)\right]^{\mathrm{T}}\,\middle|\,P_{n,\vartheta}\right)\\
&\qquad\Rightarrow \mathscr{N}\left(\begin{pmatrix} -1/2h^{\mathrm{T}}\Gamma(\vartheta)hI(f)\\ 0\end{pmatrix}, \Sigma(\vartheta)\right),
\end{aligned}
$$

where

$$\Sigma(\vartheta) = \begin{pmatrix} h^{\mathrm{T}}\Gamma(\vartheta)hI(f) & \left(-\Gamma(\vartheta)h\int\psi\frac{f'}{f}\,dF\right)^{\mathrm{T}} \\ -\Gamma(\vartheta)h\int\psi\frac{f'}{f}\,dF & \Gamma(\vartheta)\int\psi^2\,dF \end{pmatrix}.$$

To prove this, consider

$$\log\frac{dP_{n,\vartheta_n}}{dP_{n,\vartheta}} - t^{\mathrm{T}}\Psi_n(\vartheta) = o_P(1) + h^{\mathrm{T}}\Delta_n(\vartheta) - t^{\mathrm{T}}\Psi_n(\vartheta) - \frac{1}{2}h^{\mathrm{T}}\Gamma(\vartheta)I(f)h,$$

see (2.1),

for $t \in \mathbb{R}^p \setminus \{0\}$. Since

$$\mathscr{L}\left(h^{\mathrm{T}}\Delta_n(\vartheta) - t^{\mathrm{T}}\Psi_n(\vartheta) \middle| P_{n,\vartheta}\right)$$
$$\Rightarrow \mathscr{N}\left(0, h^{\mathrm{T}}\Gamma(\vartheta)hI(f) + t^{\mathrm{T}}\Gamma(\vartheta)t\int\psi^2\,dF + 2h^{\mathrm{T}}\Gamma(\vartheta)t\int\psi\frac{f'}{f}\,dF\right),$$

as can be seen from the central limit theorem for martingales [cf. Brown (1971)], we get the desired result. □

Acknowledgments. The author wishes to express his gratitude to the referees. Their reports and their criticisms considerably improved the presentation from earlier versions of the paper.

REFERENCES

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.

BASAWA, I. V. and SCOTT, D. J. (1983). *Asymptotic Optimal Inference for Non-ergodic Models. Lecture Notes in Statist.* **17**. Springer, New York.

BASAWA, I. V. and KOUL, H. L. (1983). Asymptotically minimax tests of composite hypotheses for nonergodic type processes. *Stochastic Process Appl.* **14** 41–54.

BERAN, R. J. (1974). Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist.* **2** 63–74.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BOLDIN, M. V. (1983). Estimation of the distribution of noise in an autoregressive scheme. *Theory Probab. Appl.* **27** 886–871.

BROCKWELL, P. J. and DAVIS, R. A. (1986). *Time Series: Theory and Methods*. Springer, New York.

BROWN, B. M. (1971). Martingale central limit theorems. *Ann. Math. Statist.* **42** 59–66.

DZHAPARIDZE, K. O. (1977). Tests of composite hypotheses for random variables and stochastic processes. *Theory Probab. Appl.* **22** 104–118.

FABIAN, V. and HANNAN, J. (1982). On estimation and adaptive estimation for locally asymptotically normal families. *Z. Wahrsch. Verw. Gebiete* **59** 459–478.

KREISS, J.-P. (1985a). A note on *M*-estimation in stationary ARMA-processes. *Statist. Decisions* **3** 317–336.

KREISS, J.-P. (1985b). Tests and adaptive tests in ARMA-models—the multivariate case. Unpublished manuscript.

KREISS, J.-P. (1987). On adaptive estimation in stationary ARMA-processes. *Ann. Statist.* **15** 112–123.

KREISS, J.-P. (1988). Estimation of the distribution of residuals in AR(∞)-processes. Unpublished manuscript.

LE CAM, L. (1960). Locally asymptotically normal families of distributions. *Univ. Calif. Publ. Statist.* **3** 37–98.

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.

QUENOUILLE, M. H. (1947). A large-sample test for the goodness of fit of autoregressive schemes. *J. Roy. Statist. Soc. Ser. A* **110** 123–129.

ROUSSAS, G. G. (1972). *Contiguity of Probability Measures*. Cambridge Univ. Press, Cambridge.

STRASSER, H. (1985). *Mathematical Theory of Statistics*. de Gruyter, Berlin.

WALKER, A. M. (1952). Some properties of the asymptotic power function of goodness-of-fit tests for linear autoregressive schemes. *J. Roy. Statist. Soc. Ser. B* **14** 117–134.

INSTITUT FÜR MATHEMATISCHE STOCHASTIK
TECHNISCHE UNIVERSITÄT BRAUNSCHWEIG
POCKELSSTRASSE 14
3300 BRAUNSCHWEIG
WEST GERMANY