The Annals of Statistics
1987, Vol. 15, No. 1, 112–133

ON ADAPTIVE ESTIMATION IN STATIONARY ARMA PROCESSES

BY JENS-PETER KREISS

University of Hamburg

We consider the estimation problem for the parameter ϑ_0 of a stationary ARMA(p, q) process, with independent and identically, but not necessary normally distributed errors. First we prove local asymptotic normality (LAN) for this model. Then we construct locally asymptotically minimax (LAM) estimators, which asymptotically achieve the smallest possible covariance matrix. Utilizing these, we finally obtain strongly adaptive estimators, by using usual kernel estimators for the score function $\dot{\varphi} = -f'/2f$, where f denotes the density of the error distribution. These estimates turn out to be asymptotically optimal in the LAM sense for a wide class of symmetric densities f.

1. Introduction. In this paper we consider stochastic processes $(X_t;\ t \in Z)$, $Z = \{0, \pm 1, \pm 2, \dots\}$, with discrete time. We assume that these processes satisfy the following difference equation:

$$X_t = a_1 X_{t-1} + \cdots + a_p X_{t-p} + e_t + b_1 e_{t-1} + \cdots + b_q e_{t-q}, \tag{1.1}$$

for all integers t. Here the random variables e_t form a sequence of independent and identically distributed observations with zero mean and finite variance $\sigma^2 > 0$. Such processes, called ARMA(p, q) models [autoregressive moving average models of order (p, q)], are well known in the literature, cf. Fuller (1976) and other monographs on the subject.

In the usual ARMA situation, observations of the process X_t are available and one relevant issue is the estimation of the underlying parameter $\vartheta = (a, b) = (a_1, \dots, a_p, b_1, \dots, b_q) \in \mathbb{R}^{p+q}$. On this subject there exist a lot of results, concerning consistency and central limit theorems for the proposed estimators. Often the estimates are constructed under the assumption that the errors $(e_t;\ t \in Z)$ are normally distributed, while the asymptotic results hold for quite general distributional shapes.

The aim of this paper is threefold.

First, we want to obtain an optimality criterion for sequences of estimates in the ARMA situation. For this purpose we prove in a first main part that the concept of local asymptotic normality (LAN) is applicable to ARMA models (Theorem 3.1). This concept goes back to Le Cam (1960); for a definition see also Fabian and Hannan (1982). Successful criteria for a statistical model to be LAN are contained in Roussas (1979) and Swensen (1985). Since a very appropriate concept of asymptotic efficiency of estimators in LAN models, called local

Received July 1984; revised April 1986.

AMS 1980 *subject classifications*. Primary 62F35, 62M09; secondary 62E20, 62F12.

Key words and phrases. ARMA process, log-likelihood ratio, locally asymptotically normal, locally asymptotically minimax, adaptive estimates, kernel estimators.

asymptotic minimax (LAM), was introduced by Hájek (1972) and fully exploited by Fabian and Hannan (1982), we use this as an optimality criterion.

Second, we present one possible construction of such LAM estimates if only $\sqrt{n}$-consistent initial estimators are available. For this a paper of Beran (1976) is useful. Unfortunately, these estimates appear to depend on the distribution of the white-noise process $(e_t;\ t \in Z)$.

That is why our third aim is to give estimates which are asymptotically equivalent to LAM estimates up to order $1/\sqrt{n}$ *and* which do not depend on the distribution of the white noise. Such estimators are called *adaptive* (in the strong sense). This last step contains generalizations of the results of Bickel (1982) in several directions. Bickel treats independent observations and of course that is why his proofs cannot easily be carried over to the problem considered here. While a proof of the LAN property for AR(p) models [i.e., ARMA($p, 0$) models] is contained in a paper of Akritas and Johnson (1982), the construction of adaptive sequences of estimators in the model considered herein is the main new result of the present paper. As can be seen from a small Monte Carlo study at the end of Section 5, practical application of the recommended adaptive procedure is not immediately excluded.

To make the paper more inviting to read all technical and complicated proofs are given in Section 6.

2. Notation and assumptions. First, we have to ensure that stationary solutions $(X_t,\ t \in Z)$ of (1.1) exist and that these solutions are invertible, cf. Fuller (1976), Section 2.7. This is the case if we assume that the parameter space $\theta \subset \mathbb{R}^{p+q}$ is chosen in such a way that the polynomials $A(z) = 1 - a_1 z - \cdots - a_p z^p$ and $B(z) = 1 + b_1 z + \cdots + b_q z^q$ have no zeros with magnitude less or equal to one. These latter conditions are usually denoted by *stationarity* and *invertibility* conditions. Additionally, we assume that these polynomials have no zeros in common and $a_p \neq 0$ or $b_q \neq 0$. Further let us assume:

(A.1) The distribution of the zero mean random variable e_t possesses an absolute continuous Lebesgue density f, $f(x) \neq 0$, for all $x \in \mathbb{R}$, with finite Fisher information $I(f) = \int (f'/f)^2 f\, d\lambda$. Moreover, $0 < \sigma^2 := \int x^2 f(x)\, dx < \infty$.

If $(e_{1-q}, \dots, e_0; X_{1-p}, \dots, X_0; X_1, \dots, X_n)$ denotes a sample of the stochastic process, in which, for the sake of simplicity, we use an initial part of observations of the white-noise process, then the following property is assumed to be satisfied.

(A.2) The common distribution of $(e_{1-q}, \dots, e_0; X_{1-p}, \dots, X_n)$ possesses for all $n \in \mathbb{N} = \{0, 1, 2, \dots\}$ a nowhere vanishing Lebesgue density $g_n(\cdot\,; \vartheta)$, where $\vartheta \in \theta$ is the underlying parameter.

The density of the distribution of $(e_{1-q}, \dots, e_0; X_{1-p}, \dots, X_n)$ also can be expressed in the form

$$g_0(e_{1-q}, \dots, X_0; \vartheta) \prod_{t=1}^{n} f\big(e_t\{e_{1-q}, \dots, X_t\}\big), \tag{2.1}$$

where $e_t\{e_{1-q}, \dots, X_t\}$ denotes the residual calculated from (1.1). For these calculated residuals we derive an explicit expression in order to obtain a more manageable representation of the likelihood function. To do so we need the following notation.

NOTATION 2.1. For each $\vartheta \in \theta$ there exists $\eta > 1$, such that we have for all $z \in D_\eta(0) = \{w \in \mathbb{C} \mid |w| < \eta\}$ the following power series expansions:

(i)
$$\left(1 + b_1 z + \cdots + b_q z^q\right)^{-1} = \sum_{k=0}^{\infty} \beta_k z^k,$$
$$\left(1 - a_1 z - \cdots - a_p z^p\right)^{-1} = \sum_{k=0}^{\infty} \alpha_k z^k,$$
$$\frac{1 + b_1 z + \cdots + b_q z^q}{1 - a_1 z - \cdots - a_p z^p} = \sum_{k=0}^{\infty} \rho_k z^k = \left(\sum_{k=0}^{\infty} \delta_k z^k\right)^{-1}.$$

Because of this the time series $\{X_t\}$ has the following properties [cf. Ash and Gardner (1975), Sections 2.3.4–2.3.7]:

(ii)
$$\sum_{k=0}^{\infty} \beta_k X_{t-k} = \sum_{k=0}^{\infty} \alpha_k e_{t-k},$$
$$e_t = \sum_{k=0}^{\infty} \rho_k X_{t-k} \quad \textit{and} \quad X_t = \sum_{k=0}^{\infty} \delta_k e_{t-k}, \qquad t \in Z.$$

Since $\sum_{k=0}^{\infty} |\beta_k| < \infty$ and $E|X_t|$ is finite, the series $\sum_{k=0}^{\infty} \rho_k X_{t-k}$ is almost surely absolutely convergent; see Lukacs (1968), Theorem 4.2.1. Of course, the same result holds true for the remaining series.

For the power series coefficients $\{\beta_k;\ k \geq 0\}$ of $(1 + b_1 z + \cdots + b_q z^q)^{-1}$ the following recursion formula can easily be derived:

$$\beta_s + b_1 \beta_{s-1} + \cdots + b_q \beta_{s-q} = 0, \quad \text{for all } s \geq 1 \tag{2.2}$$

(note $\beta_t = 0$ if $t < 0$ and $\beta_0 = 1$). The above Notation 2.1 together with (1.1) and (2.2) leads to (note $a_0 = -1$, $b_0 = 1$)

$$\begin{aligned} e_j &= \sum_{k=0}^{\infty} \beta_k \left(X_{j-k} - a_1 X_{j-k-1} - \cdots - a_p X_{j-k-p}\right) \\ &= \sum_{k=1}^{j} \beta_{k-1}\left(-\sum_{i=0}^{p} a_i X_{j+1-k-i}\right) + \sum_{k=j+1}^{\infty} \beta_{k-1}\left(\sum_{i=0}^{q} b_i e_{j+1-k-i}\right) \\ &= \sum_{k=1}^{j} \beta_{k-1}\left(-\sum_{i=0}^{p} a_i X_{j+1-k-i}\right) + \sum_{s=0}^{q-1} e_{-s}\left(\sum_{k=0}^{s} \beta_{j+s-k} b_k\right), \end{aligned} \tag{2.3}$$

$$j \geq 1 - q.$$

The second equality gives the explicit expression for the calculated residuals used in (2.1). For each $\vartheta \in \theta$ we can denote them more exactly by $e_j(\vartheta)$. Also the abbreviation $e_j(\vartheta_i) = e_j^i$, $i = 0, 1$, will be used. If we designate the distribution of $(e_s,\ s \leq 0;\ X_{s-p},\ s \geq 1)$ on $\mathbb{R}^Z$ by P_ϑ, when ϑ is the underlying parameter and the restriction of P_ϑ to $\mathscr{A}_j = \sigma(e_{1-q}, \dots, e_0; X_{1-p}, \dots, X_j)$ by $P_{j,\vartheta}$, then, for $\vartheta_0, \vartheta_1 \in \theta$ the likelihood ratio has the following form [cf. (2.1) and (2.3)]

$$(2.4) \qquad \frac{dP_{n,\vartheta_1}}{dP_{n,\vartheta_0}} = \frac{g_0(e_{1-q}, \dots, X_0; \vartheta_1)}{g_0(e_{1-q}, \dots, X_0; \vartheta_0)} \prod_{j=1}^{n} \frac{f\left(e_j^0 - \left(e_j^0 - e_j^1\right)\right)}{f\left(e_j^0\right)}.$$

In order to prove LAN we need derivatives of the log-likelihood ratio. To this end, we derive a simple expression for $e_j(\vartheta_0) - e_j(\vartheta)$, $\vartheta \in \theta$, which seems to be useful in other situations, too.

Lemma 2.2. *With*

$$\begin{aligned} Z(j-1; \vartheta, \vartheta_0) &= \sum_{k=1}^{j} \beta_{k-1}(\vartheta)\left(X_{j-k}, \dots, X_{j-k+1-p}; e_{j-k}^0, \dots, e_{j-k+1-q}^0\right)^T \\ &= \sum_{k=1}^{j} \beta_{k-1}\left(Y^T(j-k); E(j-k; \vartheta_0)^T\right)^T, \end{aligned}$$

$$(2.5) \qquad e_j(\vartheta_0) - e_j(\vartheta) = (\vartheta - \vartheta_0)^T Z(j-1; \vartheta, \vartheta_0)$$

holds true.

Proof. From the definition of $e_j(\vartheta)$, cf. (2.3), we obtain

$$\begin{aligned} e_j(\vartheta_0) - e_j(\vartheta) = (a - a_0)^T &\sum_{k=1}^{j} \beta_{k-1}(\vartheta) Y(j-k) \\ &+ \sum_{k=1}^{j} \left(\beta_{k-1}(\vartheta_0) - \beta_{k-1}(\vartheta)\right) \sum_{i=0}^{p} a_i^0 X_{j+1-k-i} \\ &+ \sum_{s=0}^{q-1} e_{-s} \left\{ \sum_{k=0}^{s} \beta_{j+s-k}(\vartheta_0) b_k^0 - \sum_{k=0}^{s} \beta_{j+s-k}(\vartheta) b_k \right\}. \end{aligned}$$

Since for all $\vartheta_0 \in \theta$ $\sum_{i=0}^{p} a_i^0 X_{t-i} = \sum_{i=0}^{q} b_i^0 e_{t-i}(\vartheta_0)$, $t \geq 1$, and from the recursion formula (2.2),

$$\begin{aligned} &\sum_{k=1}^{j} \beta_{k-1}(\vartheta)\left(e_{j+1-k}(\vartheta_0) + b_1 e_{j-k}(\vartheta_0) + \cdots + b_q e_{j+1-q-k}(\vartheta_0)\right) \\ &\qquad = e_j(\vartheta_0) - \sum_{s=0}^{q-1} e_{-s} \left(\sum_{k=0}^{s} \beta_{j+s-k}(\vartheta) b_k \right), \end{aligned}$$

we have the desired result. □

From (2.4) and (2.5) we get

$$(2.6) \qquad \frac{dP_{n,\vartheta}}{dP_{n,\vartheta_0}} = \frac{g_0(e_{1-q}, \dots, X_0; \vartheta)}{g_0(e_{1-q}, \dots, X_0; \vartheta_0)} \prod_{j=1}^{n} \frac{f\left(e_j^0 - (\vartheta - \vartheta_0)^T Z(j-1; \vartheta, \vartheta_0)\right)}{f\left(e_j^0\right)}.$$

With the following additional abbreviation,

$$\phi_j^2(\vartheta_0, \vartheta) = \frac{f\big(e_j(\vartheta_0) - (\vartheta - \vartheta_0)^T Z(j-1; \vartheta, \vartheta_0)\big)}{f\big(e_j(\vartheta_0)\big)}, \tag{2.7}$$

we have

$$\log \frac{dP_{n,\vartheta}}{dP_{n,\vartheta_0}} = \log \frac{g_0(e_{1-q}, \dots, X_0; \vartheta)}{g_0(e_{1-q}, \dots, X_0; \vartheta_0)} + 2 \sum_{j=1}^{n} \log \phi_j(\vartheta_0, \vartheta). \tag{2.8}$$

After these preliminaries, we are now ready to establish local asymptotic normality for the above likelihood ratio.

3. Local asymptotic normality. This section of the paper is devoted to local asymptotic normality for ARMA processes. In order to establish this property for our model we will verify assumptions (A.1)–(A.4) of Roussas (1979), who considers arbitrary stochastic processes. Similar conditions sufficient for the LAN property are given in Swensen (1985). Swensen's conditions are also valid in the case considered here. For this section we emphasize that both $\mathbf{e}_0 = (e_{1-q}, \dots, e_0)$ and $\mathbf{X}_0 = (X_{1-p}, \dots, X_0)$ are regarded as initial observations. We need the following regularity assumption.

(A.3) $g_0(\mathbf{e}_0, \mathbf{X}_0, \vartheta_n) \to g_0(\mathbf{e}_0, \mathbf{X}_0, \vartheta_0)$, in P_{ϑ_0}-probability if $\vartheta_n \to \vartheta$.

The main result of this section is

THEOREM 3.1 (LAN property for ARMA models). *Let $\{h_n\} \subset \mathbb{R}^{p+q}$ be a bounded sequence and $\vartheta_n = \vartheta_0 + n^{-1/2} h_n$. Under our assumptions* (A.1), (A.2) *and* (A.3) *we have for*

$$\Delta_n(\vartheta) = \frac{2}{\sqrt{n}} \sum_{j=1}^{n} \dot{\varphi}\big(e_j(\vartheta)\big) Z(j-1; \vartheta, \vartheta), \qquad \dot{\varphi} = -f'/2f, \tag{3.1}$$

the following two results:

$$\log\big[dP_{n,\vartheta_n}/dP_{n,\vartheta_0}\big] - h_n^T \Delta_n(\vartheta_0) + \tfrac{1}{2} h_n^T I(f) \Gamma(\vartheta_0) h_n \to 0, \tag{3.2}$$

in P_{n,ϑ_0}-probability, where $\Gamma(\vartheta_0)$ is defined in Theorem 3.5 below (approximation of the log-likelihood ratio).

$$\mathscr{L}\big(\Delta_n(\vartheta_0) | P_{n,\vartheta_0}\big) \Rightarrow \mathscr{N}(0, I(f)\Gamma(\vartheta_0)), \tag{3.3}$$

where " $\Rightarrow$ " denotes weak convergence (asymptotic normality of the approximating statistic).

REMARK. Again we would like to mention that we define LAN the same as Fabian and Hannan (1982), 2. Definition, page 461. This definition is slightly more general than the one given by Hájek (1972).

From the above theorem one can obtain by using standard arguments

COROLLARY 3.2. *Under the same assumptions as above*

(3.4) $\{P_{n,\vartheta_0}\}$ *and* $\{P_{n,\vartheta_n}\}$ *are contiguous in the sense of Definition 2.1, Roussas* (1972), *page* 7,

and

$$\mathscr{L}\big(\Delta_n(\vartheta_0) - I(f)\Gamma(\vartheta_0)h_n | P_{n,\vartheta_n}\big) \Rightarrow \mathscr{N}(0, I(f)\Gamma(\vartheta_0)). \tag{3.5}$$

In what follows we assemble the results which, together with those of Roussas (1979), lead to Theorem 3.1. Since the proofs are rather technical we defer them to Section 6.

First, we consider differentiability in quadratic mean.

THEOREM 3.3. *For each* $\vartheta_0 \in \theta$, *the random functions* $\phi_j(\vartheta_0, \cdot)$ *are differentiable in q.m.* $[P_{\vartheta_0}]$ *uniformly in* $j \geq 1$. *That is, there are* $(p+q)$-*dimensional r.v.'s* $\dot{\phi}_j(\vartheta_0) = \dot{\varphi}(e_j^0)Z(j-1; \vartheta_0, \vartheta_0) = \dot{\varphi}(e_j^0)Z^0(j-1)$ [*the q.m. derivative of* $\phi_j(\vartheta_0, \vartheta)$ *with respect to* ϑ *at* ϑ_0] *such that*

$$\frac{\phi_j(\vartheta_0, \vartheta_0 + \lambda h) - 1}{\lambda} - h^T\dot{\phi}_j(\vartheta_0) \to_{\lambda \to 0} 0, \quad \textit{in q.m.} \left[P_{\vartheta_0}\right], \tag{3.6}$$

uniformly on bounded sets of $h \in \mathbb{R}^{p+q}$ *and uniformly in* $j \in \mathbb{N}$. *Finally,* $\dot{\phi}_j(\vartheta_0)$ *is measurable with respect to* $\mathscr{A}_j$.

Next, we have

THEOREM 3.4. *For each* $\vartheta_0 \in \theta$ *and each* $h \in \mathbb{R}^{p+q}$, *the sequence* $\{(h^T\dot{\phi}_j(\vartheta_0))^2\}$, $j \in \mathbb{N}$, *is uniformly integrable with respect to* P_{ϑ_0}.

The following result deals with the asymptotic covariance matrix of $\Delta_n(\vartheta_0)$.

THEOREM 3.5. *For each* $\vartheta_0 \in \theta$ *and* $j \geq 1$ *let the* $(p+q) \times (p+q)$-*dimensional covariance matrix* $\Gamma_j(\vartheta_0)$ *be defined by*

$$\begin{aligned}\Gamma_j(\vartheta_0) &= 4E_{\vartheta_0}\big[\dot{\phi}_j(\vartheta_0)\dot{\phi}_j^T(\vartheta_0)\big] \\ &= I(f)E_{\vartheta_0}\big[Z(j-1; \vartheta_0, \vartheta_0)Z^T(j-1; \vartheta_0, \vartheta_0)\big].\end{aligned} \tag{3.7}$$

Then $\Gamma_j(\vartheta_0) \to \Gamma(\vartheta_0)I(f)$, *as* $j \to \infty$, *in any one of the standard norms in* $\mathbb{R}^{p+q}$, *and* $\Gamma(\vartheta_0)$ *is positive definite.*

To obtain some insight into the structure of $\Gamma(\vartheta_0)$ consider

EXAMPLE 3.6. An explicit representation of $\Gamma(\vartheta_0)$ is

$$\begin{aligned}\Gamma(\vartheta) = E_\vartheta\Bigg(&\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\beta_{k-1}\beta_{l-1}\begin{pmatrix} Y(1-k) \\ E(1-k; \vartheta_0)\end{pmatrix} \\ &\times\big(Y^T(1-l), E^T(1-l; \vartheta_0)\big)\Bigg).\end{aligned} \tag{3.8}$$

For purely autoregressive schemes, i.e., $q = 0$, we have

$$(3.9)\qquad \Gamma(\vartheta) = (E_\vartheta X_r X_s)_{r,s=1,\ldots,p}.$$

Finally we need some weak laws of large numbers (WLLN) for the approximating functions $\dot{\phi}_j(\vartheta_0)$.

THEOREM 3.7. (i) *For each* $\vartheta_0 \in \theta$, *each* $h \in \mathbb{R}^{p+q}$ *and for the probability measure* P_{ϑ_0}, *the WLLN holds for the sequence* $\{[h^T\dot{\phi}_j(\vartheta_0)]^2,\ j \in \mathbb{N}\}$. *Also*

$$\text{(ii)}\quad \frac{1}{n}\sum_{j=1}^{n}\left\{E_{\vartheta_0}\left[\left(h^T\dot{\phi}_j(\vartheta_0)\right)^2 \middle| \mathscr{A}_{j-1}\right] - \left[h^T\dot{\phi}_j(\vartheta_0)\right]^2\right\} \to 0, \quad \text{as } n \to \infty,$$

in P_{ϑ_0}-*probability. As is shown in Schütt* (1985) *and Swensen* (1985) *we can dispense with this latter assertion.*

Theorems 3.3–3.7 guarantee that the sufficient conditions for local asymptotic normality given in Roussas (1979) are fulfilled.

REMARK. Recall that at the beginning of Section 2 we assumed that initial observations of the white-noise process are available. Of course, this assumption is bothersome, so that we would like to mention that it is possible to dispense with it.

Define

$$\hat{e}_t(\vartheta) := \sum_{k=1}^{t}\beta_{k-1}(X_{t+1-k} - a_1X_{t-k} - \cdots - a_pX_{t+1-k-p})$$

and

$$\hat{\Delta}_n(\vartheta) := \frac{2}{\sqrt{n}}\sum_{j=1}^{n}\dot{\varphi}(\hat{e}_j(\vartheta))\sum_{k=1}^{j}\beta_{k-1}\begin{pmatrix} Y(j-k) \\ \hat{E}(j-k;\vartheta)\end{pmatrix}$$

[compare with (2.3) and (3.1)].

If, for example, $\dot{\varphi}$ satisfies a global Lipschitz condition it can be shown that $E_\vartheta|\Delta_n(\vartheta) - \hat{\Delta}_n(\vartheta)| = o(1)$ holds true (the proof is omitted, since it is not essential). From this we have that Theorem 3.1 remains valid if we replace Δ_n by $\hat{\Delta}_n$, which depends on observations of the time series, only.

Now we are ready to consider statistical inference for stochastic processes of ARMA type.

4. Existence and construction of LAM estimates. From Theorem 3.1 (LAN) we can construct sequences of estimates which are locally asymptotically minimax (LAM) as is defined in Fabian and Hannan (1982), 1. Definition, page 463. But as it turns out below these estimates depend on the distribution of the underlying white noise (e_t).

In the above paper a very useful criterion is given which ensures the LAM property, also in our context, under suitable conditions. More precisely, we have from Fabian and Hannan (1982), 3. Theorem, page 467:

LEMMA 4.1. *Under assumptions* (A.1), (A.2) *and* (A.3) *we have for any sequence* $\{Z_n\}$ *of estimates the following implication*:

$$(4.1)\quad \sqrt{n}(Z_n - \vartheta_0) - \frac{\Gamma(\vartheta_0)^{-1}}{I(f)}\Delta_n(\vartheta_0) = o_{P_{\vartheta_0}}(1) \quad (\{Z_n\} \textit{ is called } \vartheta_0\textit{-regular})$$

implies that $\{Z_n\}$ *is LAM.*

In order to construct regular estimates the existence of $\sqrt{n}$-consistent initial estimators $\{\bar{\vartheta}_n\}$ is essential. That is why we assume

(A.4) There exists a sequence $\{\bar{\vartheta}_n\}$ of estimators which satisfies $\sqrt{n}(\bar{\vartheta}_n - \vartheta_0) = O_{P_{\vartheta_0}}(1)$.

COMMENTS ON ASSUMPTION (A.4). Of course (A.4) holds for estimators for which the usual CLT is valid, i.e., for all the standard estimators. For pure autoregressive models such estimators exist under moment conditions, as is shown in Anderson (1971), Theorem 5.5.7.

In contrast to AR models, the computation of estimators in the ARMA case is more complicated. Nevertheless, (A.4) is satisfied in this situation too, as is stated in Section 8.4 of Fuller (1976).

Now, if $\dot{\varphi}$ is assumed to satisfy

$$(\text{A.5})\quad \begin{aligned} &\text{(i)}\ \lim_{h\to 0}\int\{\dot{\varphi}(x+h) - \dot{\varphi}(x)\}^2 f(x)\,dx = 0,\\ &\text{(ii)}\ \lim_{h\to 0}\int\frac{\dot{\varphi}(x-h) - \dot{\varphi}(x)}{h}f(x)\,dx = -\tfrac{1}{2}I(f), \end{aligned}$$

we can establish

THEOREM 4.2 (Existence of LAM estimators). *Assume* $\{\bar{\vartheta}_n\} \subset \theta$ *is discrete and* $\sqrt{n}$*-consistent for* $\vartheta_0 \in \theta$ (*for the definition of discreteness see below*). *Then* $\hat{\vartheta}_n$ *defined by* (4.2) *and* (4.3) *below is regular* [*cf*. (4.1)]:

$$(4.2)\quad \hat{\vartheta}_n = \bar{\vartheta}_n + \frac{1}{\sqrt{n}}\frac{\hat{\Gamma}_n(\bar{\vartheta}_n)^{-1}}{I(f)}\Delta_n(\bar{\vartheta}_n),$$

$$(4.3)\quad \hat{\Gamma}_n(\vartheta) = \frac{1}{n}\sum_{j=1}^{n} Z(j-1;\vartheta,\vartheta)Z^T(j-1;\vartheta,\vartheta).$$

For technical reasons we restrict ourselves to *discrete* sequences of estimators $\{\bar{\vartheta}_n\}$ such as (assume $\{\vartheta_n\}$ according to (A.4) is given):

$\bar{\vartheta}_n$ is given by one of the vertices of $\{\vartheta: \vartheta = n^{-1/2}(i_1, \dots, i_{p+q}), i_j \in Z\}$ nearest to ϑ_n.

Of course, $\{\bar{\vartheta}_n\}$ satisfies the following more general discreteness property.

DEFINITION 4.3. A sequence $\{\bar{\vartheta}_n\}$ of estimates is called discrete if $K \in \mathbb{N}$ exists such that independently of $n \in \mathbb{N}$, $\bar{\vartheta}_n$ takes on at most K different values in

$$Q_n = \left\{\vartheta \in \mathbb{R}^{p+q}: \sqrt{n}|\vartheta - \vartheta_0| \le c\right\}, \quad c > 0 \text{ fixed.}$$

The great advantage of discrete estimates is the following result, which goes back to Le Cam and is also used, for example, by Bickel (1982).

LEMMA 4.4. *Assume $\{S_n(\vartheta),\ n \in \mathbb{N}\}$ to be a sequence of random variables which depends on $\vartheta \in \theta$. If for each sequence $\{\vartheta_n\} \subset \theta$ satisfying*

(4.4) $\quad \sqrt{n}(\vartheta_n - \vartheta_0)$ *is bounded by a constant* $c > 0$,

we have $S_n(\vartheta_n) = o_{P_{\vartheta_0}}(1)$, then also $S_n(\bar{\vartheta}_n) = o_{P_{\vartheta_0}}(1)$ holds for discrete estimators $\{\bar{\vartheta}_n\}$ which are $\sqrt{n}$-consistent.

PROOF. For $\varepsilon > 0$ there exists $M > 0$ such that

$$P_{\vartheta_0}\left\{\left|S_n(\bar{\vartheta}_n)\right| \ge \varepsilon\right\} \le P_{\vartheta_0}\left\{\left|S_n(\bar{\vartheta}_n)\right| \ge \varepsilon, \sqrt{n}|\bar{\vartheta}_n - \vartheta_0| \le M\right\} + \frac{\varepsilon}{2}.$$

Since $\bar{\vartheta}_n$ achieves in $\{\vartheta: \sqrt{n}|\vartheta - \vartheta_0| \le M\}$ only K_M (say) different values we obtain the desired result from our assumption. □

Finally, we give a proof of Theorem 4.2.

PROOF OF THEOREM 4.2. In order to establish regularity of $\{\hat{\vartheta}_n\}$, cf. (4.2), the following *asymptotic linearity* is essential: Let $\{\vartheta_n\} \subset \theta$ be a sequence with (4.4). Then

$$\Delta_n(\vartheta_n) - \Delta_n(\vartheta_0) + \Gamma(\vartheta_0)I(f)\sqrt{n}(\vartheta_n - \vartheta_0) = o_{P_{\vartheta_0}}(1), \tag{4.5}$$

as is shown in Section 6. From this, Lemma 4.4 and

$$\hat{\Gamma}_n(\bar{\vartheta}_n) \to \Gamma(\vartheta_0), \quad \text{as } n \to \infty \text{ in } P_{\vartheta_0}\text{-probability} \tag{4.6}$$

[see proof of (4.5)] we obtain regularity of $\{\hat{\vartheta}_n\}$. □

REMARK. The regularity of $\{\hat{\vartheta}_n\}$ implies

$$\mathscr{L}\left(\sqrt{n}(\hat{\vartheta}_n - \vartheta_0)|P_{n,\vartheta_0}\right) \Rightarrow \mathscr{N}\left(0, \Gamma(\vartheta_0)^{-1}/I(f)\right), \tag{4.7}$$

where this covariance matrix is minimal for all asymptotically normally distributed estimates, cf. Kabaila (1983), who deals with classical asymptotic efficiency of estimators in the sense of Cramér and Rao.

5. Construction of adaptive estimates. In this section we give an answer to the question whether it is possible to construct estimators which are independent of the distribution of white noise but which are LAM, simultaneously for several types of error distributions. Of course, it is enough to establish regularity, cf. Lemma 4.1, under these densities. Such estimators, if they exist, are called adaptive for the specific class of densities.

To give an answer to the above question we look for estimators of the unknown score function $\dot{\varphi}$ and the Fisher information $I(f)$. Then we define the adaptive estimator similar to $\hat{\vartheta}_n$, cf. (4.2). First, we restrict our consideration to estimators of $\dot{\varphi}$.

The first step in this direction (for autoregressive models) was taken by Beran (1976). However, his resulting estimator for ϑ_0 is only adaptive for a small class of error distribution densities. We shall use instead the usual kernel density estimator for $\dot{\varphi}$.

Introduce the following additional notation:

$$(5.1)\quad \begin{aligned} &\text{(i)} \quad g(x;\sigma) = 1/\sqrt{2\pi\sigma^2}\exp(-x^2/2\sigma^2), \qquad x \in \mathbb{R},\\ &\text{(ii)} \quad f_\sigma(x) = \int g(x-y;\sigma) f(y)\,dy,\\ &\text{(iii)} \quad \hat{f}_{\sigma,j}(x;\vartheta) = \frac{1}{2(n-1)} \sum_{\substack{i=1\\ i\neq j}}^{n} \{g(x+e_i(\vartheta);\sigma) + g(x-e_i(\vartheta);\sigma)\},\\ &\qquad\qquad j = 1,\dots,n. \end{aligned}$$

Define $\hat{q}_{n,j}$ to be the following estimator of $\dot{\varphi}$:

$$(5.2)\quad \hat{q}_{n,j}(x;\vartheta) = \begin{cases} -\dfrac{1}{2}\dfrac{\hat{f}'_{\sigma(n),j}(x;\vartheta)}{\hat{f}_{\sigma(n),j}(x;\vartheta)}, & \text{if } \begin{cases} \hat{f}_{\sigma(n),j}(x;\vartheta) \geq d_n,\\ |x| \leq g_n,\\ \left|\hat{f}'_{\sigma(n),j}(x;\vartheta)\right| \leq c_n \hat{f}_{\sigma(n),j}(x;\vartheta), \end{cases}\\ 0, & \text{otherwise,} \end{cases}$$

with $c_n \to \infty$, $g_n \to \infty$, $\sigma(n) \to 0$, $d_n \to 0$. Let

$$(5.3)\qquad \tilde{\Delta}_n(\vartheta) = \frac{2}{\sqrt{n}} \sum_{j=1}^{n} \hat{q}_{n,j}(e_j(\vartheta);\vartheta) Z(j-1;\vartheta,\vartheta)$$

be the estimated version of $\Delta_n(\vartheta)$, cf. (3.1), and use in analogy to (4.2) ($\hat{I}_n$ is defined below)

$$(5.4)\qquad \tilde{\vartheta}_n = \bar{\vartheta}_n + \frac{1}{\sqrt{n}} \frac{\hat{\Gamma}_n^{-1}(\bar{\vartheta}_n)}{\hat{I}_n} \tilde{\Delta}_n(\bar{\vartheta}_n),$$

with $\{\bar{\vartheta}_n\}$ an initial estimator as in Section 4. The remaining task will be to verify that this estimator is *adaptive* for the following broad class of densities f:

$$\text{(A.6)} \qquad \begin{array}{l} \text{(i) } f \text{ is symmetric about the origin,} \\ \text{(ii) } \int x^4 f(x)\,dx < \infty. \end{array}$$

Note that we use the $\hat{q}_{n,j}(e_j(\bar{\vartheta}_n); \bar{\vartheta}_n)$ as estimators of the $\dot{\varphi}(e_j)$, $j = 1, \dots, n$, which means that the residual $e_j(\bar{\vartheta}_n)$ is not used in the estimation of the score function at that specific point.

For the proof we use ideas of Bickel (1982), who treats the regression case with symmetric errors, i.e., he has independent observations. In contrast to Bickel, we also use the full sample for both estimation procedures, so that practical applications are not immediately excluded. The proof also heavily uses the fact that f is symmetric. Results for unsymmetric but zero mean densities f will be discussed in a subsequent paper [Kreiss (1987)].

The results are as follows.

THEOREM 5.1. *Let $\{\bar{\vartheta}_n\} \subset \theta$ be a discrete and $\sqrt{n}$-consistent sequence of estimators of ϑ_0. Under our assumptions* (A.1)–(A.6)

$$\tilde{\Delta}_n(\bar{\vartheta}_n) - \Delta_n(\bar{\vartheta}_n) = o_{P_{\vartheta_0}}(1)$$

holds, if $c_n \to \infty$, $g_n \to \infty$, $\sigma(n) \to 0$, $d_n \to 0$, $\sigma(n)c_n \to 0$, $g_n\sigma(n)^{-4}/n \to 0$ and $n\sigma(n)^9$ stays bounded.

A proof is given in Section 6. Now we state the central result of the paper.

THEOREM 5.2 (Existence of adaptive estimators). *Under the assumptions of Theorem* 5.1 *we have, given a consistent estimator $\hat{I}_n$ of $I(f)$, that the estimator*

$$\tilde{\vartheta}_n = \bar{\vartheta}_n + \frac{1}{\sqrt{n}} \frac{\hat{\Gamma}_n(\bar{\vartheta}_n)^{-1}}{\hat{I}_n} \tilde{\Delta}_n(\bar{\vartheta}_n) \tag{5.5}$$

satisfies

$$\sqrt{n}\left(\tilde{\vartheta}_n - \vartheta_0\right) - \frac{\Gamma(\vartheta_0)^{-1}}{I(f)} \Delta_n(\vartheta_0) = o_{P_{\vartheta_0}}(1), \tag{5.6}$$

for all f satisfying (A.6), *which ensures that $\{\tilde{\vartheta}_n\}$ is LAM, and we obtain*

$$\mathscr{L}\left(\sqrt{n}\left(\tilde{\vartheta}_n - \vartheta_0\right) \middle| P_{n,\vartheta_0}\right) \Rightarrow \mathscr{N}\left(0, \Gamma(\vartheta_0)^{-1}/I(f)\right), \tag{5.7}$$

for all densities f satisfying (A.6).

PROOF. We only have to prove that $\{\tilde{\vartheta}_n\}$ is regular. To see this consider

$$\sqrt{n}\left(\tilde{\vartheta}_n - \vartheta_0\right) - \frac{\Gamma(\vartheta_0)^{-1}}{I(f)}\Delta_n(\vartheta_0)$$

$$= \sqrt{n}\left(\bar{\vartheta}_n - \vartheta_0\right) + \frac{\Gamma(\vartheta_0)^{-1}}{I(f)}\left(\tilde{\Delta}_n(\bar{\vartheta}_n) - \Delta_n(\vartheta_0)\right) + o_{P_{\vartheta_0}}(1)$$

(because of the consistency of $\hat{I}_n$ and $\hat{\Gamma}_n$)

$$= \frac{\Gamma(\vartheta_0)^{-1}}{I(f)}\left(\Delta_n(\bar{\vartheta}_n) - \Delta_n(\vartheta_0) + \Gamma(\vartheta_0)I(f)\sqrt{n}\left(\bar{\vartheta}_n - \vartheta_0\right)\right) + o_{P_{\vartheta_0}}(1),$$

(from Theorem 5.1)

$$= o_{P_{\vartheta_0}}(1) \quad \left(\text{because of (4.5) and the discreteness of } \bar{\vartheta}_n\right). \qquad \square$$

In defining $\tilde{\vartheta}_n$ in (5.5), we make use of a consistent estimator $\hat{I}_n$ of Fisher information. For this purpose we can use

$$\hat{I}_n = \frac{4}{n}\sum_{j=1}^{n} \hat{q}_{n,j}^2\left(e_j(\bar{\vartheta}_n); \bar{\vartheta}_n\right). \tag{5.8}$$

To see that this is consistent, note that from the WLLN we have for each $\{\vartheta_n\} \subset \theta$ satisfying (4.4),

$$\frac{4}{n}\sum_{j=1}^{n} \dot{\varphi}^2\left(e_j(\vartheta_n)\right) \to I(f), \quad \text{in } P_{\vartheta_n}\text{-probability as } n \to \infty.$$

Thus the assertion follows from Bickel (1982), Lemma 4.1, the contiguity of $\{P_{n,\vartheta_n}\}$ and $\{P_{n,\vartheta_0}\}$ [cf. (3.4)] and from Lemma 4.4.

Let us close this section with some simulation results.

We have simulated AR(1) series ($X_t = 0.5X_{t-1} + e_t$) for the following four densities of the errors:

$$f_1(x) = 1/\sqrt{2\pi}\exp(-x^2/2),$$

$$f_2(x) = 0.05/\sqrt{50\pi}\exp(-x^2/50) + 0.95/\sqrt{2\pi}\exp(-x^2/2),$$

$$f_3(x) = 0.5/\sqrt{2\pi}\exp\left(-(x-3)^2/2\right) + 0.5/\sqrt{2\pi}\exp\left(-(x+3)^2/2\right),$$

$$f_4(x) = \exp(-2|x|).$$

The densities f_1, f_2 and f_4 are commonly used for studying the behavior of the estimators in autoregressive models, while f_3 is chosen to show that the proposed adaptive procedure works quite well.

In all cases we used the usual LS estimator as an initial estimator $\bar{\vartheta}_n$, cf. Fuller (1976), Theorem 8.21, and compare the behavior of $\hat{\vartheta}_n$, cf. (4.2) and $\tilde{\vartheta}_n$, cf. (5.5). The length of the simulated series is $n = 50$.

TABLE 1
Empirical 90% *confidence intervals for* $\sqrt{n}(\bar{\vartheta}_n - 0.5)$ (*case* 1), $\sqrt{n}(\hat{\vartheta}_n - 0.5)$ (*case* 2) *and* $\sqrt{n}(\tilde{\vartheta}_n - 0.5)$ (*case* 3), *respectively, with the smoothing parameter* $\sigma(n) = 0.4$

density \ case	1	2	3
f_1	(−1.57, 1.21)	(−1.57, 1.21)	(−1.75, 1.32)
f_2	(−1.46, 1.13)	(−1.31, 1.00)	(−1.37, 1.09)
f_3	(−1.65, 1.18)	(−0.55, 0.47)	(−0.65, 0.58)
f_4	(−1.57, 1.17)	(−1.30, 1.03)	(−1.45, 1.09)

The simulations show that for extreme distributions of the white noise, e.g., f_3, the adaptive estimator has much more power than the usual LS estimator, even when the latter is used as the initial estimate. In standard situations (f_1, f_4), $\tilde{\vartheta}_n$ can compete with the LS estimator.

Also these simulations suggest that the estimators described and motivated by the theoretical results of this paper represent a first step towards the practical application of adaptive procedures in dependent situations.

Of course, some additional work is necessary to facilitate the computation of such estimators.

The remaining section of the paper is concerned with the proofs of various auxiliary results used in Sections 2–5.

6. Proofs. Let us start with three auxiliary results.

LEMMA 6.1. *For any* $\eta > 1$ *define*

$$B_\eta = \{b = (b_1, \dots, b_q) \in \mathbb{R}^q \colon b(z) = 1 + b_1 z + \cdots + b_q z^q \neq 0 \text{ if } |z| \le \eta\},$$

and, for any $b \in B_\eta$, *let* $\{\beta_j \colon j \in \mathbb{N}\}$ *denote the coefficients of the power series expansion of* $b^{-1}(z)$. *Then* B_η *is an open set in* $\mathbb{R}^q$, *and so, for any* $b^0 \in B_\eta$ *and* $\varepsilon > 0$, *is* $\{b \in B_\eta \colon \sup_j \eta^j |\beta_j - \beta_j^0| < \varepsilon\}$.

PROOF. The assertion follows from Cauchy's estimates [Ahlfors (1966), page 122]. □

LEMMA 6.2. *From assumption* (A.1):

(i). $\sqrt{f}$ *is absolutely continuous with* $\sqrt{f}\,' = f\,'/2\sqrt{f}$.

(ii) *We have*

$$\varphi(\cdot, t) = \left(\frac{f(\cdot - t)}{f(t)}\right)^{1/2} \text{ is } L_2(P_f)\text{-differentiable}, \tag{6.1}$$

at $t = 0$ *with* L_2-*derivative* $\dot{\varphi} = -f\,'/2f$, $dP_f/d\lambda = f$, *and*

$$\int_{\mathbb{R}} \left(\frac{\varphi(x, s) - \varphi(x, t)}{s - t}\right)^2 f(x)\,dx \le \int_{\mathbb{R}} \dot{\varphi}^2(x) f(x)\,dx. \tag{6.2}$$

PROOF. (i) is shown in Ibragimov and Has'minskii (1981), Lemma 2.1, page 121. We restrict our further consideration to (6.2). From (i)

$$\int_{\mathbb{R}} \left(\frac{\varphi(x,s) - \varphi(x,t)}{s-t} \right)^2 f(x)\, dx$$

$$= \frac{1}{(s-t)^2} \int_{\mathbb{R}} \left(\int_t^s \frac{f'(x-\lambda)}{2\sqrt{f(x-\lambda)}}\, d\lambda \right)^2 dx$$

$$\leq \frac{1}{|s-t|} \int_t^s \int_{\mathbb{R}} \left(\frac{f'(x-\lambda)}{2f(x-\lambda)} \right)^2 f(x-\lambda)\, dx\, d\lambda$$

$$= \int \dot{\varphi}^2(x) f(x)\, dx. \qquad \square$$

LEMMA 6.3. *Let $g(\vartheta)$: $(\Omega, \mathscr{A}, P) \to (\mathbb{R}, \mathbb{B})$ be square-integrable and let θ be an open subset of $\mathbb{R}^k$. If $\dot{g}$: $(\Omega, \mathscr{A}, P) \to (\mathbb{R}^k, \mathbb{B}_k)$ exists with*

(i) $$E\left[\frac{g(\vartheta + \lambda h_1) - g(\vartheta + \lambda h_2)}{\lambda} \right]^2 \leq E\left[(h_1 - h_2)^T \dot{g}(\vartheta) \right]^2, \quad h_1, h_2 \in \mathbb{R}^k, \quad \lambda \neq 0,$$

(ii) $$E\|\dot{g}(\vartheta)\|^2 < \infty,$$

(iii) $$\frac{g(\vartheta + \lambda h) - g(\vartheta)}{\lambda} \to_P h^T \dot{g}(\vartheta), \quad \text{as } \lambda \to 0,$$

then g is differentiable in q.m. $[P]$ at $\vartheta \in \theta$.

For a proof of Lemma 6.3 we refer to Kreiss (1984), Lemma 2.4.

Now we can verify the assumptions made by Roussas (1979) which ensure LAN for our model.

PROOF OF THEOREM 3.3. First define

$$Z^\infty(j-1; \vartheta, \vartheta_0) = \sum_{k=1}^{\infty} \beta_{k-1}(\vartheta) \left(Y^T(j-k), E(j-k; \vartheta_0)^T \right)^T$$

and note that $Z(j-1; \vartheta, \vartheta_0)$, $Z(j-1; \vartheta_0)$ $(= Z(j-1; \vartheta_0, \vartheta_0))$ and $Z^\infty(j-1; \vartheta_0)$ $(= Z^\infty(j-1; \vartheta_0, \vartheta_0))$ are (under P_{ϑ_0}) independent of $e_j(\vartheta_0)$. From the definition of φ, recall (6.1), we have

$$\phi_j(\vartheta_0, \vartheta_0 + \lambda h) = \varphi\left(e_j^0, \lambda h^T Z(j-1; \vartheta, \vartheta_0) \right).$$

Now let $H \subset \mathbb{R}^{p+q}$ be a bounded set, i.e., $\|h\| \leq M$ for all $h \in H$. We have to verify that for all $\varepsilon > 0$ there exists $\lambda_0 > 0$ such that for all λ, $|\lambda| \leq \lambda_0$, $h \in H$ and $j \in \mathbb{N}$ $(\vartheta = \vartheta_0 + \lambda h)$:

$$(6.3) \quad E_{\vartheta_0} \left[\frac{\varphi\left(e_j^0, \lambda h^T Z(j-1; \vartheta, \vartheta_0) \right) - 1}{\lambda} - h^T \dot{\varphi}\left(e_j^0 \right) Z(j-1; \vartheta_0) \right]^2 \leq \varepsilon.$$

It is enough to show that (6.4) and (6.5) below hold.

$\forall\, \varepsilon > 0\; \exists\, j_1 \in \mathbb{N}$ and $\lambda_1 > 0$ such that $\forall\, j \geq j_1,\ h \in H,\ |\lambda| \leq \lambda_1$:

$$\text{(i)}\quad E_{\vartheta_0}\left[\frac{\varphi\big(e_j^0, \lambda h^T Z(j-1; \vartheta, \vartheta_0)\big) - \varphi\big(e_j^0, \lambda h^T Z^\infty(j-1; \vartheta_0)\big)}{\lambda}\right]^2 \leq \varepsilon,$$

$$\text{(6.4)}\qquad \text{(ii)}\quad E_{\vartheta_0}\left[h^T\{Z(j-1; \vartheta, \vartheta_0) - Z^\infty(j-1; \vartheta_0)\}\dot{\varphi}\big(e_j^0\big)\right]^2 \leq \varepsilon$$

and

$$\text{(iii)}\quad E_{\vartheta_0}\left[\frac{\varphi\big(e_j^0, \lambda h^T Z^\infty(j-1; \vartheta_0)\big) - 1}{\lambda} - h^T Z^\infty(j-1; \vartheta_0)\dot{\varphi}\big(e_j^0\big)\right]^2 \leq \varepsilon.$$

$\forall\, \varepsilon > 0\; \exists\, \lambda_2 > 0$, such that $\forall\, |\lambda| \leq \lambda_1,\ h \in H,\ j \leq j_1 - 1$:

$$\text{(6.5)}\qquad E_{\vartheta_0}\left[\frac{\varphi\big(e_j^0, \lambda h^T Z(j-1; \vartheta_0, \vartheta)\big) - 1}{\lambda} - h^T Z(j-1; \vartheta_0)\dot{\varphi}\big(e_j^0\big)\right]^2 \leq \varepsilon.$$

To see this we have from (6.2)

$$E_{\vartheta_0}\left[\frac{\varphi\big(e_j^0, \lambda h^T Z(j-1; \vartheta, \vartheta_0)\big) - \varphi\big(e_j^0, \lambda h^T Z^\infty(j-1; \vartheta_0)\big)}{\lambda}\right]^2$$
$$= E_{\vartheta_0}\int\left[\frac{\varphi\big(x, \lambda h^T Z(j-1; \vartheta, \vartheta_0)\big) - \varphi\big(x, \lambda h^T Z^\infty(j-1; \vartheta_0)\big)}{\lambda}\right]^2 f(x)\,dx$$
$$\leq E_{\vartheta_0}\left[h^T\{Z(j-1; \vartheta, \vartheta_0) - Z^\infty(j-1; \vartheta_0)\}\right]^2 I(f)/4,$$

so that (6.4)(i) and (ii) are consequences of Lemma 6.1. The third part of (6.4) is mainly based on

$$E_{\vartheta_0}\left[\frac{\varphi\big(e_j^0, \lambda h^T Z^\infty(j-1; \vartheta_0)\big) - 1}{\lambda} - h^T Z^\infty(j-1; \vartheta_0)\dot{\varphi}\big(e_j^0\big)\right]^2$$
$$= E_{\vartheta_0}\left[\frac{\varphi\big(e_1^0, \lambda h^T Z^\infty(0; \vartheta_0)\big) - 1}{\lambda} - h^T Z^\infty(0; \vartheta_0)\dot{\varphi}\big(e_1^0\big)\right]^2,$$

since $\{Z^\infty(j; \vartheta_0);\ j \in Z\}$ is strictly stationary and $Z^\infty(j-1; \vartheta_0)$ is independent of e_j^0. Now it is an easy task to verify (i)–(iii) of Lemma 6.3 by using Lemma 6.2 to establish (6.4)(iii). For fixed $j \in \mathbb{N}$, one can finally prove (6.5) in the same way. □

PROOF OF THEOREM 3.4. The following inequality holds because of the strict stationarity of $\{Z^\infty(j; \vartheta_0);\ j \in Z\}$:

$$\int 1_{\{[h^T\dot{\phi}_j(\vartheta_0)]^2 > c\}}\left[h^T\dot{\phi}_j(\vartheta_0)\right]^2 dP_{\vartheta_0}$$
$$\leq E_{\vartheta_0}\left|\big(h^T\dot{\phi}_j(\vartheta_0)\big)^2 - \big(h^T Z^\infty(j-1; \vartheta_0)\dot{\varphi}\big(e_j^0\big)\big)^2\right|$$
$$+ \int 1_{\{[h^T Z^\infty(0; \vartheta_0)\dot{\varphi}(e_1^0)]^2 > d\}}\left[h^T Z^\infty(0; \vartheta_0)\dot{\varphi}\big(e_1^0\big)\right]^2 dP_{\vartheta_0}$$
$$+ \frac{d}{c}E_{\vartheta_0}\left[h^T\dot{\phi}_j(\vartheta_0)\right]^2, \qquad d > 0,$$

where $1_{\{\cdot\}}$ denotes the indicator function of the set. Now (6.4)(ii) leads to the assertion. □

For a proof of Theorem 3.5 we refer to Kreiss (1984), Theorem 2.6. It remains to prove Theorem 3.7 in order to ensure LAN.

Proof of Theorem 3.7. Note that [under P_{ϑ_0}] $\{e_t^0; t \in Z\}$ is an ergodic sequence [cf. Doob (1953), page 460] and that $\{h^T Z^\infty(j; \vartheta_0)\}^2$ can be written as an integrable function which depends on e_t^0, $t \le j$, only. By shifting $\{e_t^0\}$ in the argument of this function we obtain $(h^T Z^\infty(j+1; \vartheta_0))^2$. That is why the ergodic theorem [see, for example, Ash and Gardner (1975), Section 3.3.6] implies, as $n \to \infty$, that

$$\frac{1}{n}\sum_{j=1}^{n}\left[\dot{\varphi}\left(e_j^0\right)h^T Z^\infty(j-1; \vartheta_0)\right]^2 \to E_{\vartheta_0}\left[\dot{\varphi}\left(e_1^0\right)h^T Z^\infty(0; \vartheta_0)\right]^2 P_{\vartheta_0} \quad \text{a.e.}$$

and ($\mathscr{A}_{j-1}^* := \sigma(e_t^0, t \le 0; X_{1-p}, \dots, X_{j-1})$)

$$\frac{1}{n}\sum_{j=1}^{n} E_{\vartheta_0}\left[\left\{\dot{\varphi}\left(e_j^0\right)h^T Z^\infty(j-1; \vartheta_0)\right\}^2 | \mathscr{A}_{j-1}^*\right]$$

$$= \frac{1}{4}I(f)\frac{1}{n}\sum_{j=1}^{n}\left[h^T Z^\infty(j-1; \vartheta_0)\right]^2 \to E_{\vartheta_0}\left[\dot{\varphi}\left(e_1^0\right)h^T Z^\infty(0; \vartheta_0)\right]^2 P_{\vartheta_0} \quad \text{a.e.}$$

The result follows by using (6.4) because of

$$E_{\vartheta_0}\left[\left(h^T \dot{\phi}_j(\vartheta_0)\right)^2 | \mathscr{A}_{j-1}\right] = E_{\vartheta_0}\left[\left(\dot{\varphi}\left(e_j^0\right)h^T Z(j-1; \vartheta_0)\right)^2 | \mathscr{A}_{j-1}^*\right]$$

[note that $Z(j-1; \vartheta_0)$ is measurable with respect to $\mathscr{A}_{j-1}$]. □

The next task is to verify the asymptotic linearity of $\Delta_n(\vartheta_0)$, that is to prove (4.5). The proof follows from a paper of Beran (1976), who treats the AR(p) case. Using (2.5) we are able to carry his idea over to more complicated ARMA models. As in Beran's paper we first need

Lemma 6.4. *Define*

$$(6.6) \quad T_n(\vartheta) = \frac{2}{\sqrt{n}}\sum_{j=1}^{n}\left\{\dot{\varphi}\left(e_j(\vartheta)\right) - E_{\vartheta_0}\left[\dot{\varphi}\left(e_j(\vartheta)\right)|\mathscr{A}_{j-1}\right]\right\}Z(j-1; \vartheta).$$

Then

$$(6.7) \quad T_n(\vartheta_n) - T_n(\vartheta_0) = o_{P_{\vartheta_0}}(1)$$

holds, where $\{\vartheta_n\}$ *is a sequence satisfying* (4.4).

Proof. For $\delta > 0$ and $t \ge 1$ let

$$X_t^\delta := \begin{cases} X_t, & |X_t| \le \delta\sqrt{n}, \\ 0, & \text{otherwise.} \end{cases}$$

Let $e_t^{0,\delta}$ denote the similarly truncated version of e_t^0, and set $Z^\delta(j-1;\vartheta,\vartheta_0) = \sum_{k=1}^{j}\beta_{k-1}(X_{j-k}^\delta,\dots,e_{j-k+1-q}^{0,\delta})^T$. Further, define [cf. (2.5)]

$$e_{t,\delta}(\vartheta_n) := (\vartheta_0-\vartheta_n)^T Z^\delta(t-1;\vartheta_n,\vartheta_0) + e_t^0, \tag{6.8}$$

and $T_n^\delta(\vartheta)$ as $T_n(\vartheta)$ with $\dot\varphi(e_{j,\delta}(\vartheta_n))$ instead of $\dot\varphi(e_j(\vartheta_n))$. Since

$$\frac{1}{\sqrt{n}}\max_{t=1-p,\dots,n}|X_t| + \frac{1}{\sqrt{n}}\max_{t=1-q,\dots,n}|e_t^0| = o_{P_{\vartheta_0}}(1), \tag{6.9}$$

by stationarity, we have $T_n^\delta(\vartheta_n) - T_n(\vartheta_n) = o_{P_{\vartheta_0}}(1)$, $\forall\ \delta>0$. To obtain (6.7) it suffices to consider $E_{\vartheta_0}\|T_n^\delta(\vartheta_n)-T_n(\vartheta_0)\|^2$. If we consider the last q components $T_n^{p+1},\dots,T_n^{p+q}$ of T_n only, we obtain, for $t=0,\dots,q-1$,

$$\begin{aligned}
&E_{\vartheta_0}\big[T_n^{p+t+1,\delta}(\vartheta_n) - T_n^{p+t+1}(\vartheta_0)\big]^2\\
&\quad\le \frac{12}{n}E_{\vartheta_0}\Bigg\{\sum_{j=1}^{n}\Bigg[\Bigg(\sum_{k=1}^{j}\beta_{k-1}^n e_{j-k-t}(\vartheta_n)\Bigg)\Big(\dot\varphi\big(e_{j,\delta}(\vartheta_n)\big)-\dot\varphi\big(e_j^0\big)\Big)\\
&\qquad\qquad - E_{\vartheta_0}\Bigg[\Bigg(\sum_{k=1}^{j}\beta_{k-1}^n e_{j-k-t}(\vartheta_n)\Bigg)\Big(\dot\varphi\big(e_{j,\delta}(\vartheta_n)\big)-\dot\varphi\big(e_j^0\big)\Big)\Big|\mathscr{A}_{j-1}\Bigg]\Bigg]\Bigg\}^2\\
&\qquad + \frac{12}{n}E_{\vartheta_0}\Bigg\{\sum_{j=1}^{n}\sum_{k=1}^{j}\beta_{k-1}^n\big(e_{j-k-t}(\vartheta_n)-e_{j-k-t}^0\big)\\
&\qquad\qquad\times\Big(\dot\varphi\big(e_j^0\big)-E_{\vartheta_0}\big[\dot\varphi\big(e_j^0\big)|\mathscr{A}_{j-1}\big]\Big)\Bigg\}^2\\
&\qquad + \frac{12}{n}E_{\vartheta_0}\Bigg\{\sum_{j=1}^{n}\sum_{k=1}^{j}\big(\beta_{k-1}^n-\beta_{k-1}^0\big)e_{j-k-t}^0\Big(\dot\varphi\big(e_j^0\big)-E_{\vartheta_0}\big[\dot\varphi\big(e_j^0\big)|\mathscr{A}_{j-1}\big]\Big)\Bigg\}^2.
\end{aligned}$$

If we observe that $E_{\vartheta_0}[\dot\varphi(e_j^0)|\mathscr{A}_{j-1}] = E_{\vartheta_0}\dot\varphi(e_j^0) = 0$ and that $e_j(\vartheta_n)-e_j^0 = (\vartheta_0-\vartheta_n)^T Z(j-1;\vartheta_n,\vartheta_0)$, then Lemma 6.1 enables us to show that both the second and third summand converge to zero. By using properties of conditional expectations we can bound the first term by

$$\begin{aligned}
&\frac{12}{n}\sum_{j=1}^{n}E_{\vartheta_0}\Bigg[\sum_{k=1}^{j}\beta_{k-1}^n e_{j-k-t}(\vartheta_n)\Bigg]^2\Big[\dot\varphi\big(e_{j,\delta}(\vartheta_n)\big)-\dot\varphi\big(e_j^0\big)\Big]^2\\
&\quad = \frac{12}{n}\sum_{j=1}^{n}E_{\vartheta_0}\Bigg[\sum_{k=1}^{j}\beta_{k-1}^n e_{j-k-t}(\vartheta_n)\Bigg]^2\\
&\qquad\times\int\Big[\dot\varphi\big(x+(\vartheta_0-\vartheta_n)^T Z^\delta(j-1;\vartheta_n,\vartheta_0)\big)-\dot\varphi(x)\Big]^2 f(x)\,dx.
\end{aligned}$$

Since $E_{\vartheta_0}(\sum_{k=1}^{j}\beta_{k-1}^n e_{j-k-t}(\vartheta_n))^2$ is bounded uniformly in n, j and t (use Lemma 6.1 again) and since $|e_{j,\delta}(\vartheta_n)-e_j(\vartheta_0)| = O(\delta)$ uniformly in j, n by (6.8), there exists [because of (A.5)(i)] a $\delta_0>0$ such that this last term is less than ε for all $\delta\le\delta_0$ and $n, j\in\mathbb{N}$. Components 1 to p of T_n are dealt with similarly but more simply. □

Now we are ready to complete the

PROOF OF (4.5). Using (A.5)(ii), (6.7) and the same truncation technique as above we obtain first

$$\Delta_n(\vartheta_n) - \Delta_n(\vartheta_0) + \hat{\Gamma}_n(\vartheta_n) I(f)\sqrt{n}(\vartheta_n - \vartheta_0) = o_{P_{\vartheta_0}}(1),$$
$$\text{(6.10)} \qquad \text{where } \hat{\Gamma}_n(\vartheta) := \frac{1}{n}\sum_{j=1}^{n} Z(j-1;\vartheta) Z^T(j-1;\vartheta).$$

If we can show that

$$\text{(6.11)} \qquad \hat{\Gamma}_n(\vartheta_n) \to \Gamma(\vartheta_0), \quad \text{in } P_{\vartheta_0}\text{-probability as } n \to \infty,$$

holds, the desired result follows. (6.11) itself is implied by

$$\text{(i)} \qquad E_{\vartheta_0}\left(\sum_{k=j}^{\infty} \beta_{k-1}^n X_{j-k-t}\right)^2 \to_{j\to\infty} 0, \quad \text{uniformly in } n \geq 1,$$

$$\text{(ii)} \quad E_{\vartheta_0}\left(\sum_{k=1}^{\infty} (\beta_{k-1}^n - \beta_{k-1}^0) X_{j-k-t}\right)^2 \to_{n\to\infty} 0, \quad \text{uniformly in } j \geq 1,$$

$$\text{(iii)} \qquad \left\{\sum_{k=1}^{\infty} \beta_{k-1}^0 X_{j-k-t};\ j \in Z\right\} \text{ is ergodic and square integrable,}$$

and analogous results for $\{e_j^0\}$ in place of $\{X_j\}$. To prove (i)–(iii) Lemma 6.1 is basic. For details we refer to Kreiss (1984), Lemma 4.4. □

Finally, it remains to prove Theorem 5.1. For the rest of this paper we assume that (A.1)–(A.6) hold and that $\{\vartheta_n\} \subset \theta$ is a sequence satisfying (4.4). As a first step we note that

$$E_{\vartheta_n}\|\tilde{\Delta}_n(\vartheta_n) - \Delta_n(\vartheta_n)\|^2$$
$$\text{(6.12)} \qquad = \frac{4}{n}\sum_{j=1}^{n} E_{\vartheta_0}\left[\|Z(j-1;\vartheta_n)\|^2 \int_{\mathbb{R}} \{\hat{q}_{n,j}(x;\vartheta_n) - \dot{\varphi}(x)\}^2 f(x)\,dx\right],$$

because of the following symmetry property of $\hat{q}_{n,j}$ and $\dot{\varphi}$:

$$\hat{q}_{n,j}(-x;\vartheta) = -\hat{q}_{n,j}(x;\vartheta) \quad \text{and} \quad \dot{\varphi}(-x) = -\dot{\varphi}(x).$$

Along the lines of Bickel (1982) we will now establish a number of auxiliary results which, together, ensure that the right-hand side of (6.12) converges to zero as $n \to \infty$.

LEMMA 6.5. *For each* $x \in \mathbb{R}$

$$E_{\vartheta_n}\left[f_{\sigma(n)}^{-1}(x)\{\hat{f}_{\sigma(n),j}(x;\vartheta_n) - f_{\sigma(n)}(x)\}^2 \|Z(j-1;\vartheta_n)\|^2\right]$$
$$\text{(6.13)} \qquad \leq \frac{\sigma(n)^{-1}}{n-1}\{\kappa_0 + \kappa_1 x^2/(n-1)\}, \qquad 1 \leq j \leq n,$$

where κ_0, κ_1 *denote suitable constants.*

PROOF. Observe that

$$\|Z(j-1;\vartheta_n)\|^2 = \sum_{t=0}^{p-1}\left(\sum_{k=1}^{j}\beta_{k-1}^n X_{j-k-t}\right)^2 + \sum_{t=0}^{q-1}\left(\sum_{k=1}^{j}\beta_{k-1}^n e_{j-k-t}(\vartheta_n)\right)^2,$$

that there exists a sequence $\{\gamma_k^n\}_{k\in\mathbb{N}}$ such that

$$\sum_{k=1}^{j}\beta_{k-1}^n X_{j-k-t} = \sum_{k=1}^{\infty}\gamma_{k-1}^n e_{j-k-t}(\vartheta_n)$$

(cf. Notation 2.1), and that $E_{\vartheta_n}(\sum_{k=1}^{\infty}\gamma_{k-1}^n e_{j-k}(\vartheta_n))^4$ is bounded because of (A.6)(ii). Thus, it is enough to consider terms of the form

$$E_{\vartheta_n}\left[f_{\sigma(n)}^{-1}(x)\left\{\hat{f}_{\sigma(n),j}(x;\vartheta_n) - f_{\sigma(n)}(x)\right\}^2\left(\sum_{k=1}^{\infty}\gamma_{k-1}^n e_{i-k}^n(\vartheta_n)\right)^2\right] = T_n,$$

for $i \le j$.

Noting that $\hat{f}_{\sigma(n),j}$ depends on $e_s(\vartheta_n)$, $1 \le s \le n$, only, we obtain

$$(6.14)\qquad \begin{aligned} T_n \le 2E_{\vartheta_n}&\left[f_{\sigma(n)}^{-1}(x)\left\{\hat{f}_{\sigma(n),j}(x;\vartheta_n) - f_{\sigma(n)}(x)\right\}^2\left(\sum_{k=1}^{i-1}\gamma_{k-1}^n e_{i-k}(\vartheta_n)\right)^2\right] \\ &+\sigma(n)^{-1}/(n-1)O(1) \end{aligned}$$

[compare with Bickel (1982), (6.7)]. Next

$$\begin{aligned} &E_{\vartheta_n}\left[f_{\sigma(n)}^{-1}(x)\left\{\hat{f}_{\sigma(n),j}(x;\vartheta_n) - f_{\sigma(n)}(x)\right\}^2 e_{i-k}^2(\vartheta_n)\right] \\ &\quad\le 2E_{\vartheta_n}\Bigg[f_{\sigma(n)}^{-1}(x)\Bigg[\frac{1}{2(n-1)}\sum_{\substack{t\ne j\\ t\ne i-k}}\{g(x+e_t(\vartheta_n);\sigma(n)) - f_{\sigma(n)}(x) \\ &\qquad\qquad +g(x-e_t(\vartheta_n);\sigma(n)) - f_{\sigma(n)}(x)\}\Bigg]^2 e_{i-k}^2(\vartheta_n)\Bigg] \\ &\quad\quad +2E_{\vartheta_n}\Bigg[f_{\sigma(n)}^{-1}(x)\Bigg[\frac{1}{2(n-1)}\{g(x+e_{i-k}(\vartheta_n);\sigma(n)) - f_{\sigma(n)}(x) \\ &\qquad\qquad +g(x-e_{i-k}(\vartheta_n);\sigma(n)) - f_{\sigma(n)}(x)\}\Bigg]^2 e_{i-k}^2(\vartheta_n)\Bigg] \end{aligned}$$

(note that the second part occurs only if $i-k \ne j$).

Since $y^2 g(x+y;\sigma) \le 2\sigma + 2x^2\sigma^{-1}$, these last expressions can be bounded by $(\sigma(n)^{-1}/(n-1))(O(1) + (x^2/(n-1))O(1))$. Similar methods lead to

$$\begin{aligned} &\left|E_{\vartheta_n}\left[f_{\sigma(n)}^{-1}(x)\left\{\hat{f}_{\sigma(n),j}(x;\vartheta_n) - f_{\sigma(n)}(x)\right\}^2 e_{i-k}(\vartheta_n)e_{i-l}(\vartheta_n)\right]\right| \\ &\quad\le \sigma^2\sigma(n)^{-1}/(n-1)^2, \qquad k \ne l. \end{aligned}$$

Because $\sum_{k=0}^{\infty}|\gamma_k^n|^\alpha = O(1)$, $\alpha = 1,2$, the assertion follows. □

In the same way one can establish

LEMMA 6.6. *For each $x \in \mathbb{R}$*

$$E_{\vartheta_n}\left[f_{\sigma(n)}^{-1}(x)\left\{ \hat{f}'_{\sigma(n),j}(x;\vartheta_n) - f'_{\sigma(n)}(x)\right\}^2 \|Z(j-1;\vartheta_n)\|^2\right] \le \frac{\sigma(n)^{-3}}{n-1}\left(K_0 + K_1 x^2/(n-1)\right), \qquad 1 \le j \le n, \tag{6.15}$$

holds, where K_0, K_1 denote suitable constants.

To follow Bickel (1982), Lemmas 6.1–6.3, we now prove the following three results. First, we have

LEMMA 6.7. *If $n \to \infty$,*

$$\frac{1}{n}\sum_{j=1}^{n} E_{\vartheta_n}\left[\int\left\{\frac{f'_{\sigma(n)}(x)}{\sqrt{f_{\sigma(n)}(x)}} - \frac{f'(x)}{\sqrt{f(x)}}\right\}^2 dx \|Z(j-1;\vartheta_n)\|^2\right] = o(1). \tag{6.16}$$

PROOF. Because of Lemma 6.2 [Bickel (1982)] the result follows since $E_{\vartheta_n}\|Z(j-1;\vartheta_n)\|^2$ is bounded, independently of $j \in \mathbb{N}$. □

Next we prove a more complicated result.

LEMMA 6.8.

$$\frac{1}{n}\sum_{j=1}^{n} E_{\vartheta_n}\left[\int\left\{\hat{q}_{n,j}(x;\vartheta_n) + \frac{1}{2}\frac{f'_{\sigma(n)}(x)}{f_{\sigma(n)}(x)}\right\}^2 \times \|Z(j-1;\vartheta_n)\|^2 f_{\sigma(n)}(x)\,dx\right] = o(1). \tag{6.17}$$

PROOF. We consider two parts of the expression under consideration,

$$I_1^{n,j} = \int_{A_{n,j}B_nC_{n,j}}\left\{\hat{q}_{n,j}(x;\vartheta_n) + \frac{1}{2}\frac{f'_{\sigma(n)}(x)}{f_{\sigma(n)}(x)}\right\}^2 \|Z(j-1;\vartheta_n)\|^2 f_{\sigma(n)}(x)\,dx$$

and

$$I_2^{n,j} = \int_{(A_{n,j}B_nC_{n,j})^c}\left\{\hat{q}_{n,j}(x;\vartheta_n) + \frac{1}{2}\frac{f'_{\sigma(n)}(x)}{f_{\sigma(n)}(x)}\right\}^2 \|Z(j-1;\vartheta_n)\|^2 f_{\sigma(n)}(x)\,dx,$$

where

$$A_{n,j} := \left\{x \mid \hat{f}_{\sigma(n),j}(x;\vartheta_n) \ge d_n\right\}, \quad B_n := \left\{x \mid |x| \le g_n\right\}$$

and

$$C_{n,j} := \left\{x \mid |\hat{f}'_{\sigma(n),j}(x;\vartheta_n)| \le c_n \hat{f}_{\sigma(n),j}(x;\vartheta_n)\right\}.$$

From Lemmas 6.5 and 6.6 and the assumptions on c_n, g_n, $\sigma(n)$ and d_n, we have

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n} E_{\vartheta_n} I_1^{n,j} \le \frac{1}{2n}\sum_{j=1}^{n}\Bigg\{ & E_{\vartheta_n}\Bigg[\int_{A_{n,j}B_nC_{n,j}}\Bigg(\frac{\hat{f}'_{\sigma(n),j}(x;\vartheta_n)}{\hat{f}_{\sigma(n),j}(x;\vartheta_n)} - \frac{\hat{f}'_{\sigma(n),j}(x;\vartheta_n)}{f_{\sigma(n)}(x)}\Bigg)^2 \\ & \qquad \times \|Z(j-1;\vartheta_n)\|^2 f_{\sigma(n)}(x)\,dx\Bigg] \\ & + E_{\vartheta_n}\Bigg[\int_{A_{n,j}B_nC_{n,j}}\Bigg(\frac{\hat{f}'_{\sigma(n),j}(x;\vartheta_n)}{f_{\sigma(n)}(x)} - \frac{f'_{\sigma(n)}(x)}{f_{\sigma(n)}(x)}\Bigg)^2 \\ & \qquad \times \|Z(j-1;\vartheta_n)\|^2 f_{\sigma(n)}(x)\,dx\Bigg]\Bigg\} \\ = o(1). \end{aligned} \tag{6.18}$$

To obtain (6.17) it is now enough, because of (6.16), to prove $n^{-1}\sum_{j=1}^{n} E_{\vartheta_n} J_2^{n,j} = o(1)$, where

$$J_2^{n,j} = \int_{(A_{n,j}B_nC_{n,j})^c}\left(\frac{f'(x)}{f(x)}\right)^2 f(x)\,dx \|Z(j-1;\vartheta_n)\|^2.$$

We start by choosing $j(n) \in \{1,\dots,n\}$ so that $\max[E_{\vartheta_n} J_2^{n,j};\ j = 1,\dots,n] = E_{\vartheta_n} J_2^{n,j(n)}$ is satisfied. Then, for example,

$$\begin{aligned} & E_{\vartheta_n}\int\left[\left(\frac{f'(x)}{f(x)}\right)^2 \|Z(j-1;\vartheta_n)\|^2 1_{A^c_{n,j(n)}}\right] f(x)\,dx \\ & \quad = O(1)\left(\int\left(\frac{f'(x)}{f(x)}\right)^2 E_{\vartheta_n} 1_{A^c_{n,j(n)}} f(x)\,dx\right)^{1/2} = o(1), \end{aligned}$$

because of (A.6)(ii) and because $E_{\vartheta_n} 1_{A^c_{n,j(n)}} \to 0$ for all $x \in \mathbb{R}$, since $\hat{f}_{\sigma(n),j(n)}(x;\vartheta_n) \to_{n\to\infty} f(x)$ in P_{ϑ_n}-probability; cf. Bickel (1982), (6.12).

By similar arguments which take account of (6.13) of Bickel (1982), one can obtain the same result for the remaining two parts of $J_2^{n,j}$. □

Finally we have

Lemma 6.9. *If $n \to \infty$*

$$\frac{1}{n}\sum_{j=1}^{n} E_{\vartheta_n}\left[\int \hat{q}^2_{n,j}(x;\vartheta_n)\left(\sqrt{f_{\sigma(n)}(x)} - \sqrt{f(x)}\right)^2 dx \|Z(j-1;\vartheta_n)\|^2\right] \to 0. \tag{6.19}$$

The proof is exactly the same as that of Lemma 6.3 in Bickel (1982), because $\hat{q}_{n,j}$ possesses the upper bound $c_n/2$ and $E_{\vartheta_n}\|Z(j-1;\vartheta_n)\|^2$ is bounded independently of $j, n \in \mathbb{N}$. Combining (6.12) and Lemmas 6.7–6.9, we obtain

$$E_{\vartheta_n}\|\tilde{\Delta}_n(\vartheta_n) - \Delta_n(\vartheta_n)\|^2 = o(1),$$

from which we can obtain Theorem 5.1, because of contiguity of $\{P_{n,\vartheta_n}\}$ and $\{P_{n,\vartheta_0}\}$ [cf. (3.4)], the measurability of $\tilde{\Delta}_n(\vartheta_n)$ and $\Delta_n(\vartheta_n)$ with respect to $\mathscr{A}_n$, and Lemma 4.4.

Acknowledgments. The author is grateful to an Associate Editor and the referees. Their reports led to considerable improvement of the presentation of the original version of the paper.

REFERENCES

AHLFORS, L. V. (1966). *Complex Analysis*, 2nd ed. McGraw-Hill, New York.

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

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

ASH, R. B. and GARDNER, M. F. (1975). *Topics in Stochastic Processes*. Academic, New York.

BERAN, R. (1976). Adaptive estimates for autoregressive processes. *Ann. Inst. Statist. Math.* **28** 77–89.

BICKEL, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671.

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

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

FULLER, W. A. (1976). *Introduction to Statistical Time Series*. Wiley, New York.

HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 175–194. Univ. California Press.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation, Asymptotic Theory*. Springer, Berlin.

KABAILA, P. (1983). On the asymptotic efficiency of estimators of the parameters of an ARMA process. *J. Time Ser. Anal.* **4** 37–47.

KREISS, J.-P. (1984). Existenz und Konstruktion von adaptiven Schätzfolgen in ARMA(p, q) Modellen. Dissertation, Univ. Hamburg.

KREISS, J.-P. (1987). On adaptive estimation in autoregressive models when there are nuisance functions. To appear in *Statist. Decisions*.

LE CAM, L. (1960). Locally asymptotically normal families of distributions. *Univ. California Publ. Statist.* **3** 27–98.

LUKACS, E. (1968). *Stochastic Convergence*. Heath, Lexington, Mass.

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

ROUSSAS, G. G. (1979). Asymptotic distribution of the log-likelihood function for stochastic processes. *Z. Wahrsch. verw. Gebiete* **47** 31–46.

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

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.

INSTITUT FÜR MATHEMATISCHE STOCHASTIK
UNIVERSITÄT HAMBURG
BUNDESSTRASSE 55
2000 HAMBURG 13
FEDERAL REPUBLIC OF GERMANY