The Annals of Statistics
1992, Vol. 20, No. 3, 1426–1440

TESTING STATIONARITY IN THE MEAN OF AUTOREGRESSIVE PROCESSES WITH A NONPARAMETRIC REGRESSION TREND

By Hartmut Milbrodt

Universität Dortmund

In this paper, we suggest tests of stationarity in the mean of autoregressive time series versus arbitrary trend alternatives. As an intermediate, though essential, step local asymptotic normality of autoregressive models with a nonparametric regression trend is established. Moreover, a functional central limit theorem for the underlying likelihood ratio processes is derived. These results then offer a general construction principle by which every goodness of fit test (*case* 0), which is based on comparing the empirical distribution function and the hypothetical distribution function, corresponds to a test of stationarity in the mean of AR processes. The asymptotic power of these tests is derived. A small simulation study illustrates the performance of Kolmogorov–Smirnov and Cramér–von Mises type tests of stationarity in the mean at hand of a particular AR(2) process.

1. Introduction. In time series analysis, there is a large variety of procedures that are a priori only applicable to stationary series. This concerns statistical problems in both the time domain and the frequency domain. However, many time series occurring in practice, particularly those arising from economic areas, are nonstationary. To apply the usual methods of time series analysis one has to convert them into a stationary series, for example, by transforming the data, by estimation and elimination of a trend and seasonal components, by repeated differencing or whatsoever [see, e.g., Brockwell and Davis (1987), Sections 1.4 and 9, or Wei (1989), Section 4, in connection with ARIMA models]. This naturally raises the question whether the resulting process is stationary. Due to a lack of alternative methods, this is often decided just by inspecting the series or some of its characteristics, such as the sample autocorrelation function. It is the main goal of the present paper, to provide reliable and easy-to-use tests in order to detect any kind of departures from stationarity in the mean. The procedures proposed are developed in Section 3 (see in particular Theorem 3.4 and Remarks 3.5), whereas Section 2 contains a functional version of a local asymptotic normality result, which is our most important tool.

In a series of recent papers, the usefulness of the notion of local asymptotic normality for the statistical analysis of time series has been demonstrated. Hallin, Ingenbleek and Puri (1985) and Hallin and Puri (1988) proved LAN of

---

Received May 1990; revised September 1991.

AMS 1980 *subject classifications*. Primary 62M10; secondary 62G10, 62G25.

Key words and phrases. Autoregressive processes, nonparametric regression, local, asymptotic normality (LAN), LAW structure, tests of stationarity in the mean.

ARMA models in order to develop a theory of rank tests within time series analysis. Kreiss established LAN of ARMA processes (1987b) and AR(∞) processes (1988, 1990) and gave applications to adaptive and locally asymptotically minimax (LAM) estimation of functionals of the structural parameters (1987a, b; 1988; 1990) and to testing linear hypotheses concerning these parameters (1989). An early predecessor of the testing results is Akritas and Johnson (1982). All these papers refer to stationary processes. Among the papers considering not necessarily stationary time series we mention the thesis of Schütt (1985) and Swensen (1985). In both of these, LAN of AR(p) processes with an additional *linear* regression term was established under conditions including the "stationarity assumption" for the structural parameters and Hájek-type assumptions on the density of the i.i.d. innovations. Swensen used the result to derive the asymptotic power of the Durbin–Watson test of positive dependence of regression residuals under local alternatives. [For similar investigations in the stationary case, see Moussatat (1976), Chapter 9.] Analogous results for MA(q) processes with a linear regression trend have recently been obtained by Garel (1989).

While Akritas and Johnson (1982) rely on the classical technique of differentiability in quadratic mean, which is more suited to the i.i.d. case, later papers employ variants of martingale central limit theorems especially designed for log-likelihoods of dependent observations [Le Cam (1986), 10.5, Theorem 1; Schütt (1985), Satz 3.1.2; Swensen (1985), Lemma 1]. The theorem of Schütt, which has also been used in the papers of Kreiss (1988, 1990), turned out to be particularly useful. Relying on this result, we prove LAN of AR(p) processes with a regression trend in Section 2 of this paper. In fact, an invariance principle for the underlying likelihood ratio processes, which essentially is a variant of the local asymptotic Wiener (LAW) structure of So and Sen (1981) is established along these lines. Of course, this entails LAN. In contrast to Swensen (1985), we impose a nonparametric regression model. More precisely, we use L_2-generated triangular arrays of regression constants. This particular way to look at nonparametric regression, which is inspired by the "approximate scores" of Hájek and Šidák (1967), is made explicit and investigated by Moussatat [(1976), Chapter 8] and Millar (1982) [see also Milbrodt (1985) for its use in the asymptotic theory of sampling from finite populations]. Within this setup, we suggest tests of stationarity in the mean of autoregressive time series versus fairly general trend alternatives in Section 3. Their asymptotic power will be derived from our LAW results. The proposed tests are similar to the classical tests of goodness of fit of a single univariate cdf based on comparison of the empirical distribution function edf with the theoretical distribution function. To every such edf test there corresponds a test of stationarity in the mean of AR(p) models. These tests may be employed to decide whether heuristic methods to generate stationary data, such as differencing [see, e.g., Anderson (1971), Section 3.4, or Wei (1989), Section 4], have been successful with certain data. In contrast to Epps (1988), who suggested χ^2-type tests of stationarity of Gaussian processes, we require some knowledge of the model structure (though not that it be Gaussian), but no prior knowledge of possible change points. This makes our approach inherently nonparametric. A recent paper dealing with a somewhat related problem is Barry and Hartigan (1990), where an omnibus test for departures from constant mean in a nonparametric normal regression model is investigated.

Now, let us describe the situation dealt with throughout. Fix $p \in \mathbb{N}$ and let $\Theta \subset \mathbb{R}^p$ denote the open subset of those ϑ for which $\vartheta_p \neq 0$ and the polynomial $\sum_{i=0}^p \vartheta_i z^i$ ($\vartheta_0 := 1$) has only zeros outside the closed complex unit disk ("stationarity assumption"). Let $(\varepsilon_t)_{t \in \mathbf{Z}}$ denote a white noise process consisting of centered, square integrable i.i.d. variables. It is well known that for every $\vartheta \in \Theta$ the stochastic difference equation

$$\sum_{i=0}^{p} \vartheta_i V_{t-i} = \varepsilon_t, \qquad t \in \mathbb{Z}, \vartheta_0 := 1 \tag{1.1}$$

has a unique strictly stationary solution $(V_t)_{t \in \mathbf{Z}}$, which can be represented as an L_2-convergent MA(∞) series

$$V_t = \sum_{r=0}^{\infty} \delta_r \varepsilon_{t-r}, \qquad t \in \mathbb{Z};$$

hence, V_t is independent of $(\varepsilon_s)_{s>t}$ [Anderson (1971), Section 5.2]. We are interested in the process

$$X_t = m_t + V_t, \qquad t \in \mathbb{N}, \tag{1.2}$$

defined by the autoregression parameter ϑ and an unknown mean function m: $\mathbb{N} \to \mathbb{R}^1$.

2. Local parametrization and local asymptotic Wiener structure. Let $\vartheta \in \Theta$ and $a \in \mathbb{R}^1$ be fixed. We reparametrize our model for the observations (1.2) locally around ϑ and $m \equiv a$. For this, we introduce L_2-generated triangular arrays of regression constants:

$$b_{nt} := \sqrt{n} \int_{(t-1)/n}^{t/n} b(x)\, dx, \qquad 1 \leq t \leq n, n \in \mathbb{N}, b \in L_2[0,1]. \tag{2.1}$$

Every $b \in L_2[0,1]$ generates a sequence of mean functions $(m_n(b))_{n \in \mathbb{N}}$ according to

$$m_n(b) := a + \sum_{t=1}^{n} b_{nt} 1_{\{t\}}.$$

Let $\Theta_n(\vartheta) := \{\tau \in \mathbb{R}^p |\ \vartheta + n^{-1/2}\tau \in \Theta\}$, $n \in \mathbb{N}$. For $b \in L_2[0,1]$, $n \in \mathbb{N}$ and $\tau \in \Theta_n(\vartheta)$ the distribution of $(X_1, \dots, X_n)^T$ under the parameter values $\vartheta + n^{-1/2}\tau$ and $m_n(b)$ is denoted by $P_{\tau, b, n}$. No reference is made to (ϑ, a), the center of localization. Subsequently, we work with the coordinate representation models

$$E_n := \left(\mathbb{R}^{\mathbb{N}}, \mathscr{F}_n, \{P_{\tau,b,n} |\ \tau \in \Theta_n(\vartheta), b \in L_2[0,1]\}\right), \qquad n \in \mathbb{N}, \tag{2.2}$$

where $\mathscr{F}_n$ denotes the sub-σ-field of $\mathscr{B}(\mathbb{R}^{\mathbb{N}})$, generated by the first n coordinates. Within these models, the variables under consideration are the coordinate projections $(x_t)_{t\in\mathbb{N}}$, which correspond to the observations $(X_t)_{t\in\mathbb{N}}$, the centered variables $v_t := x_t - a$, $t \in \mathbb{N}$, and the innovations $e_t := \sum_{i=0}^{p} \vartheta_i v_{t-i}$, $t \geq p+1$, which under $P_{0,0,n}$ are i.i.d. according to the law of ε_1.

Our assumptions on the innovations will be a slight generalization of the Hájek-type assumptions used by Schütt [(1985), Sections 3.2 and 3.4], by Swensen [(1985), (A.2)] and by Kreiss [(1987), (A.1); (1990), (1.5)].

ASSUMPTION 1. The distribution of ε_1 possesses a strictly positive and bounded Lebesgue density f. The shift-parameter family

$$2(f(\cdot - a)/f)^{1/2}, \qquad a \in \mathbb{R}^1,$$

is quadratic-mean differentiable with $L_2(f)$-derivative φ and Fisher information

$$I := \int (\varphi(x))^2 f(x)\, dx < \infty.$$

Let γ denote the autocovariance function of V [(1.1)] and $\Gamma := (\gamma(|i-j|))_{1\leq i,j\leq p}$. For every $n \in \mathbb{N}$, $\tau \in \Theta_n(\vartheta)$, $b \in L_2[0,1]$, we introduce variables

$$z_{nt} := \begin{cases} 0, & 1 \leq t \leq p, \\ -\varphi(e_t)\left(n^{-1/2}\sum_{i=1}^{p}\tau_i v_{t-i} - \sum_{i=0}^{p}\vartheta_i b_{n,t-i}\right), & p+1 \leq t \leq n. \end{cases}$$

The following LAW theorem is the main result of this section.

2.1. THEOREM. *Suppose Assumption* 1 *holds. Then* $(E_n)_{n\in\mathbb{N}}$ *is* LAW *on* $H := \mathbb{R}^p \times L_2[0,1]$ *with covariance*

$$K_s\colon ((\tau_1,b_1),(\tau_2,b_2)) \mapsto I\left(\left(\sum_{i=0}^{p}\vartheta_i\right)^2 \int_0^s b_1(x)b_2(x)\,dx + s\langle\tau_1,\Gamma\tau_2\rangle\right),$$
$$0 \leq s \leq 1,$$

and centered log-likelihoods

$$L_{n,s}(\tau,b)\colon (x_1,\dots,x_n) \mapsto \sum_{t=1}^{[ns]} z_{nt}(x),$$
$$(\tau,b) \in \Theta_n(\vartheta) \times L_2[0,1],\ 0 \leq s \leq 1:$$

For every $(\tau,b) \in H$,

$$\mathscr{L}\left((L_{n,s}(\tau,b))_{0\leq s\leq 1}\middle|P_{0,0,n}\right) \mapsto \mathscr{L}\left(W(K_s((\tau,b),(\tau,b)))\right), \tag{2.3}$$

weakly in $D[0,1]$

(W *a standard Brownian motion on* $[0,1]$), *and*

$$\sup_{0\le s\le 1}\left|\log\frac{dP_{\tau,b,n}|\mathscr{F}_{[ns]}}{dP_{0,0,n}|\mathscr{F}_{[ns]}} - L_{n,s}(\tau,b) + \frac{1}{2}K_s((\tau,b),(\tau,b))\right| \to 0 \qquad (P_{0,0,n}). \tag{2.4}$$

PROOF. Since the proof follows ideas of Schütt (1985) and Swensen (1985), which have also been used in several papers of Kreiss, we shall restrict ourselves to a brief outline. Let $(\tau, b) \in H$ be fixed and $n \in \mathbb{N}$ be such that $\tau \in \Theta_n(\vartheta)$. Put

$$q_{nt} := \frac{dP_{\tau,b,n}|\mathscr{F}_t}{dP_{0,0,n}|\mathscr{F}_t}, \qquad p \le t \le n,$$

and

$$y_{nt} := 2\left(\sqrt{\frac{q_{nt}}{q_{n,t-1}}} - 1\right), \qquad p+1 \le t \le n.$$

In order to establish our theorem, we shall apply Satz 3.1.2 and Satz 3.1.5 of Schütt (1985) [cf. also Swensen (1985), Lemma 1]. Hence, we have to show that:

(i) $(z_{nt}, \mathscr{F}_t)_{1\le t\le n, n\in\mathbb{N}}$ is a square integrable martingale difference array under $(P_{0,0,n})$.

(ii) $\max_{1\le t\le n}|z_{nt}| \to 0$ $(P_{0,0,n})$.

(iii) $\limsup_{n\to\infty} P_{0,0,n}(\sum_{t=1}^n z_{nt}^2) < \infty$.

(iv) $\sum_{t=1}^n P_{0,0,n}(z_{nt}^2 1_{\{|z_{nt}|>C\}}|\mathscr{F}_{t-1}) \to 0$ $(P_{0,0,n})$ for some $C > 0$ $(\mathscr{F}_0 := \{\varnothing, \mathbb{R}^{\mathbb{N}}\})$.

(v) $\sum_{t=1}^{[ns]} P_{0,0,n}(z_{nt}^2|\mathscr{F}_{t-1}) \to K_s((\tau,b),(\tau,b))$ $(P_{0,0,n})$, $0 < s \le 1$.

(vi) $\lim_{n\to\infty} P_{0,0,n}(\sum_{t=p+1}^n (y_{nt} - z_{nt})^2) = 0$.

(vii) $q_{np} \to 1$ $(P_{0,0,n})$.

For a detailed proof of (i)–(vii), see Milbrodt (1990). □

2.2. REMARKS.

(i) Since Θ is an open subset of $\mathbb{R}^p$, the parameter sets of E_n, $n \in \mathbb{N}$, converge to H. Moreover, (H, K_1) is a Hilbert space: For every $\vartheta \in \Theta$ we have $\sum_{i=0}^p \vartheta_i \ne 0$, and the spectral density of V is bounded away from 0 [Neuhaus and Kreiss (1985), Korollar 7.2]. It follows that Γ is strictly positive definite, and that the norm induced by K_1 is equivalent to the sum of the usual Euclidean norms on the factors of H. Using the terminology of Strasser [(1985), Section 80], Theorem 2.1 entails that the experiments $(E_n)_{n\in\mathbb{N}}$ are asymptotically normal on (H, K_1) with central sequence $(L_{n,1})_{n\in\mathbb{N}}$.

(ii) Because of the monotonicity of $s \mapsto K_s((\tau, b), (\tau, b))$ the limiting process in (2.3) is the centered Gaussian process starting at 0 with covariance structure

$$(s_1, s_2) \mapsto K_{s_1 \wedge s_2}((\tau, b), (\tau, b)).$$

(iii) Theorem 2.1 refers to the case that the density f of the innovations is completely known. An analogous result also holds in the unknown variance model, where f is known apart from its variance $\sigma^2 > 0$. If the center of localization is $(\vartheta, \sigma, m \equiv a) \in \Theta \times (0, \infty) \times \mathbb{R}^1$, the modifications required are as follows. Additional assumption $J := \int (1 + x\varphi(x))^2 f(x)\, dx < \infty$; local parameter space $H := \mathbb{R}^p \times \mathbb{R}^1 \times L_2[0, 1]$; covariance formally unchanged, however with a $(p + 1) \times (p + 1)$ matrix Γ which is obtained from the original Γ by placing J/I in position $(p + 1, p + 1)$, filling up with zeros and dividing by σ^2; for $p + 1 \le t \le n$, z_{nt} has to be modified by subtracting $n^{-1/2}\tau_{p+1}(1 + e_t\varphi(e_t))$ and then dividing by σ. The changes required in the proof are now more or less obvious [cf. also Schütt (1985), Section 3.3].

(iv) Although no explicit attempt in this direction has been made yet, it seems to be reasonable to expect an LAN–LAW result similar to Theorem 2.1 also to hold for invertible ARMA models with a nonparametric regression trend. This is suggested by the fact that Swensen's (1985) argument ensuring LAN in the parametric AR case carry over to the parametric MA case [Garel (1989)], and that the general ARMA case will not lead to substantially new problems compared with these two special cases [cf. Hallin, Ingenbleek and Puri (1985), Appendix 2].

Concluding this section, we remark that similar to asymptotic normality, the invariance principle (2.3) also carries over from the hypothesis to contiguous alternatives:

2.3. Corollary. *Suppose Assumption* 1 *holds. Then for every* $(\tau, b) \in H$, $\mathscr{L}((L_{n,s}(\tau, b) - K_s((\tau, b), (\tau, b)))_{0 \le s \le 1} | P_{\tau, b, n}) \to \mathscr{L}(W(K_s((\tau, b), (\tau, b))))$, *weakly, in the Skorohod space* $D[0, 1]$.

Proof. By standard arguments involving Le Cam's third lemma and the Cramér–Wold device, one gets weak convergence of finite-dimensional marginal distributions from Theorem 2.1. Moreover, since the processes converge under the hypothesis, their $(P_{0,0,n})$ distributions are uniformly tight. Using contiguity, this tightness carries over to the alternative $(P_{\tau, b, n})$. □

3. Testing stationarity in the mean of autoregressive time series. In this section, we shall apply the LAW result Theorem 2.1 to the construction of tests for the hypothesis of stationarity in the mean. The observations (1.2) are stationary if and only if the mean function $m_t \equiv a$ is constant. We want to test this hypothesis against alternatives with arbitrary nonconstant mean function and arbitrary autoregressive structure (1.1).

3.1. DISCUSSION. In a certain sense, the following discussion parallels the derivation of omnibus tests of the hypothesis of randomness against the alternative that the observations are independent but not identically distributed in Strasser [(1989), Section 4].

Reparametrizing locally around fixed parameters $(\vartheta, m \equiv a) \in \Theta \times \mathbb{R}^1$, the hypothesis of stationarity in the mean is given by $\mathbb{R}^p \times H_0$, where

$$H_0 := \{b \in L_2[0,1] \mid b \equiv \text{const}\}.$$

The whole regression parameter space is the orthogonal sum

$$L_2[0,1] = H_0 \oplus \{b \in L_2[0,1] \mid \bar{b} = 0\}$$

(here $\bar{b} := \int_0^1 b(x)\,dx$ denotes the mean of the regression function). Since $b \mapsto b - \bar{b}$ is the projection on $H_0^\perp$, it seems reasonable to reject the hypothesis of stationarity if and only if

$$\sup_{b \in \mathscr{H}} \left| L_{n,1}(0, b - \bar{b}) \right| > c_\alpha, \qquad n \in \mathbb{N}, \tag{3.1}$$

where $\mathscr{H} \subset L_2[0,1]$ is "sufficiently large" and the critical value c_α is chosen as to guarantee asymptotic level α. Having in mind Kolmogorov–Smirnov-type tests, we take

$$\mathscr{H} := \{1_{[0,r]} \mid 0 \le r \le 1\}.$$

The tests (3.1) may be rewritten as follows. Let

$$U_{\vartheta,a,n}(s) := \frac{1}{\sqrt{n}} \sum_{i=0}^{p} \vartheta_i \begin{cases} 0, & 0 \le s < \dfrac{p+1}{n}, \\ \displaystyle\sum_{t=p+1}^{[ns]} \varphi(e_t) - s \sum_{t=p+1}^{n} \varphi(e_t), & \dfrac{p+1}{n} \le s \le 1. \end{cases}$$

Then it is not hard to see that

$$\sup_{b \in \mathscr{H}} \left| L_{n,1}(0, b - \bar{b}) - \int_0^1 b(s) U_{\vartheta,a,n}(ds) \right| \le \text{const}(\vartheta) \cdot n^{-1/2} \max_{p+1 \le t \le n} |\varphi(e_t)|.$$

Due to Lindeberg's condition, the right-hand side tends to 0 stochastically under $(P_{0,0,n})$; hence also under contiguous alternatives. This shows that the tests (3.1) are asymptotically equivalent to the tests $\psi_{\vartheta,a,n}$ defined by the critical regions

$$\sup_{0 \le s \le 1} |U_{\vartheta,a,n}(s)| > c_\alpha, \qquad n \in \mathbb{N}.$$

We also consider Cramér–von Mises-type tests $\chi_{\vartheta,a,n}$ given by critical regions of the form

$$\int_0^1 U_{\vartheta,a,n}(s)^2\,ds > c_\alpha'^2, \qquad n \in \mathbb{N}.$$

To evaluate the asymptotic power of these tests, let $(\Omega_0, \mathscr{A}_0, Q_0, (W_0(s))_{0 \le s \le 1})$ denote a standard Brownian bridge on [0, 1].

3.2. THEOREM. *Suppose Assumption* 1 *holds. Then for every* $(\tau, b) \in H$,

$$\lim_{n\to\infty} P_{\tau,b,n}(\psi_{\vartheta,a,n}) = Q_0\left(\sup_{0\le s\le 1}\left|W_0(s) + C(\vartheta)\int_0^s \big(b(r) - \bar{b}\big)\,dr\right| > k_\alpha\right)$$

and

$$\lim_{n\to\infty} P_{\tau,b,n}(\chi_{\vartheta,a,n}) = Q_0\left(\int_0^1\left(W_0(s) + C(\vartheta)\int_0^s \big(b(r) - \bar{b}\big)\,dr\right)^2 ds > k_\alpha'^2\right),$$

where $C(\vartheta) := \sqrt{I}\sum_{i=0}^p \vartheta_i$, $k_\alpha := c_\alpha |C(\vartheta)|^{-1}$ *is the critical value of the two-sided one-sample Kolmogorov–Smirnov test of level* α *and* $k'_\alpha := c'_\alpha|C(\vartheta)|^{-1}$ *is the critical value of the corresponding Cramér–von Mises test.*

PROOF. Let $(\tau, b) \in H$. Employing Lemma 80.11 of Strasser [(1985), page 414], the Cramér–Wold device and the fact that

$$\sup_{0\le s\le 1}\left|L_{n,s}(0,1) - L_{n,1}(0, 1_{[0,s]})\right| \to 0 \qquad (P_{0,0,n})$$

(hence also under contiguous alternatives), we obtain convergence of finite-dimensional marginals under $(P_{\tau,b,n})$:

$$L_{n,s}(0,1) - K_s((0,1),(\tau,b)) \to W\big(C(\vartheta)^2 s\big), \qquad 0 \le s \le 1.$$

The same tightness argument as in the proof of Corollary 2.3 then entails weak convergence in $D[0,1]$. Hence,

$$\mathscr{L}\left(\left(\frac{1}{\sqrt{nI}}\sum_{t=p+1}^{[ns]}\varphi(e_t) - C(\vartheta)\int_0^s b(r)\,dr\right)_{0\le s\le 1}\middle| P_{\tau,b,n}\right) \to \mathscr{L}(W),$$

showing that the $P_{\tau,b,n}$ distributions of

$$C(\vartheta)^{-1}U_{\vartheta,a,n}(s) - C(\vartheta)\int_0^s\big(b(r) - \bar{b}\big)\,dr, \qquad 0 \le s \le 1,$$

converge to $\mathscr{L}(W_0|Q_0)$. □

3.3. REMARKS. The asymptotic power functions given in Theorem 3.2 neither depend on τ nor on $\bar{b}$. On $\{b \in L_2[0,1] |\ \bar{b} = 0\}$ they coincide with the well known asymptotic powers of the corresponding two-sided one-sample edf tests of level α up to the normalizing factor $C(\vartheta)$ [cf., e.g., Milbrodt and Strasser (1990), (6)]. Hence, for every $(\vartheta, a) \in \Theta \times \mathbb{R}^1$:

(i) $(\psi_{\vartheta,a,n})$ and $(\chi_{\vartheta,a,n})$ are strictly asymptotically unbiased at (ϑ, a):

$$\lim_{n\to\infty} P_{\tau,b,n}(\psi_{\vartheta,a,n}) \wedge \lim_{n\to\infty} P_{\tau,b,n}(\chi_{\vartheta,a,n}) > \alpha, \qquad b \not\equiv \bar{b}$$

[cf. Milbrodt and Strasser (1990), Remark 2.3].

(ii) $(\psi_{\vartheta,a,n})$ and $(\chi_{\vartheta,a,n})$ are locally asymptotically admissible at (ϑ, a) [consult Strasser (1985), Example 82.23].

(iii) The curvature of the asymptotic power function at the hypothesis has—up to the normalizing constant $C(\vartheta)$—the same principal components' decomposition as the power of the corresponding two-sided edf test [see Neuhaus (1976) in the CM case and Milbrodt and Strasser (1990) for KS tests].

Evidently, results corresponding to Theorem 3.2 and the preceding remarks also hold for other tests of stationarity in the mean, which reject the hypothesis, if the observed sample path of $U_{\vartheta,a,n}$ is "too-large"—its size being measured by an appropriate seminorm on the Skorohod space $D[0,1]$ [see Shorack and Wellner (1986) for a survey of edf tests which thus may be carried over to our present situation].

The tests $\psi_{\vartheta,a,n}$ and $\chi_{\vartheta,a,n}$, $n \in \mathbb{N}$, constructed in (3.1) still depend on the global parameter (ϑ, a) and the unobservable white noise (e_t). It remains to remove this dependence by estimating the center of localization. To be more explicit, our final variants of the KS tests and the CM tests will be

$$\mathrm{KS}_n := \psi_{T_n,n} \quad \text{and} \quad \mathrm{CM}_n := \chi_{T_n,n}, \qquad n \in \mathbb{N}, \tag{3.2}$$

where $(T_n)_{n\in\mathbb{N}}$ is a sequence of estimates of the reference point (ϑ, a). This construction will work under the following additional conditions:

ASSUMPTION 2. $f \in C_2(\mathbb{R}^1)$, $\dot{\varphi}$ is uniformly continuous and $\dot{\varphi}(\varepsilon_1)$ is integrable.

This assumption of Cramér–Wald type is trivially satisfied in the most important case, that of Gaussian white noise (then $\dot{\varphi}$ is constant). It also holds for centered logistic errors. Double exponential densities, however, satisfy Assumption 1 but not Assumption 2.

ASSUMPTION 3. Under the hypothesis of stationarity in the mean, $T_n = (\hat{\vartheta}_n, \hat{a}_n)^T$, $n \in \mathbb{N}$, is a $\sqrt{n}$-consistent sequence of estimators of $(\vartheta, m \equiv a)$: For every $(\vartheta, a) \in \Theta \times \mathbb{R}^1$ the distributions $\mathscr{L}(\sqrt{n}(T_n - (\vartheta, a)^T)|P_{0,0,n})$, $n \in \mathbb{N}$, are uniformly tight; that is,

$$\sqrt{n}\left(T_n - (\vartheta, a)^T\right) = O_{P_{0,0,n}}(1).$$

Note that we neither require efficiency nor asymptotic normality of (T_n). Of course, Assumption 3 holds if $\sqrt{n}(\hat{\vartheta}_n - \vartheta)$ and $\sqrt{n}(\hat{a}_n - a)$ both satisfy a clt. In particular, we may use the sample mean $\hat{a}_n := (1/n)\sum_{t=1}^{n} x_t$ [cf. Brockwell and Davis (1987), Theorem 7.1.2] together with Yule–Walker estimators or maximum likelihood estimators $\hat{\vartheta}_n$, $n \in \mathbb{N}$ [see Brockwell and Davis (1987), Chapter 8 and Anderson (1971), Section 5.5], or the least absolute deviations estimators discussed in Bloomfield and Steiger (1983) and Dunsmuir and Spencer (1991).

3.4. THEOREM. *Under Assumptions* 1–3, *for every* $(\vartheta, a) \in \Theta \times \mathbb{R}^1$ *the pairs of sequences of tests*

$$(3.3)\quad \begin{aligned} \psi_{\vartheta,a,n} &= 1_{\{\sup_{(p+1)/n \le s \le 1}(1/\sqrt{nI})|\sum_{t=p+1}^{[ns]}\varphi(e_t) - s\sum_{t=p+1}^{n}\varphi(e_t)| > k_a\}}, & n \in \mathbb{N},\\ \mathrm{KS}_n &:= 1_{\{\sup_{(p+1)/n \le s \le 1}(1/\sqrt{nI})|\sum_{t=p+1}^{[ns]}\varphi(\hat{e}_{nt}) - s\sum_{t=p+1}^{n}\varphi(\hat{e}_{nt})| > k_a\}}, & n \in \mathbb{N} \end{aligned}$$

and

$$(3.4)\quad \begin{aligned} \chi_{\vartheta,a,n} &= 1_{\{(1/nI)\int_{(p+1)/n}(\sum_{t=p+1}^{[ns]}\varphi(e_t) - s\sum_{t=p+1}^{n}\varphi(e_t))^2\,ds > k_a'^2\}}, & n \in \mathbb{N},\\ \mathrm{CM}_n &:= 1_{\{(1/nI)\int_{(p+1)/n}(\sum_{t=p+1}^{[ns]}\varphi(\hat{e}_{nt}) - s\sum_{t=p+1}^{n}\varphi(\hat{e}_{nt}))^2\,ds > k_a'^2\}}, & n \in \mathbb{N} \end{aligned}$$

[$\hat{e}_{nt} := \sum_{i=0}^{p}\hat{\vartheta}_{ni}(x_{t-i} - \hat{a}_n)$, $\hat{\vartheta}_{n0} := 1$, *"estimated white noise"*] *are both asymptotically stochastically equivalent with respect to* $(P_{0,0,n})$.

3.5. REMARKS. Before entering the proof of Theorem 3.4, let us point out some of its consequences.

(i) To implement our KS test and the CM test proposed in (3.2) for a given level α and a given choice of φ satisfying Assumptions 1 and 2 proceed as follows:

- Look up critical values k_a and k_a' in tables for the usual two-sided one-sample KS test and CM test, respectively. If the sample size is small (≤ 50), simulated critical values may be used.
- Use your favorite $\sqrt{n}$-consistent estimates $\hat{a}_n$ of the sample mean and $\hat{\vartheta}_n$ of the structural parameters to calculate estimated residuals

$$\hat{e}_{nt} := \sum_{i=0}^{p}\hat{\vartheta}_{ni}(x_{t-i} - \hat{a}_n), \qquad t = p+1, \dots, n.$$

- Plug these critical values and estimated residuals in formulae (3.3) and (3.4) for KS_n and CM_n to decide whether the time series is stationary in the mean. Of course, for the evaluation of KS_n it suffices to calculate the sup over $s = (p+1)/n, (p+2)/n, \dots, 1$. The CM_n-test statistic is asymptotically equivalent to

$$\frac{1}{nI}\left(\frac{1}{n}\sum_{k=p+1}^{n-1}\left(\sum_{t=p+1}^{k}\varphi(\hat{e}_{nt})\right)^2 - \frac{2}{3}\left(\sum_{t=p+1}^{n}\varphi(\hat{e}_{nt})\right)^2 + \frac{1}{n^2}\left(\sum_{t=p+1}^{n}\varphi(\hat{e}_{nt})\right)\left(\sum_{u=p+1}^{n-1}\varphi(\hat{e}_{nu})u^2\right)\right)$$

as can be seen be lengthy but elementary calculations.

(ii) For reasons of contiguity, the stochastic equivalences claimed in Theorem 3.4 carry over to arbitrary local alternatives $(P_{\tau,b,n})$. Hence, for every $(\vartheta, a) \in \Theta$, the asymptotic power functions of $(\psi_{\vartheta,a,n})$ and KS_n as well as those of $(\chi_{\vartheta,a,n})$ and CM_n near (ϑ, a) coincide, showing that assertions

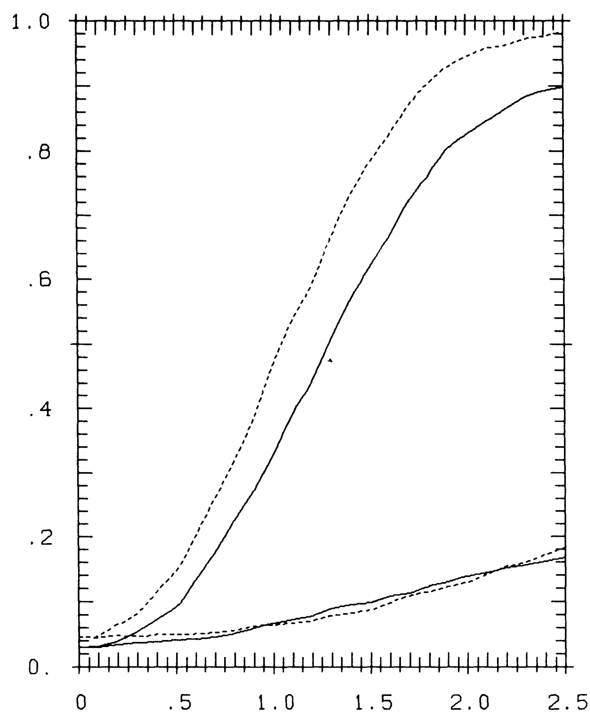

FIG. 1. *Power of KS_{100} (solid) and CM_{100} (dotted), monotone trend (upper curves) and cosine trend (lower curves); 5%-level.*

Theorem 3.2 and Remark 3.3(i)–(iii) remain valid for (KS_n) and (CM_n). In particular, the principal components' decompositions mentioned in Remark 3.3(iii) hold, indicating that both testing procedures are rather sensitive to monotone trend alternatives whereas they rapidly loose their power for oscillating trends with increasing frequency. This is also supported by the following small simulation study.

(iii) Our simulations have been done for the AR(2) model with standard Gaussian white noise and structural parameters $\vartheta_1 = -1.3$ and $\vartheta_2 = 0.6$. The levels were $\alpha = 1\%$ and $\alpha = 5\%$, and the sample sizes $n = 75$, $n = 100$ and $n = 150$. Figures 1 and 2 refer to the 5% level and the sample size $n = 100$. Figure 1 shows simulated powers of KS_{100} (solid) and CM_{100} (dotted) based on 5000 Monte Carlo replicas, and for directions of alternatives determined by the first two principal components of the asymptotic power of the CM test. To be precise, the trend alternatives considered were

$$m_{A,\nu}(t) = A * \cos\left(\frac{\nu\pi}{100}t\right), \qquad t = 1,\dots,100,$$

with amplitude $A = 0.1; 0.2; \dots; 2.5$ varying for each curve and frequency

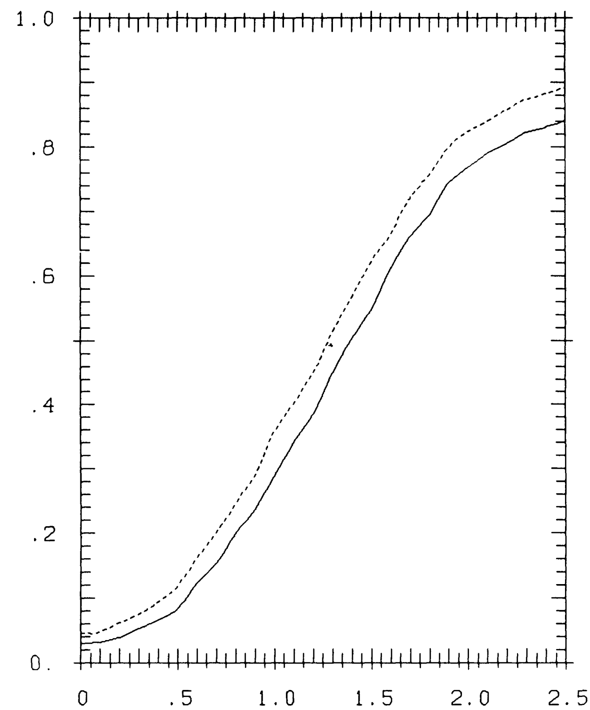

FIG. 2. *Power of KS_{100} (solid) and CM_{100} (dotted), sine trend; 5%-level.*

parameter ν fixed: $\nu = 1$ (upper curves) and $\nu = 2$ (lower curves). The test statistics were calculated using the sample mean and Yule–Walker estimators. Both tests show a very good performance in case of the monotone trend alternative ($\nu = 1$), and they still work for $\nu = 2$, whereas their power against cosine trends, which oscillate stronger ($\nu = 3$, $\nu = 4, \dots$), is extremely poor. Also, they tend to be slightly conservative. The CM test is a little bit superior to KS in the situations studied here (the reason being that the cosine functions are the exact principal components of the asymptotic power of CM, but only approximations in case of KS). The behaviour of both tests for the 1% level and for the sample sizes $n = 75$ and $n = 150$ is *qualitatively* the same. To illustrate that the power for periodic trend alternatives is not only a matter of frequency but also strongly depends on their phase, Figure 2 exhibits simulated powers of KS_{100} (solid) and CM_{100} (dotted) against alternatives $t \mapsto A * \sin((\pi/50)t)$; compare with the two lower curves in Figure 1.

This simulation experiment is "preliminary" in the sense that we did not investigate the effect the choice of the estimators $T_n = (\hat{\vartheta}_n, \hat{a}_n)^T$ has on the finite sample behaviour of the tests. Additionally, it would be interesting to examine and compare the small sample performance of our CM tests and our KS tests.

PROOF OF THEOREM 3.4. Let $(\vartheta, a) \in \Theta \times \mathbb{R}^1$ be fixed. For every $0 \le s \le 1$, a Taylor expansion around (θ, a) gives

$$\frac{1}{\sqrt{n}}\left(\sum_{t=p+1}^{[ns]} (\varphi(\hat{e}_{nt}) - \varphi(e_t)) - s\sum_{t=1}^{n} (\varphi(\hat{e}_{nt}) - \varphi(e_t))\right)$$

$$= \sum_{i=1}^{p} \sqrt{n}\left(\hat{\vartheta}_{ni} - \vartheta_i\right)\left(\frac{1}{n}\sum_{t=p+1}^{[ns]} \dot{\varphi}(e_t)v_{t-i} - \frac{s}{n}\sum_{t=p+1}^{n} \dot{\varphi}(e_t)v_{t-i}\right)$$

$$-\left(\sum_{i=0}^{p} \vartheta_i\right)\sqrt{n}\left(\hat{a}_n - a\right)\left(\frac{1}{n}\sum_{t=p+1}^{[ns]} \dot{\varphi}(e_t) - \frac{s}{n}\sum_{t=p+1}^{n} \dot{\varphi}(e_t)\right)$$

$$+\frac{1}{\sqrt{n}}\left(\sum_{t=p+1}^{[ns]} r_{nt} - s\sum_{t=p+1}^{n} r_{nt}\right),$$

with remainders

$$r_{nt} = \langle \mathrm{grad}|_{\vartheta_n^*, a_n^*}\varphi(e_t) - \mathrm{grad}|_{\vartheta, a}\varphi(e_t), T_n - (\vartheta, a)^T \rangle,$$

where for some $\zeta_{nt} \in [0, 1]$,

$$(\vartheta_n^* - \vartheta, a_n^* - a) = \zeta_{nt}\big(T_n^T - (\vartheta, a)\big), \qquad t = p+1, \dots, n.$$

We show that $\sup_{0 \le s \le 1}| \cdots |$ of the right-hand side tends to 0 $(P_{0,0,n})$-stochastically. For this, we use the following assertion:

Let $(Y_n)_{n \in \mathbb{N}}$ be a stationary shift-ergodic sequence of integrable random variables on some probability space $(\Omega, \mathscr{A}, P)$. Then

$$\sup_{0 \le s \le 1}\left|\frac{1}{n}\sum_{k=1}^{[ns]} Y_k - \frac{s}{n}\sum_{k=1}^{n} Y_k\right| \to 0, \qquad P\text{-a.e.}$$

(Splitting up $\sup_{0 \le s \le 1} \cdots = \sup_{0 \le s \le c} \cdots \vee \sup_{c \le s \le 1} \cdots$ and first letting $n \to \infty$ and then $c \searrow 0$, this follows from the ergodic theorem.)

Employing this fact and the consistency Assumption 3, we obtain that $\sup_{0 \le s \le 1}| \cdots |$ of the first two summands in the expansion tends to 0 $(P_{0,0,n})$. As to the remainders, the claim will follow from Assumption 3 and the Cauchy–Schwarz inequality, once

$$(3.5) \qquad \frac{1}{n}\sum_{t=p+1}^{n} \left\|\mathrm{grad}|_{\vartheta_n^*, a_n^*}\varphi(e_t) - \mathrm{grad}|_{\vartheta, a}\varphi(e_t)\right\| \to 0 \qquad (P_{0,0,n})$$

is established. However, using Assumptions 2 and 3, it is more routine to verify (3.5). □

Acknowledgments. I would like to thank an Associate Editor and the referees for helpful suggestions leading to a reorganization of the paper and to various other improvements.

REFERENCES

AKRITAS, M. G. and JOHNSON, R. A. (1982). Efficiencies of tests and estimators for p-order autoregressive processes when the error distribution is nonnormal. *Ann. Inst. Statist. Math. Ser. A* **34** 579–589.

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

BARRY, D. and HARTIGAN J. A. (1990). An omnibus test for departures from constant mean. *Ann. Statist.* **18** 1340–1357.

BLOOMFIELD, P. and STEIGER, W. L. (1983). *Least Absolute Deviations: Theory, Applications and Algorithms*. Birkhäuser, Boston.

BROCKWELL, P. J. and DAVIS, R. A. (1987). *Time Series: Theory and Methods*. Springer, New York.

DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York.

DUNSMUIR, W. T. M. and SPENCER, N. M. (1991). Strong consistency and asymptotic normality of $l1$ estimates of the autoregressive moving-average model. *J. Time Ser. Anal.* **12** 95–104.

EPPS, T. W. (1988). Testing that a Gaussian process is stationary. *Ann. Statist.* **16** 1667–1683.

GAREL, B. B. (1989). The asymptotic distribution of the likelihood ratio for M. A. processes with a regression trend. *Statist. Decisions* **7** 169–184.

HÁJEK, J. and ŜIDÁK, Z. (1967). *Theory of Rank Tests*. Academic, New York.

HALLIN, M., INGENBLEEK, J.-F. and PURI, M. L. (1985). Linear serial rank tests for randomness against ARMA alternatives. *Ann. Statist.* **13** 1156–1181.

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

KREISS, J.-P. (1987a). On adaptive estimation in autoregressive models when there are nuisance functions. *Statist. Decisions* **5** 59–76.

KREISS, J.-P. (1987b). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15** 112–133.

KREISS, J.-P. (1988). Asymptotic statistical inference for a class of stochastic processes. Habilitationsschrift, Univ. Hamburg.

KREISS, J.-P. (1989). Testing linear hypotheses in ARMA-models. Preprint, Univ. Hamburg.

KREISS, J.-P. (1990). Local asymptotic normality for autoregression with infinite order. *J. Statist. Plann. Inference* **26** 185–219.

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

MILBRODT, H. (1985). Local asymptotic normality of sampling experiments. *Z. Wahrsch. Verw. Gebiete* **70** 131–156.

MILBRODT, H. (1990). Testing stationarity in the mean of autoregressive models with a non-parametric regression trend. Preprint, Fachbereich Mathematik der Univ. Dortmund.

MILBRODT, H. and STRASSER, H. (1990). On the asymptotic power of the two-sided Kolmogorov–Smirnov test. *J. Statist. Plann. Inference* **26** 1–23.

MILLAR, P. W. (1982). Optimal estimation of a general regression function. *Ann. Statist.* **10** 717–740.

MOUSSATAT, M. W. (1976). On the asymptotic theory of stastistical experiments and some of its applications. Ph.D. dissertation, Univ. California, Berkeley.

NEUHAUS, G. (1976). Asymptotic power properties of the Cramér–von Mises test under contiguous alternatives. *J. Multivariate Anal.* **6** 95–110.

NEUHAUS, G. and KREISS, J.-P. (1985). Einführung in die Zeitreihenanalyse. *Skriptum zur Mathematischen Statistik* **10**. Univ. Münster.

SCHÜTT, W. (1985). Einige Untersuchungen über das asymptotische Verhalten der Log-Likelihood-Quotienten Statistik. Dissertation, Univ. Hamburg.

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

SO, Y. C. and SEN, P. K. (1981). Repeated significance tests based on likelihood ratio statistics. *Commun. Statist. A—Theory Methods* **10** 2149–2176.

STRASSER, H. (1985). *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory*. de Gruyter, Berlin.

STRASSER, H. (1989). Tangent vectors for models with independent but not identically distributed observations. *Statist. Decisions* **7** 127–152.

SWENSEN, A. R. (1985). The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *J. Multivariate Anal.* **16** 54–70.

WEI, W. S. (1989). *Time Series Analysis*. Addison-Wesley, Reading, Mass.

FACHBEREICH MATHEMATIK
UNIVERSITÄT DORTMUND
POSTFACH 50 05 00
W-4600 DORTMUND 50
GERMANY