The Annals of Statistics
1983, Vol. 11, No. 3, 872–885

ORDER ESTIMATION IN ARMA-MODELS BY LAGRANGIAN MULTIPLIER TESTS

BY B. M. PÖTSCHER[1]

Technische Universität Wien

A stepwise testing procedure using Lagrangian multiplier tests is developed to determine the order of an ARMA-process. The strong consistency of this procedure under slightly weaker assumptions than in Hannan (1980) (proof of the consistency of the order estimators obtained via BIC) is proved.

1. Introduction. In a series of papers Dunsmuir and Hannan [5] and Deistler, Dunsmuir and Hannan [3], [8] proved the consistency and asymptotic normality of maximum likelihood estimators in the ARMA and ARMAX model. They showed that if the data are generated from an ARMA (or ARMAX) process which is "within" the parameterspace then the transferfunction can always be estimated consistently (in a suitable topology) using a (Gaussian) ML-criterion or an approximation to it. The estimators of the parameters, however, need not converge (they need not even exist) to the true parameters if the true transferfunction is not uniquely parameterized ("not identified") in the given parameterspace. The processes at which this problem arises belong to those which can already be described by a lower dimensional parameterspace, i.e. processes which are of lower order. See also Deistler [2] for a discussion of this problem. Therefore it is necessary to set up procedures to determine the true order of the process. One way to do this is using criteria like the AIC, BIC criterion. Hannan [7] recently proved the consistency of the order estimators obtained via BIC in the scalar ARMA case (given an upper bound for the true order), Hannan and Quinn [6] in the multivariate AR-case, see also [9] for a recursive algorithm and its asymptotic properties.

Another way to determine the order is to use a hypothesis testing framework. There are some recent papers proposing this approach, favouring especially the Lagrange multiplier (LM) test for this sort of testing problems as a "diagnostic checking device", see e.g. [15]–[17]. In [19] we have shown that this way of using the LM-test has some flaws, due to the special topological structure of the parameterspaces of ARMA-models. In the present paper, however, we give a procedure which determines the order of a scalar ARMA process consistently using LM-tests in a suitable way. Moreover the assumption for the consistency proof in Hannan [7] namely that the zeroes of all the MA-polynomials in the parameterspace have to be bounded away uniformly from the unit circle can be relaxed (see Theorem 5.10). Note furthermore that in [13] there are some gaps in some proofs but they do not affect the scalar case without exogenous variables; they are corrected in [14].

2. Preliminary and notation. Consider a scalar weakly stationary process $x(t)$, $t \in \mathbb{Z}$ with zero mean defined on some probability space $(\Omega, \mathscr{A}, P)$. Its spectral measure F is assumed to be rational, i.e. of the form $dF = (\sigma^2/2\pi)|h(z)|^2\, d\lambda(z)$, $\sigma^2 > 0$, where λ is the Haarmeasure on the unit circle $S = \{z \in \mathbb{C} : |z| = 1\}$ with $\lambda(S) = 2\pi$ and $h(z)$ is a rational function, i.e. $h \in \mathbb{R}(z)$, with $h(0) = 1$ having no poles on S. Then it is well known that h can

Received January 1982; revised January 1983.

[1] Inst. of Econometrics and Operations Research, University of Technology, A-1040, Argentinierstr. 8. The support by "Fonds zur Förderung der wissenschaftlichen Forschung", Project No. 3639, is gratefully acknowledged. I am grateful to M. Deistler for his guidance and help during the time this research was done. I also thank the referees for their helpful comments. All remaining errors are mine.

AMS 1980 *subject classifications.* Primary 62M10, 62M15; secondary 60G10, 93E12.

Key words and phrases. Linear systems, ARMA models, order estimation, law of the iterated logarithm, prediction errors estimation, Lagrange multiplier test.

be chosen in a unique way as having no poles in $D \cup S$ and no zeroes in D where $D = \{z \in \mathbb{C} : |z| < 1\}$, having again $h(0) = 1$ and such that $x(t)$ is the (stationary) solution of

$$a_0(L)x(t) = b_0(L)\varepsilon(t) \quad \text{"ARMA-model"} \tag{1}$$

where $a_0(z)$, $b_0(z)$ are polynomials (with real coefficients) with $a_0(0) = b_0(0) = 1$, $a_0^{-1}(z)b_0(z) = h(z)$ and $a_0(z) \neq 0$ on $D \cup S$, $b_0(z) \neq 0$ on D; $\varepsilon(t)$ is white noise with variance σ^2 and is called the (linear) innovation process of x. From (1) we can obtain

$$x(t) = h(L)\varepsilon(t) \tag{2}$$

where the right-hand side is to be understood in mean-square. h is called the transfer function. Note that the polynomials a_0, b_0 are not unique because of possible common factors. A further assumption is that $(1/T)\sum_{t=1}^{T} x(t)x(t-s)$ converges to $E(x(t)x(t-s))$ almost surely for every $s \in \mathbb{Z}$. Clearly there is then a set $E \subseteq \Omega$ with $P(E) = 1$ such that for every $\omega \in E$ the above mentioned sums converge to the corresponding expectations for all $s \in \mathbb{Z}$. In the sequel we will mainly work on the set $E' \subseteq E$, $P(E') = 1$, defined in the appendix, for convenience and forget about the rest of Ω. With $\tilde{x}(t)$ we denote the process $\tilde{x}(t) = x(t)$ for $t \geq 1$ and $\tilde{x}(t) = 0$ otherwise. Next we define our parameter-spaces which we will use in estimating h. Note, however, that our parameterspaces may include rational functions k with poles on S and therefore they are not itself transferfunctions of ARMA-processes. However, their inverses k^{-1} can be interpreted as prediction error transferfunctions.

Denote by U the set of all rational functions $k \in \mathbb{R}(z)$ with $k(0) = 1$, having no poles and zeroes in D. Denote by $U^1 \subseteq U$ the set consisting of rational functions with the additional property of possessing no poles on S, and similarly let $U^2 \subseteq U$ consist of those k having no zeroes on S. If $V \subseteq U$ we will write often $V^1 := V \cap U^1$, $V^2 := V \cap U^2$ and $V^{1,2} := V \cap U^1 \cap U^2$. We view U as a subset of $\mathbb{R}^N$ via the embedding $k \in U \to (T_1(k), T_2(k), \cdots) \in \mathbb{R}^N$ where $k(z) = \sum_{i=0}^{\infty} T_i(k)z^i$ is the Taylor series expansion. The topology U inherits in this way is called the "pointwise" topology (see Deistler [2], Dunsmuir and Hannan [5]).

If $k \in U$ then define $p(k)$, $q(k)$ respectively, as the minimum of the degrees of all denominators, all numerators respectively, of k. Clearly there are uniquely determined polynomials a, b with $a(0) = b(0) = 1$, degree $a = p(k)$, degree $b = q(k)$ and $k = a^{-1}b$. Trivially a and b are prime and are called *the* denominator and *the* numerator of k. Now we are interested in special subsets of U which can be parameterized in an Euclidean way: for $p, q \in \mathbb{N} \cup \{0\}$ let $\theta_{(p,q)}$ be the set of pairs (a, b) of prime polynomials such that $a(0) = b(0) = 1$, a and b have no zeroes in D and $\deg a \leq p$, $\deg b \leq q$ but not both $\deg a < p$ and $\deg b < q$. $\theta_{(p,q)}$ is embedded in $\mathbb{R}^{p+q}$ in a natural way and carries the Euclidean (subspace) topology. The closure of $\theta_{(p,q)}$ in $\mathbb{R}^{p+q}$ is denoted by $\bar{\theta}_{(p,q)}$ and contains all pairs (a, b) of polynomials with $a(0) = b(0) = 1$, a and b have no zeroes in D and $\deg a \leq p$, $\deg b \leq q$. $\theta_{(p,q)}$ is open in $\bar{\theta}_{(p,q)}$ (cf. Deistler [2]). Similarily as before if $A \subseteq \bar{\theta}_{(p,q)}$ we denote by A^1(A^2 respectively) that subset of A consisting only of those pairs (a, b) where $a(z)$ has no zeroes in $D \cup S$ ($b(z)$ has no zeroes in $D \cup S$, respectively). Again $A^{1,2} = A^1 \cap A^2$. Consider now the mapping $\pi_{(p,q)}: \bar{\theta}_{(p,q)} \to U$ defined by $(a, b) \to a^{-1}b$. It is easy to see (cf. Deistler et al. [3]) that $\pi_{(p,q)}$ is continuous with respect to the Euclidean topology and the pointwise topology. Furthermore it can be shown (cf. Deistler [2]; note that in this paper no stability nor invertibility condition is used. The results obviously carry over to our case) that $\pi_{(p,q)}$ restricted to $\theta_{(p,q)}$ is a homeomorphism between $\theta_{(p,q)}$ and its image $\pi_{(p,q)}(\theta_{(p,q)})$ which we abbreviate by $U_{(p,q)}$. The closure $\bar{U}_{(p,q)}$ in U can be shown to be $\pi_{(p,q)}(\bar{\theta}_{(p,q)})$ (but this is true only in the scalar case discussed here) and furthermore $\bar{U}_{(p,q)} = \cup\{U_{(r,s)}: 0 \leq r \leq p, 0 \leq s \leq q\}$ but $U_{(r,s)} \cap U_{(p,q)} \neq \emptyset$ is possible, see Deistler [2]. One should remark that the set $U_{(p,q)}$ is exactly the set of all $k \in \bar{U}_{(p,q)}$, which have a unique parameterization in $\bar{\theta}_{(p,q)}$, i.e. $U_{(p,q)}$ is the set of all "identified" transfer functions. For a fuller account of the topological and algebraic structure of this parameterization see Deistler [2], Deistler and Hannan [4].

For every $k \in U$ define now $\tilde{e}(t, k) = k^{-1}(L)\tilde{x}(t)$ the prediction error made when one uses $k^{-1}(z)$ to predict the process one step ahead from his finite past $x(1), \cdots, x(t)$. Our criterion function is now defined as

$$\tilde{l}_T(k, x(1), \cdots, x(T)) = \frac{1}{T} \sum_{t=1}^{T} \tilde{e}(t, k)^2. \tag{3}$$

$\tilde{l}_T$ is well-defined for all $k \in U$ and sometimes we will omit the arguments $x(1), \cdots, x(T)$. We will also write $\tilde{l}_T$ for the function $\tilde{l}_T(\pi_{(p,q)}(\cdot), x(1), \cdots, x(T))$. Also the notations $\tilde{l}_T(k, \omega)$ or $\tilde{l}_T(a, b, \omega)$ where $a^{-1} b = k$ and $\omega \in \Omega$ will be used. If we replace $\tilde{x}$ by x we will then write l_T and $e(t, k)$ insofar these functions are then defined. Our estimators will be obtained by minimizing $\tilde{l}_T$ over suitable subsets of U. This criterion comes from the theory of prediction error estimation and produces (under some additional assumptions) estimators which are asymptotically as efficient as the Gaussian estimators (i.e. estimators obtained by maximizing a Gaussian likelihood but without necessarily assuming the process to be Gaussian) are; see Ljung and Caines [12], Ploberger [13], [14] and also the appendix for further information. Now as already stated, if one minimizes $\tilde{l}_T$ over some $\bar{U}_{(p,q)}$ one gets estimators $\hat{k}_T$ which converge to h if $h \in \bar{U}_{(p,q)}$ and to the set of those rational functions which minimize the expected prediction error in the case that $h \notin \bar{U}_{(p,q)}$; see Dunsmuir and Hannan [5], Ploberger [13], [14], Ljung and Caines [12]. For a *precise definition* of $\hat{k}_T$ see the appendix. Now in the first case but if $h \in \bar{U}_{(p,q)} \backslash U_{(p,q)}$ the parameter estimator $(\hat{a}_T, \hat{b}_T)$ will not converge, but "search" along the set $\pi_{(p,q)}^{-1}(h)$, i.e. along the equivalence class of "observational equivalent structures". These equivalence classes can be proved to be affine subspaces (intersected with the set given by the conditions for the poles and zeroes, of course), see Deistler [2], and the criterion function is clearly constant on them. Now $h \in \bar{U}_{(p,q)} \backslash U_{(p,q)}$ means that there are smaller integers (r, s) such that $h \in U_{(r,s)}$, so the problem of nonconvergence of the parameter estimators stems from an overparameterization. This is the starting point of the present paper to give a procedure which allows one to determine the "correct" (r, s). The assumptions of this section stated above will be maintained throughout the paper except the converse is stated. The following assumptions, however, will be used only when they are explicitly mentioned:

(A) The process x is strictly stationary and ergodic. Its innovations ε are a martingale difference sequence, i.e. $E(\varepsilon(t) \mid \mathscr{A}_{t-1}) = 0$ where $\mathscr{A}_{t-1}$ is the σ-algebra generated by the past $\{x(s): s < t\}$. Also $E(\varepsilon(t)^2 \mid \mathscr{A}_{t-1}) = \sigma^2$.

(B) $h \in U^2$ (note that by definition as a transfer function, h is in U^1 anyway)

(C) $E\varepsilon(t)^4 < \infty$.

Note that we sometimes use the same symbols for different objects, but there should be no confusion possible. Subsequences will often be indexed by the same index as the original sequences; also $T(\omega)$ is used as a generic notation for a random time not always the same and not necessarily measurable. The symbol λ is used for the Haar-measure on S as well as for the Lebesque measure on $[-\pi, \pi]$, which parametrizes S in an obvious way; in the first case the integrals are meant to be taken over S, in the second case over $[-\pi, \pi]$ which is then always indicated; the argument $e^{i\lambda}$ is sometimes omitted.

3. Some asymptotics. In this section we state some results on the asymptotic behaviour of the score vector and its asymptotic covariance matrix which will be useful later on. Suppose now that $P \geq 0$, $Q \geq 0$, $P, Q \in \mathbb{Z}$ are given. For $(a, b) \in \bar{\theta}_{(P,Q)}$ we write also $\tau(a, b) = (a_1, a_2, \cdots, a_P, b_1, \cdots, b_Q)'$ and conversely if $\tau \in \bar{\theta}_{(P,Q)}$ we write $a_\tau(z)$, $b_\tau(z)$ for the polynomials which are associated to τ in that way. Now for $\tau \in \bar{\theta}_{(P,Q)}$ the function $\tilde{l}_T(\tau)$ is clearly smooth and its vector of first derivatives is denoted by $\tilde{d}_T(\tau)$ (or by $\tilde{d}_T(\tau, \omega)$ to emphasize the dependence on the data). Note that $\tilde{l}_T(\tau)$ is defined and smooth in an open neighborhood of $\bar{\theta}_{(P,Q)}$ so that the derivatives are also defined for boundary points. Denote by $\tilde{f}(t, \tau)$ the vector

$$\left[\frac{L}{b(L)} \tilde{x}(t), \cdots, \frac{L^P}{b(L)} \tilde{x}(t), -\frac{a(L)L}{b^2(L)} \tilde{x}(t), \cdots, -\frac{a(L)L^Q}{b^2(L)} \tilde{x}(t)\right]'$$

of dimension $(P+Q)\times 1$, which is nothing else than $\partial\tilde{e}(t,\tau)/\partial\tau$. If we replace $\tilde{x}(t)$ by $x(t)$ in the definition of $\tilde{f}(t,\tau)$ we obtain a vector we will call $f(t,\tau)$, which is at least defined for $\tau\in\bar{\theta}^2_{(P,Q)}$. We have $\tilde{d}_T(\tau) = (2/T)\sum_{t=1}^T \tilde{e}(t,\tau)\tilde{f}(t,\tau)$, where $\tau=(a,b)$, by an easy computation and the fact that $(\partial/\partial\tau)T_i(\pi_{(P,Q)}(\tau)) = T_i((\partial/\partial\tau)\pi_{(P,Q)}(\tau))$.

Now if $\tau\in\bar{\theta}^2_{(P,Q)}$ then by Lemma A2 of the appendix we see that $\sqrt{T}\,\tilde{d}_T(\tau)$ differs from $\sqrt{T}\,d_T(\tau)$ only by a term which goes to zero in probability, where $d_T(\tau)$ is obtained by replacing $\tilde{x}(t)$ by $x(t)$ in the formula for $\tilde{d}_T(\tau)$ above. Consequently they have the same asymptotic distribution. The proofs of the following statements are given in [19]:

LEMMA 3.1. *If $\tau\in\bar{\theta}^2_{(P,Q)}$ then $Ed_T(\tau) = 2Ee(t,\tau)\cdot f(t,\tau) = (\sigma^2/2\pi)\partial l(\tau)/\partial\tau =: \mu(\tau)$ where $l(\tau)=\int_{-\pi}^{\pi}|b^{-1}(e^{i\lambda})a(e^{i\lambda})h(e^{i\lambda})|^2\,d\lambda$. (Note that $l(\tau)$ is also differentiable in points where $a(z)$ has zeroes on S).*

THEOREM 3.2. *If* (A) *holds and if $\tau_0\in\bar{\theta}^2_{(P,Q)}$ is such that $\pi_{(P,Q)}(\tau_0)=h$ (this clearly implies that h fulfills* (B)) *then $\sqrt{T}\,d_T(\tau_0)$ is asymptotically normal $N(0, 4\sigma^2A(\tau_0))$ where $A(\tau_0)=Ef(t,\tau_0)f(t,\tau_0)'$.*

REMARK. Note that $A(\tau)$ is defined for all $\tau\in\bar{\theta}^2_{(P,Q)}$ even if h is a completely arbitrary transfer function. Furthermore the assumptions in Theorem 3.2 imply that (B) must hold.

THEOREM 3.3. *Suppose that $V\subseteq\bar{\theta}^2_{(P,Q)}$ is such that there is $M>0$, $0<\lambda<1$ such that $|T_i(b_\tau^{-1})|\le M.\lambda^i$ for all $\tau\in V$ and $i\in\mathbb{N}$. Then*

(a) *$A(\tau)=\lim_{T\to\infty}\tilde{A}_T(\tau)$ for all $\omega\in E$ and uniformly in $\tau\in V$, where*

$$\tilde{A}_T(\tau) = \frac{1}{T}\sum_{t=1}^T \tilde{f}(t,\tau)\tilde{f}(t,\tau)' \text{ and } \tau=(a,b).$$

(b) *$B(\tau)=\lim_{T\to\infty}\tilde{B}_T(\tau)$ for all $\omega\in E$ uniformly in $\tau\in V$ where $B(\tau)$ is the expectation of the expression $(\partial^2\tilde{e}(t,\tau)^2/\partial\tau\partial\tau')$ after replacing $\tilde{x}$ by x in it and where $\tilde{B}_T(\tau) = (1/T)\sum_{t=1}^T(\partial^2\tilde{e}(t,\tau)^2/\partial\tau\partial\tau')$.*

(c) *If* (A) *holds and there is $\tau_0\in\bar{\theta}^2_{(P,Q)}$ with $\pi_{(P,Q)}(\tau_0)=h$ then $A(\tau_0)=(1/2)\,B(\tau_0)$.*

(d) *$A(\tau)$, $B(\tau)$ are continuous for $\tau\in\bar{\theta}^2_{(P,Q)}$.*

$A(\tau)$ plays a central role in the construction of the LM-statistic, since every generalized inverse of $A(\tau_0)$ is up to a constant factor a generalized inverse of the asymptotic covariance matrix of the score vector and the rank of $A(\tau_0)$ determines the degrees of freedom of the test statistic, see [19].

4. The procedure. Suppose $P\ge 0$, $Q\ge 0$ are given integers such that $p(h)\le P$, $q(h)\le Q$, i.e. the process x can be described as an ARMA(P, Q)-process. If we want to decide whether it can already be modeled as an ARMA(P', Q')-process with $P'\le P$, $Q'\le Q$ we could try to set up a test for the null hypothesis $h\in\bar{U}^1_{(P',Q')}$. Using the parameterspaces $\bar{\theta}^1_{(P,Q)}$ and $\bar{\theta}^1_{(P',Q')}$ this can be expressed as zero restrictions on the parametervector $(a,b)\in\bar{\theta}^1_{(P,Q)}$ in an obvious way, where (a,b) are *the* denominator and *the* numerator of h. If one tried, however, to use the LM-test for this problem, one would see that the test statistic has a pathological behaviour in several points of the null hypothesis due to the fact first that some points in the null hypothesis need not be identified even there and second that at some points identified within the null hypothesis the matrix $A(\tau)$ changes its rank, making the usual estimators of a generalized inverse of $A(\tau_0)$ inconsistent, as has been discussed in [19]. This behaviour of the LM-statistic seems to have been overlooked in the literature, cf. [15], [16], [17]. As a consequence, one has to be more specific about what one's null hypothesis really is and the use of a *single* LM-test as a diagnostic checking device in the context of testing the orders of an ARMA-model may be challenged on the basis of what was said above, see again [19].

A natural way out of this is to use a whole sequence of LM-tests, starting with the lowest possible order and then testing "upwards" as long as the test rejects. Loosely speaking, the orders p, q where the test does not reject for the first time are then estimates for $p(h)$, $q(h)$. The precise definition will be given below. The question arising then is under which circumstances these estimators are consistent. It will be shown below that strong consistency of the estimators is achieved if the significance levels of all the tests involved tend to zero with increasing sample size at a rate specified below. Next we define the procedure:

First variant:

Step 1. Choose a (P, Q) which is big enough to make you believe that the true transfer function h is in $\bar{U}_{(P,Q)}$.

Step 2. Choose a chain $(0, 0) = (p_0, q_0), (p_1, q_1), \cdots, (p_K, q_K) = (P, Q)$ such that either $p_{i+1} = p_i$, $q_{i+1} = q_i + 1$ or $p_{i+1} = p_i + 1$, $q_{i+1} = q_i$ (necessarily $K = P + Q$).

Step 3. Assign a number $\alpha(p_i, q_i)$ between 0 and 1 to every (p_i, q_i).

Then do for $k = 0, \cdots, K - 1$:

Step 4. Calculate $\hat{k}_T$ for the (p_k, q_k)-specification (i.e. fit an ARMA(p_k, q_k)-model, see the appendix for a precise definition of $\hat{k}_T$) and set $\hat{\tau}_T = \pi^{-1}_{(p_k,q_k)}(\hat{k}_T)$ (more precisely, take a measurable selection of $\pi^{-1}_{(p_k,q_k)}(\cdot)$, which clearly exists). (We will show below that $\pi^{-1}_{(p_k,q_k)}(\hat{k}_T)$ consists only of one point for large T as long as $h \notin \bar{U}_{(p_k,q_k)}$. If $h \in U^{1,2}_{(p_k,q_k)}$ then this definition of $\hat{\tau}_T$ agrees with the one given before Theorem 4.2 in [19] and again $\pi^{-1}_{(p_k,q_k)}(\hat{k}_T)$ is a singleton for large T.)

Step 5. Calculate $R_T^{(p_k,q_k)}(\hat{\tau}_T) = (T/4\hat{\sigma}_T^2)\,\hat{\tilde{d}}'_T\tilde{A}_T(\hat{\tau}_T)^+\hat{\tilde{d}}_T$ where $+$ denotes the Moore-Penrose inverse, see e.g. [20]. If it is bigger than $c_{\alpha(p_k,q_k)}$ then do the $(k + 1)$th-loop; otherwise stop and set $\hat{p}_T = p_k$, $\hat{q}_T = q_k$ (here $c_{\alpha(p_k,q_k)}$ is the upper $\alpha(p_k,q_k)$-quantile of a Chi squared distribution with $\max(P - p_k, Q - q_k)$ degrees of freedom). Here $\hat{\sigma}_T^2 := (1/T)\sum_{t=1}^{T}\tilde{e}(t, \hat{k}_T)^2 = \tilde{l}_T(\hat{\tau}_T)$ and $\hat{\tilde{d}}_T$ is short for $\tilde{d}_T(\hat{\tau}_T)$.

The *second variant* of the procedure is obtained as follows:

Step 1'. As Step 1 above.

Step 2'. Do steps 2–5 of the above procedure for all possible chains. Among the resulting $\hat{p}$'s, $\hat{q}$'s consider those pairs which have the smallest $\hat{p} + \hat{q}$ and select the one which has also the smallest $\hat{p}$ (this last condition is imposed only to make the estimator unique).

Note that if one is willing to assign to a (p, q) in every chain where it appears the same α, then one need not test every chain separately.

In order not to have to fit a too large number of models, one might be interested only in looking at models where $p = q$. The third variant takes this into account:

Third variant:

Step 1''. As step 1 but $P = Q$.

Step 2''. Do steps 3–5 as above but only for the models where $p = q$, $p = 0, 1, \cdots, P$.

If one uses the second variant, one has of course to fit a large number of models, at most all models in the range $0 \le p \le P$, $0 \le q \le Q$. Using criteria like AIC or BIC, one has to fit all the models within this range in any case. Hannan and Rissanen recently proposed an easy algorithm which gives "approximate" ML-estimators and showed that if one uses these estimators instead of the ML-estimators in calculating the values of the BIC-criterion, one also gets consistent estimates of the orders p, q of an ARMA-process, see [9]. Of course their algorithm can be used to calculate an approximation to $\hat{\tau}_T$ and one can use this instead of $\hat{\tau}_T$ in calculating $R_T^{(p_k,q_k)}$. We are sure that the consistency results below would hold true also in this situation; this will be discussed in a separate paper.

The estimators obtained by means of the third variant will be shown to converge to $\max(p(h), q(h))$, i.e. the procedure selects the most parsimonious model of the form $p = q$ which is able to describe the process x. Note that h is identified in this parameterization, since $\pi_{(s,s)}^{-1}(h)$ consists only of one point where $s = \max(p(h), q(h))$.

5. Weak and strong consistency. In this section we give the weak and strong consistency results (Theorem 5.7 and 5.10) for the order estimators derived from the procedure defined in the previous section.

THEOREM 5.1. (*Aström-Söderström*). *Suppose* $h \in \bar{U}^{1,2}_{(P,Q)}$. *Then the function* $l(\tau(a, b)) = \int_{-\pi}^{\pi} |b^{-1}a|^2 |h|^2 \, d\lambda$ *with* $\tau \in \bar{\theta}_{(P,Q)}$ *has its minimum exactly on the set* $\pi^{-1}_{(P,Q)}(h)$. *The gradient of* l *(restricted to* $\bar{\theta}^2_{(P,Q)}$*) vanishes only there.*

REMARK. First note that the Gaussian assumption in [1] is nowhere used in their proofs. If $b^{-1}a$ has a pole on S we set $l = \infty$. Furthermore the proofs in Aström and Söderström allow $a(z)$ also to have zeroes inside or on S. Consequently there is no problem for the gradient in points (a, b) such that $a(z)$ has a zero on S.

THEOREM 5.2. *Let* $h \in U^{1,2}$ *but* $h \notin \bar{U}_{(p,q)}$. *The set*

$$D = \left\{ k \in \bar{U}_{(p,q)} : \int_{-\pi}^{\pi} |k^{-1}|^2 |h|^2 \, d\lambda = \inf\left\{ \int_{-\pi}^{\pi} |k^{-1}|^2 |h|^2 \, d\lambda : k \in \bar{U}_{(p,q)} \right\} \right\}$$

is nonempty, compact and lies in $U^2_{(p,q)}$.

PROOF. Since $h \in U^{1,2}$ we have $\inf\{\int_{-\pi}^{\pi} |k^{-1}|^2 |h|^2 \, d\lambda : k \in \bar{U}^2_{(p,q)}\} = \inf\{\int_{-\pi}^{\pi} |k^{-1}|^2 |h|^2 \, d\lambda : k \in \bar{U}_{(p,q)}\}$. Consequently D is the same as $D_{(p,q)}$ in the appendix and the compactness and nonemptiness follows from the references cited there. To show the second statement of the theorem, take a $k \in D$ and let us assume that $k \notin U^2_{(p,q)}$. k must be in $\bar{U}^2_{(p,q)}$ since otherwise the integral would be infinite which is in contradiction to $k \in D$. Consequently k is not identified which implies that $((1 - az)/(1 - bz))k(z)$ is still in $\bar{U}^2_{(p,q)}$ for all $|a| < 1$, $|b| < 1$, $a, b \in \mathbb{R}$. By definition of k this implies that

$$F(a, b) = \int_{-\pi}^{\pi} \frac{|1 - be^{i\lambda}|^2}{|1 - ae^{i\lambda}|^2} |k^{-1}|^2 |h|^2 \, d\lambda$$

has a minimum at all points $a = b$, $|b| < 1$. Consequently $\partial F/\partial b_{|a=b} = 0$ for all $|b| < 1$.

This entails

$$\int_{-\pi}^{\pi} \frac{2b - (e^{i\lambda} + e^{-i\lambda})}{1 + b^2 - b(e^{i\lambda} + e^{-i\lambda})} |k^{-1}h|^2 \, d\lambda = 0 \quad \text{for all} \quad b \in \mathbb{R}, |b| < 1. \tag{16}$$

Now

$$\frac{2b - (e^{i\lambda} + e^{-i\lambda})}{1 + b^2 - b(e^{i\lambda} + e^{-i\lambda})} = \frac{-e^{i\lambda}}{1 - be^{i\lambda}} - \frac{e^{-i\lambda}}{1 - be^{-i\lambda}} = f_b(e^{i\lambda}) + f_b(e^{-i\lambda})$$

and f_b is quadratic integrable. Also $g(e^{i\lambda}) = |k^{-1}h|^2$ is quadratic integrable since it is continuous. The Fourier series of $f_b(e^{i\lambda})$ equals $-\sum_{j\geq 1} b^{j-1}e^{ij\lambda}$. Write $g(e^{i\lambda}) = g_1(e^{i\lambda}) + g_1(e^{-i\lambda}) + a_0$ where $g_1(e^{i\lambda}) = \sum_{j\geq 1} a_j e^{ij\lambda}$ and the a_j are the Fourier-coefficients of g. This is possible as $\bar{g} = g$. Then (16) becomes

$$\int_{-\pi}^{\pi} [f_b(e^{i\lambda}) + f_b(e^{-i\lambda})][g_1(e^{i\lambda}) + g_1(e^{-i\lambda}) + a_0]\, d\lambda = 0. \tag{17}$$

Observing that $f_b(e^{i\lambda})$ lies in the closure of the span of $e^{ij\lambda}, j \geq 1$ we get

$$\int_{-\pi}^{\pi} f_b(e^{i\lambda})g_1(e^{i\lambda})\, d\lambda = 0 \quad \text{for all} \quad |b| < 1. \tag{18}$$

We now want to show that $B := \{f_b(e^{i\lambda}) : |b| < 1, b \in \mathbb{R}\}$ spans a dense subspace of the set $\mathrm{cl}(\mathrm{span}\{e^{in\lambda}: n \geq 1\})$: first we have $e^{i\lambda} \in \mathrm{span}\, B$ by setting $b = 0$. Now adding $e^{i\lambda}$ to f_b and dividing by b we get $(1/b)(f_b + e^{i\lambda}) = -\sum_{j\geq 2} b^{j-2}e^{ij\lambda}$ which is therefore in span B. Letting b tend to zero, we see that this last expression tends to $-e^{2i\lambda}$ in quadratic mean, consequently $e^{2i\lambda} \in \mathrm{cl\ span}\, B$. Then adding $e^{2i\lambda}$ to $(1/b)(f_b + e^{i\lambda})$, dividing by b and proceeding as above we get successively $e^{in\lambda} \in \mathrm{cl\ span}\, B$ for all $n \geq 1$.

But now (18) implies $g_1(e^{i\lambda}) \equiv 0$ λ – almost everywhere or in other words $|k^{-1}(e^{i\lambda})h(e^{i\lambda})|^2 = a_0$ λ – a.e. But this implies $k = h$ (since both are rational functions) which is impossible since $h \notin \bar{U}_{(p,q)}$. So we have proved $D \subseteq U^2_{(p,q)}$.

Remark. The same proof shows that the above result is true also for arbitrary L_2-spectral densities which are bounded away from zero, and nearly the same proof shows that the result is true also for h with zeroes on S if one replaces $U^2_{(p,q)}$ by $U_{(p,q)}$ in Theorem 5.2.

The proofs of the following two lemmata can be found in the appendix:

Lemma 5.3. *For $\tau \in \bar{\theta}^2_{(P,Q)}$ let $A(\tau)^-$ be any g-inverse of $A(\tau)$. If $\mu(\tau) \neq 0$ then $\mu(\tau)'A(\tau)^-\mu(\tau) > 0$.*

Lemma 5.4. *If $x_n \in \mathbb{R}^k$ converges to x and A_n to A where A_n, A are $k \times k$ matrices which are symmetric and positive semidefinite then* $\liminf x_n'A_n^+x_n \geq x'A^+x$ *(A^+ Moore-Penrose inverse).*

Lemma 5.5. *Let $h \in \bar{U}^{1,2}_{(P,Q)}$ but $h \notin \bar{U}_{(p,q)}$. If $c_T \geq 0$ is a sequence such that $T^{-1}c_T \to 0$ then for all $\omega \in E'$ there is a $T(\omega)$ such that for all $T > T(\omega)$ we have $R_T^{(p,q)}(\hat{\tau}_T) > c_T$.*

Proof. From Theorem 5.2 we conclude that $D = D_{(p,q)}$ is compact and $D \subseteq U^2_{(p,q)}$. Since $\pi_{(p,q)}$ is a homeomorphism between $\theta^2_{(p,q)}$ and $U^2_{(p,q)}$, also $\pi^{-1}_{(p,q)}(D)$ is compact. Therefore it is possible to find a compact neighborhood K of $\pi^{-1}_{(p,q)}(D)$ in $\theta^2_{(p,q)}$ such that K has a positive distance to the set of points $\tau = (a, b)$ such that b has zeroes on S. Consequently the requirements of Lemma A4 in [19] are fulfilled and we have for all $\omega \in E$ $\lim_{T\to\infty}\sup_{\tau\in K}\| \tilde{d}_T(\tau) - \mu(\tau)\| = 0$ since $Ed_T(\tau) = \mu(\tau)$ by virtue of Lemma 3.1. Now since $\hat{\tau}_T$ converges to $\pi^{-1}_{(p,q)}(D)$ (see the appendix) we have for all $\omega \in E'$ a $T(\omega)$ such that for $T > T(\omega)$ the sequence $\hat{\tau}_T$ enters K. We now claim that $\liminf_{T\to\infty} \hat{\tilde{d}}_T'\tilde{A}_T(\hat{\tau}_T)^+\hat{\tilde{d}}_T \geq \inf_{\tau\in K}\mu(\tau)'A(\tau)^+\mu(\tau)$. Suppose it would be false then we could find a subsequence (also indexed by T for ease of notation) such that $\hat{\tilde{d}}_T'\tilde{A}_T(\hat{\tau}_T)^+\hat{\tilde{d}}_T < \inf \mu(\tau)'A(\tau)^+\mu(\tau) - \varepsilon$ for some $\varepsilon > 0$. Since $\hat{\tau}_T$ lies in K (for large T) which is compact, we can select a subsequence (of the first subsequence) which converges to some point say $\tau_1 \in K$. But then

$$\| \tilde{d}_T(\hat{\tau}_T) - \mu(\tau_1)\| \leq \sup_{\tau\in K}\| \tilde{d}_T(\tau) - \mu(\tau)\| + \|\mu(\tilde{\tau}_T) - \mu(\tau_1)\|.$$

The first term now on the right hand side goes to zero as shown above, the second one by continuity of μ. Also by Theorem 3.3, $\tilde{A}(\hat{\tau}_T)$ tends to $A(\tau_1)$. Applying now Lemma 5.4 we get $\liminf_{T\to\infty} \hat{d}'_T\tilde{A}_T(\hat{\tau}_T)^+\hat{d}_T \geq \mu(\tau_1)'A(\tau_1)^+\mu(\tau_1)$. But this is a contradiction to the construction of the subsequence. Now using again Lemma 5.4 we see that $\mu(\tau)'A(\tau)^+\mu(\tau)$ is lower semicontinuous on K and hence attains its minimum. From Lemma 5.3 and Theorem 5.1 we conclude that $\inf_{\tau\in K}\mu(\tau)'A(\tau)^+\mu(\tau) > 0$. Consequently $\hat{d}'_T\tilde{A}_T(\hat{\tau}_T)^+\hat{d}_T > m > 0$ for large T. Next, applying Lemma A4 of |19| to $\tilde{l}_T(\tau)$, we easily (as in Theorem 3.3) get $\tilde{l}_T(\tau) \to (l(\tau).\sigma^2)/2\pi$ uniformly in K. Suppose now $\limsup \tilde{l}_T(\hat{\tau}_T) > \liminf \tilde{l}_T(\hat{\tau}_T)$. Choose a subsequence such that $\tilde{l}_T(\hat{\tau}_T) \to \limsup \tilde{l}_T(\hat{\tau}_T)$. By compactness there is a sub-subsequence $\hat{\tau}_T \to \tau_1 \in \pi^{-1}_{(p,q)}(D)$ and consequently by uniformity of the convergence of $\tilde{l}_T$ and the continuity of $l(\tau)$ we get $\tilde{l}_T(\hat{\tau}_T) \to l(\tau_1)\sigma^2/2\pi$. Doing the same for another subsequence with $\tilde{l}_T(\hat{\tau}_T) \to \liminf \tilde{l}_T(\hat{\tau}_T)$ we get $\tilde{l}_T(\hat{\tau}_T) \to l(\tau_2)\sigma^2/2\pi$ with $\tau_2 \in \pi^{-1}_{(p,q)}(D)$. Now by definition of D, $l(\tau_1) = l(\tau_2)$ and hence lim sup = lim inf. In other words $\hat{\sigma}^2_T \to$ constant > 0. Hence $(1/T)R_T^{(p,q)}(\hat{\tau}_T)$ stays bounded away from zero and therefore $(1/T)R_T^{(p,q)}(\hat{\tau}_T) > T c_T^{-1}$ for $T > T(\omega)$ for all $\omega \in E'$.

THEOREM 5.6. *Assume that $h \in \bar{U}^{1,2}_{(P,Q)}$ and choose a chain $(0, 0) = (p_0, q_0), \cdots, (p_K, q_K) = (P, Q)$. Let $\alpha_T(p_i, q_i)$ be between 0 and 1 such that we have $T^{-1}c_{\alpha_T(p_i,\, q_i)} \to 0$ when $T \to \infty$ for $i = 0, \cdots, K$. Then the estimators $\hat{p}_T$, $\hat{q}_T$ derived from the first variant of the procedure fulfill $\hat{p}_T \geq p(h)$, $\hat{q}_T \geq q(h)$ for $T > T(\omega)$ and $\omega \in E'$ (i.e. almost surely). For the estimators of the third variant we have $\hat{p}_T \geq \max(p(h), q(h))$ for $T > T(\omega)$ and $\omega \in E'$, if $\alpha_T(p, p)$ has the same property as $\alpha_T(p_i, q_i)$ above.*

PROOF. Suppose it would be false; then one could choose a subsequence such that $(\hat{p}_T, \hat{q}_T) = (p_{i_0}, q_{i_0})$ with $p_{i_0} < p(h)$ or $q_{i_0} < q(h)$. But then $h \notin \bar{U}^{1,2}_{(p_{i_0},q_{i_0})}$ and applying Lemma 5.5 we get a contradiction. The second statement is proved similarly.

DEFINITION 5.1. $W_{(p,q)} = \{k \in U: (p(k), q(k)) = (p, q)\}$, $V^{(P,Q)}_{(p,q)} = \cup\{W_{(r,s)}: r \leq p,\ s \leq q,\ \min(P - p, Q - q) = \min(P - r, Q - s)\}$. It is shown in [19] that $V^{(P,Q)}_{(p,q)}$ is the generic, maximal set over whose preimage under $\pi_{(p,q)}$ the matrices $A(\tau)$ have constant rank.

THEOREM 5.7. *Assume* (A) *holds and $h \in \bar{U}^{1,2}_{(P,Q)}$ (implies* (B)*). Then:*

a) *If one assigns (in every chain) $\alpha_T(p, q)$ between 0 and 1 to all (p, q), $p \leq P$, $q \leq Q$ (note that $\alpha_T(p, q)$ may depend on the chain too) such that (for all assignments) $T^{-1} c_{\alpha_T(p,q)} \to 0$ and $\alpha_T(p, q) \to 0$ (i.e. $c_{\alpha_T(p,q)} \to \infty$). Then the estimators $\hat{p}_T$, $\hat{q}_T$ obtained from the second variant are weakly consistent.*
b) *If $P = Q$ and one assigns $\alpha_T(p, p)$ to every $0 \leq p \leq P$ such that $\alpha_T(p, p) \to 0$ and $T^{-1}c_{\alpha_T(p,p)} \to 0$ then the estimator $\hat{p}_T$ obtained from the third variant of the procedure converges to $\max(p(h), q(h))$ in probability.*

PROOF. First choose a chain $(p_0, q_0), \cdots, (p_i, q_i), \cdots, (p_K, q_K)$ such that $p_i = p(h)$, $q_i = q(h)$. Now by Corollary 4.14 in [19] $R_T^{(p_i,q_i)}(\hat{\tau}_T)$ is asymptotically Chi squared. Since $\alpha_T(p_i, q_i) \to 0$, $P(R_T^{(p_i,q_i)}(\hat{\tau}_T) > c_{\alpha_T(p_i,q_i)})$ tends to zero by an easy argument. In other words for every $\delta > 0$ and T large enough there is $E_{\delta,T} \subseteq E'$ such that $P(E_{\delta,T}) \geq 1 - \delta$ and on $E_{\delta,T}$ we have $R_T^{(p_i,q_i)}(\hat{\tau}_T) < c_{\alpha_T(p_i,q_i)}$. Also for large T the tests corresponding to (p_j, q_j), $j < i$, reject for $\omega \in E'$ by Lemma 5.5. Consequently on $E_{\delta,T}$ we have $\hat{p}_T = p_i$, $\hat{q}_T = q_i$, i.e. the estimators obtained by the first variant with respect to this special chain are weakly consistent. Now since on every other chain which does not contain $(p(h), q(h))$ the estimator eventually must be bigger than $(p(h), q(h))$ at least in one component, we have that the estimators of the second variant are on $E_{\delta,T}$ determined by the estimators of the first variant for the special chosen chain. And these are weakly consistent as just shown above. To prove (b) observe that Theorem 5.6 implies that $\hat{p}_T \geq \max(p(h), q(h))$ eventually on E'. Now for $s = \max(p(h), q(h))$ we see that $h \in (V^{(P,P)}_{(s,s)})^{1,2}$, defined above, and hence $R_T^{(s,s)}(\hat{\tau}_T)$ is asymptotically Chi squared again by Corollary 4.14 of [19], and one proceeds as above.

The next theorem relates the convergence speed of the $\alpha_T(p, q)$'s to zero to the convergence speed of the c_{α_T}'s. Note that all asymptotic properties of the three variants of the procedure depend only on the convergence rate of the c_{α_T}'s. Consequently we could have defined the procedure without reference to α_T's and to Chi squared distributions, merely letting $c_T(p, q)$ then be an arbitrary sequence of non-negative numbers; but of course there is no loss of generality in not doing so.

THEOREM 5.8. *Let $0 < \alpha_T < 1$ and c_{α_T} be the corresponding upper quantile of a Chi squared distribution with ν degrees of freedom, i.e. $P(\chi^2_\nu > c_{\alpha_T}) = \alpha_T$. Then for any $f: \mathbb{N} \to \mathbb{R}$, $\lim_{T\to\infty} f(T) = \infty$ we have $(f(T))^{-1} c_{\alpha_T} \to 0$ if and only if $-\log \alpha_T / f(T) \to 0$. (Note that this is independent of ν.)*

The proof of Theorem 5.8. can be found in the appendix.

The next theorems will give the strong consistency results for the second and third variant of the procedure under some mild additional assumptions.

LEMMA 5.9. *Suppose $h \in \bar{U}^{1,2}_{(P,Q)}$ (which implies (B)) and also (A), (C) hold. Let $\hat{\tau}_T$ be the estimator—as in the definition of the procedure—corresponding to a (p, q)-specification such that $h \in V^{(P,Q)}_{(p,q)}$ and denote the true parameter by τ_0, i.e. $\tau_0 = \pi^{-1}_{(p,q)}(h)$. Then $\sqrt{T}\tilde{d}_T(\hat{\tau}_T(\omega))/g(T)$ tends to zero almost surely for all functions $g: \mathbb{R}^+ \to \mathbb{R}^+$ such that $g(T)/\sqrt{\log\log T} \to \infty$ and $\sum_{T=1}^\infty g(T)^{-1} T^{-1} < \infty$.*

PROOF. First note that the assumptions of Theorem 4.2 of [19] are fulfilled since a fortiori we have $h \in U^{1,2}_{(p,q)}$. Then recall equation (8) of [19] which is of the form $\sqrt{T}\tilde{d}_T(\hat{\tau}_T) = \sqrt{T}\tilde{d}_T(\tau_0) + D_T(\omega)\sqrt{T}(\hat{\tau}_T - \tau_0)$. Since $\hat{\tau}_T$ is a restricted estimator and by construction of τ_0 we see that both have their $p+1, \cdots, P$ and $P+q+1, \cdots, P+Q$ coordinate equal to zero. Now we skip the corresponding rows in (8) by premultiplying it with J', where J contains the columns $1, \cdots, p, P+1, \cdots P+q$ of the identity matrix of order $P+Q$. Note that $J'\hat{\tilde{d}}_T$ is zero since $\hat{\tau}_T$ is a (restricted) local minimum of $\tilde{l}_T$ over a neighborhood of τ_0 intersected with $\theta_{(p,q)}$ (for large T); see [19] and the appendix. So we get

$$
\begin{aligned}
0 &= \sqrt{T}J'\tilde{d}_T(\tau_0) + J'D_T(\omega)\sqrt{T}(\hat{\tau}_T(\omega) - \tau_0) \\
&= \sqrt{T}\,J'\tilde{d}_T(\tau_0) + J'D_T(\omega)J\sqrt{T}J'(\hat{\tau}_T(\omega) - \tau_0)
\end{aligned} \tag{24}
$$

the last equality being true since the coordinates in $\hat{\tau}_T - \tau_0$ which become deleted by J' are already zero. Now by Lemma A2, $\sqrt{T}\tilde{d}_T(\tau_0)$ differs from $\sqrt{T}d_T(\tau_0)$ only by a term say U_T which is of the order $o(g(T))$ (for every g as above) almost surely. Consequently $0 = J'U_T + \sqrt{T}J'd_T(\tau_0) + J'D_T(\omega)J\sqrt{T}J'(\hat{\tau}_T(\omega) - \tau_0)$. Now from the proof of Theorem 3.2 and by assumption (A) we see that every component of $\sqrt{T}d_T(\tau_0)$ is a martingale with strictly stationary square-integrable ergodic martingale differences and hence it obeys the law of the iterated logarithm (LIL); see for example [10] or [22]. (24) therefore implies

$$
0 = \frac{J'U_T}{g(T)} + \frac{\sqrt{T}J'd_T(\tau_0)}{g(T)} + \frac{J'D_T(\omega)J\sqrt{T}J'(\hat{\tau}_T(\omega) - \tau_0)}{g(T)} \tag{25}
$$

where the first and second term converge to zero a.s. Now as is shown before Theorem 4.2 in [19], $D_T(\omega) \to 2A(\tau_0) = B(\tau_0)$ for all $\omega \in E'$ and by virtue of the proof of Lemma 4.8 in [19] we know that $J'B(\tau_0)J$ is regular. Consequently we have $\sqrt{T}J'(\hat{\tau}_T(\omega) - \tau_0)/g(T) \to 0$ almost surely and since the remaining coordinates are identically zero we can replace it by $\sqrt{T}(\hat{\tau}_T(\omega) - \tau_0)/g(T) \to 0$. Using again equation (8) of [19] we get

$$
\sqrt{T}\tilde{d}_T(\hat{\tau}_T) = U_T + \sqrt{T}d_T(\tau_0) + D_T(\omega)\sqrt{T}(\hat{\tau}_T(\omega) - \tau_0). \tag{26}
$$

Observing the second term on the right hand side of (26) obeying the LIL, we get

$\sqrt{T}\tilde{d}_T(\hat{\tau}_T)/g(T) \to 0$ almost surely since $U_T/g(T)$ goes to zero by Lemma A2 and the third term of (26) divided by $g(T)$ does as just shown.

THEOREM 5.10. *If $h \in \bar{U}^{1,2}_{(P,Q)}$ and if* (A), (C) *hold then the estimators $\hat{p}_T$, $\hat{q}_T$ obtained from the second variant of the procedure are strongly consistent provided that for all assignments of α_T (to all (p, q)'s in all chains, compare Theorem* 5.7) *we have $T^{-1}c_{\alpha_T(p,q)} \to 0$ and $\liminf g(T)^{-2}c_{\alpha_T(p,q)} > 0$ for some g as in Lemma* 5.9. *Secondly, if $P = Q$ and under the same assumptions (here $\alpha_T(p, q)$ needs only be defined for $p = q$) the estimator $\hat{p}_T$ of the third variant of the procedure converges to* $\max(p(h), q(h))$ *almost surely.*

PROOF. In view of Theorem 5.6, it remains only to prove that the first variant of the procedure applied to a chain which passes through the point $(p(h), q(h))$ stops at this point almost surely, i.e. $R_T^{(p(h),q(h))}(\hat{\tau}_T(\omega)) < c_{\alpha_T(p(h),q(h))}$ for almost all $\omega \in E'$ and for large T. Now in the proof of Corollary 4.14 in [19] we have shown that under the given circumstances $\tilde{A}_T(\hat{\tau}_T)^+ \to A(\tau_0)^+$ and since $\hat{\sigma}_T^2 = \tilde{l}_T(\hat{\tau}_T)$ is a consistent estimator of σ^2 we have from Lemma 5.9 that $R_T^{(p(h),q(h))}(\hat{\tau}_T)/g(T)^2 \to 0$ almost surely. Since we have assumed $\liminf g(T)^{-2}c_{\alpha_T(p,q)} > 0$ for all $p \le P$, $q \le Q$ we clearly obtain the desired result. The second statement is proved in completely the same way by looking now on $R_T^{(s,s)}(\hat{\tau}_T)$, $s = \max(p(h), q(h))$ and again using Corollary 4.14 in [19] and Lemma 5.9 observing that $h \in (V^{(P,P)}_{(s,s)})^{1,2}$.

It is well known that the estimators obtained by minimizing Akaike's criterion AIC tend to overestimate p, q. The following theorem gives conditions under which our procedure gives estimates overestimating the true orders:

THEOREM 5.11. a) *Under the same conditions as in Theorem* 5.7 *but now $\alpha_T(p, q) > c > 0$ for all T and all p, q (and in all chains) and if $(p(h), q(h)) \neq (P, Q)$ then we have that the estimators $\hat{p}_T$, $\hat{q}_T$ obtained by the second variant fulfill $\lim_{T\to\infty} P(\hat{p}_T > p(h), \hat{q}_T \ge q(h)$ or $\hat{p}_T \ge p(h), \hat{q}_T > q(h)) > 0$. If $\alpha_T(p, q) \to 1$ for all p, q then this limit equals one.*

b) *The same as* a) *is true for the estimators obtained by the third variant if $P \neq \max(p(h), q(h))$, the limit under consideration being now $\lim_{T\to\infty} P(\hat{p}_T > \max(p(h), q(h)))$.*

The proof is easy and similar to the proof of Theorem 5.7.

6. Conclusion. The proposed procedure for estimating the orders of an ARMA-process provides an alternative to the usual information criteria like AIC, BIC and so on. The consistency of the resulting estimators is shown under similar conditions as the corresponding result for the estimators obtained via BIC. The assumption that P, Q must be known and fixed is not so severe a problem as it might seem since by an easy argument we see that the consistency result holds also true if P, Q are allowed to increase slowly enough with the sample size T, but we do not know at the moment how slow this increase must be. Nevertheless we expect that the same rates of increase as in [9] will work. We are pretty sure that the same results are true also for other criteria than the prediction error criterion used here. It is well known that the estimators obtained via AIC overestimate the true order asymptotically, but recent results of Shibata [21] show on the other hand that these estimators are optimal in a certain sense if the underlying process is not an AR-process and one fits only AR-models. It is not clear if the estimators considered in Theorem 5.11 have also such a property; moreover it is not clear whether one could find a behaviour of the α_T's such that the resulting estimators combine both desirable properties, namely of being consistent whenever $h \in \bar{U}^{1,2}_{(P,Q)}$ and of being optimal in the sense mentioned above whenever $h \notin \bar{U}^{1,2}_{(P,Q)}$ (or $x(t)$ is even not an ARMA-process). Looking at Theorem 5.11, we see that $\alpha_T \to 1$ provides estimators which overestimate $p(h)$, $q(h)$with probability one asymptotically which is also true for the AIC-estimators. From a testing point of view, however, α_T tending to one does not make much sense. In general, the behaviour of all the order estimation procedures is not very well understood in the case $h \notin \bar{U}^{1,2}_{(P,Q)}$. However, the pathological behaviour of a single LM-test can also occur in this case, as has been discussed in [19].

From a practical point of view all the order estimation procedures involve a lot of computational burden. The third procedure given in this paper is one way of reducing the number of models to be fitted. Of course the Hannan-Rissanen algorithm, [9], can be used to obtain (approximations to) the estimators $\hat{\tau}_T$ on which the test statistics $R_T^{(p,q)}(\hat{\tau}_T)$ are based, compare Section 4.

The assumption that h has no MA-zeroes on S seems hardly to be removable. The fourth moment condition seems to be minor.

APPENDIX

LEMMA A1. *The function $i: U \to U$ defined by $i(k) = k^{-1}$ is a homeomorphism with respect to the pointwise topology.*

For a proof see the appendix of [19].

Now let x be a scalar weakly stationary process as in Section 2 but whose spectral measure F (defined on the unit circle S) need not be necessarily rational. Following [13], [14] we define $d = \inf\{\int |k^{-1}|^2 \, dF : k \in \bar{U}^2_{(p,q)}\}$. Note that if $dF = (\sigma^2/2\pi)\, h d\lambda h^*$ then $d = \inf\{\int |k^{-1}|^2 \, dF : k \in \bar{U}_{(p,q)}\}$ where the integral is assigned the value infinity in the case k^{-1} is not in $L_2(F)$, which can only happen if k^{-1} has a pole on S. (To be precise in this case we would also have to define the value of k^{-1} on that point of S where the pole occurs. This assignment is clearly irrelevant if F gives no mass to this point as in the case of rational spectral measures). Now the set $D_{(p,q)} = \{k \in \bar{U}_{(p,q)} : \int |k^{-1}|^2 \, dF \leq d\}$ can be shown to be nonempty and compact, see [13], [14]. Note that if x is an ARMA-process and the true transfer function $h \in \bar{U}_{(p,q)}$ then $D_{(p,q)} = \{h\}$.

For technical reasons we define as in [13], [14] the following criterion

$$\bar{l}_T(k) = \int_{-\pi}^{\pi} \left(\textstyle\sum_{j=0}^{T} T_j(k^{-1}) e^{ij\lambda}\right) dI_T(\omega) \left(\textstyle\sum_{j=0}^{T} T_j(k^{-1}) e^{-ij\lambda}\right)$$

where $dI_T(\omega)/d\lambda$ is the periodogram $(1/T) \left| \sum_{t=1}^{T} x(t) e^{it\lambda} \right|^2$.

Define the set

$$C_T(\omega) = \left\{k \in \bar{U}_{(p,q)} : \lambda_{\min}(k^{-1}) \geq 1 + \frac{b_T}{T},\ |\tilde{l}_T(k) - \bar{l}_T(k)| \leq g(\lambda_{\min}(k^{-1}))\right\}$$

where $\lambda_{\min}(k^{-1})$ is the modulus of the absolutely smallest pole of k^{-1}, b_T is an arbitrary sequence of real numbers such that $T^{-1} b_T \downarrow 0$, $b_T/\ln T \to \infty$ and g is an arbitrary continuous, strictly monotone function on $\mathbb{R}$ with $g(1) = 0$. Then it follows from Lemma A1 and from the results of Sections 4 and 8 in [13] as well as Sections 4 and 5 of [14] that there is at least one sequence of (measurable) estimators $\hat{k}_T \in \bar{U}_{(p,q)}$ for $T \geq 1$ such that for almost all $\omega \in \Omega$ there is a $T(\omega)$ such that for $T \geq T(\omega)$ the following holds: (i) $\hat{k}_T(\omega) \in C_T(\omega)$, (ii) $\hat{k}_T(\omega)$ is a minimum of $\tilde{l}_T$ over $C_T(\omega)$. Furthermore *every* such sequence $\hat{k}_T$ (termed "*M*-estimator" in [13], [14]) converges almost surely to $D_{(p,q)}$ (see Theorem 8.7 in [13] or Theorem 5.7 in [14]). Loosely speaking, one gets a "consistent" estimator regardless which minimum of $\tilde{l}_T$ over $C_T(\omega)$ is chosen. From now on we will work (for every p, q) with an arbitrary but fixed sequence $\hat{k}_T$. Denote by $E_{(p,q)} \subseteq \Omega$ a set of measure one such that for $\omega \in E_{(p,q)}$, (i), (ii) and the convergence to $D_{(p,q)}$ is fulfilled for the fixed sequence $\hat{k}_T$. For later use let $E' = E \cap \bigcap \{E_{(p,q)} : p, q \in \mathbb{N} \cup \{0\}\}$. Then $P(E') = 1$ and on E' our estimators converge to $D_{(p,q)}$ for all p, q (E is defined in Section 2). Note that our assumptions imply the ones made in [13], [14]. We want to remark that the requirement of minimizing $\tilde{l}_T$ over $C_T(\omega)$ rather than over $\bar{U}_{(p,q)}$ itself is only of technical importance and seems not to be very restrictive, since if the minima over $\bar{U}_{(p,q)}$ lie all in $\bar{U}^2_{(p,q)}$ then they could have been obtained as $\hat{k}_T$ by the above procedure by a suitable choice of b_T and g. Furthermore $C_T(\omega) \uparrow \bar{U}^2_{(p,q)}$. Note also that if x is ARMA and $h \in U^2_{(p,q)}$ then there is an open neighborhood O of the unique parameters (a, b) of h (i.e. $a^{-1}b = h$) in $\theta_{(p,q)}$ such that $\pi_{(p,q)}(O) \subseteq C_T(\omega)$ for large T. That ensures that the estimators $\hat{k}_T$ can be found as local minima in the parameterspace by use of first derivatives. For a proof of this statement see [13], Theorem 8.10.

For the proof of the following lemma see the appendix of [19].

LEMMA A2. *If c and d are rational functions with no poles in $D \cup S$ then $U_T = (1/\sqrt{T}) \sum_{t=1}^T c(L)\tilde{x}(t)d(L)\tilde{x}(t) - (1/\sqrt{T}) \sum_{t=1}^T c(L)x(t)d(L)x(t)$ tends to zero in L^1 and probability. If additionally $Ex^4(t) = c_4 < \infty$ then U_T converges to zero in meansquare and is $o(g(T))$ almost surely for every $g(T)$ such that $\sum_{T=1}^\infty (g(T))^{-1}T^{-1} < \infty$ and $g(T) \to \infty (g: \mathbb{R}^+ \to \mathbb{R}^+)$.*

PROOF OF LEMMA 5.3. $\mu(\tau)$ being $\sigma^2/2\pi$ times the gradient of $l(\tau)$ is orthogonal to the set $\{\bar{\tau} \in \bar{\theta}_{(P,Q)} : l(\bar{\tau}) = l(\tau)\} \supseteq \pi^{-1}_{(P,Q)}(\pi_{(P,Q)}(\tau))$ in the point τ. To be precise $\mu(\tau)$ is orthogonal to the set obtained from $\pi^{-1}_{(P,Q)}(\pi_{(P,Q)}(\tau))$ by translating τ into the origin. But this is exactly $\ker A(\tau)$, see Lemma 4.6 in [19], (intersected with the conditions for the zeroes and poles which do not reduce the dimension!) so we have $\mu(\tau)$ orthogonal to $\ker A(\tau)$ which implies $\mu(\tau) \in \operatorname{Im} A(\tau)$, i.e. $\mu(\tau) = A(\tau) \cdot z$ and $z \neq 0$ since $\mu(\tau) \neq 0$. Now $\mu(\tau)'A(\tau)^-\mu(\tau) = z'A(\tau)A(\tau)^-A(\tau)z = z'A(\tau)\, z \geq 0$ by positive definiteness which implies also that equality can only hold iff $z \in \ker A(\tau)$ which is impossible since $0 \neq \mu(\tau) = A(\tau)z$. Consequently $\mu(\tau)'A(\tau)^-\mu(\tau) > 0$.

PROOF OF LEMMA 5.4. Suppose the lemma is false. Then there must be a subsequence (also denoted by n) such that $x_n'A_n^+x_n < x'A^+x - \varepsilon$. Now the eigenvalues of A_n, say $(\lambda_1(n), \cdots, \lambda_k(n))$ converge to the eigenvalues $(\lambda_1, \cdots, \lambda_k)$ of A. By the symmetry of A_n there is an orthonormal basis $(z_1(n), \cdots, z_k(n)) =: Z(n)$ of eigenvectors corresponding to $(\lambda_1(n), \cdots, \lambda_k(n))$. Now it is possible to find a sub-subsequence such that $Z(n)$ converges to some matrix $Z = (z_1, \cdots, z_k)$. Clearly the columns of Z are an orthonormal basis of eigenvectors corresponding to $(\lambda_1, \cdots, \lambda_k)$. For the quadratic forms we find

$$\begin{aligned} x_n'A_n^+x_n &= x_n'Z(n)Z(n)'A_n^+Z(n)Z(n)'x_n \\ &= x_n'Z(n)(Z(n)'A_nZ(n))^+Z(n)'x_n = x_n'Z(n)\begin{bmatrix}\lambda_1^+(n) & & 0\\ & \ddots & \\ 0 & & \lambda_k^+(n)\end{bmatrix}Z(n)'x_n \end{aligned}$$

where $\lambda^+ = \dfrac{1}{\lambda}$ for $\lambda \neq 0$ and 0 otherwise. Similarly

$$x'A^+x = x'Z\begin{bmatrix}\lambda_1^+ & & 0\\ & \ddots & \\ 0 & & \lambda_k^+\end{bmatrix}Z'x.$$

From that $x_n'A_n^+x_n - x'A^+x = \sum [\lambda_i^{-1}(n)u_i^2(n) - \lambda_i^{-1}u_i^2] + \sum \lambda_j^+(n)u_j^2(n)$ where $u(n) = Z(n)'x_n$, $u = Z'x$ and the first sum ranges over all i such that $\lambda_i \neq 0$, the second one over those j where $\lambda_j = 0$. Note that in the first sum $\lambda_i^{-1}(n)$ is well defined for $n \geq N$ for some N. Observing now that the first sum goes to zero and the second one is always non-negative by positive semidefiniteness, we have arrived at a contradiction to $x_n'A_n^+x_n < x'A^+x - \varepsilon$.

PROOF OF THEOREM 5.8. For ease of notation let $c_T = c_{\alpha_T}$. Now c_T is bounded iff α_T is bounded away from zero. In this case the above statement is clearly true. Next consider the case where $\alpha_T \to 0$ and then clearly $c_T \to \infty$. Now if $\nu > 1$ we have

$$P(\chi_K^2 \geq c_T) \leq \alpha_T \leq P(\chi_L^2 \geq c_T) \tag{19}$$

where K is the biggest even number less than or equal to ν and L is the smallest even number bigger than ν. Using the result of [11] page 173 we get

$$\exp\left(-\frac{1}{2}c_T\right)\sum_{j=0}^{K/2-1}\left(\frac{c_T}{2}\right)^j \Big/ j! \leq \alpha_T \leq \exp\left(-\frac{1}{2}c_T\right)\sum_{j=0}^{L/2-1}\left(\frac{c_T}{2}\right)^j \Big/ j! \tag{20}$$

From that we obtain

$$D_1\exp\left(-\frac{1}{2}c_T\right) \leq \alpha_T \leq D_2\exp\left(-\frac{1}{2}c_T\right)\cdot\left(\frac{c_T}{2}\right)^{\frac{L}{2}-1} \leq D_3\exp\left(-\frac{1}{4}c_T\right) \quad \text{with } D_i > 0. \tag{21}$$

This implies

$$\frac{-\log D_1}{f(T)} + \frac{c_T}{2f(T)} \geq -\frac{\log \alpha_T}{f(T)} \geq -\frac{\log D_3}{f(T)} + \frac{c_T}{4f(T)} \tag{22}$$

from which the claimed equivalence follows. To deal also with the remaining case $\nu = 1$ we use the result of [11] page 57 equation (29).

$$\begin{aligned} 2\left\{1 - \frac{1}{2}[1 + (1 - \exp(-c_T))^{1/2}]\right\} &\leq \alpha_T = 2\{1 - \phi(\sqrt{c_T})\} \\ &\leq 2\left\{1 - \frac{1}{2}\left[1 + \left(1 - \exp\left(-\frac{1}{2}c_T\right)\right)^{1/2}\right]\right\} \end{aligned} \tag{23}$$

where ϕ is the standard normal cumulative function. This implies

$$1 - \sqrt{1 - \exp(-c_T)} \leq \alpha_T \leq 1 - \sqrt{1 - \exp(-\tfrac{1}{2}c_T)}$$

which gives $\exp(-2c_T) \leq \alpha_T \leq \exp(-\frac{1}{2}c_T)$ for large T. From this last inequality we immediately get the required result.

REFERENCES

[1] ÅSTRÖM, K. J., and SÖDERSTRÖM, T. (1974). Uniqueness of the maximum likelihood estimates of the parameters of an ARMA model. *IEEE Trans. Autom. Control*, AC-19, No. 6.

[2] DEISTLER, M. (1983). The properties of the parameterization of ARMAX systems and their relevance for structural estimation and dynamic specification. *Econometrica* **51**.

[3] DEISTLER, M., DUNSMUIR, W., and HANNAN, E. J. (1978). Vector linear time series models: corrections and extensions. *Adv. Appl. Probab.* **10** 360–372.

[4] DEISTLER, M. and HANNAN, E. J. (1981). Some properties of the parameterization of ARMA systems with unknown order. *J. Multivariate Anal.* **11** 479–489.

[5] DUNSMUIR, W. and HANNAN, E. J. (1976). Vector linear time series models. *Adv. Appl. Probab.* **8** 339–364.

[6] HANNAN, E. J. and QUINN, B. G. (1979). The determination of the order of an autoregression. *J. Roy. Statist. Soc. Ser. B* **41** 2, 190–195.

[7] HANNAN, E. J. (1980). The estimation of the order of an ARMA process. *Ann. Statist.* **8** 1071–1081.

[8] HANNAN, E. J., DUNSMUIR, W. and DEISTLER, M. (1980). Estimation of vector ARMAX models. *J. Multivariate Anal.* **10** 275–295.

[9] HANNAN, E. J. and RISSANEN, J. (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika* **69** 81–94.

[10] HEYDE, C. C. and SCOTT, D. J. (1973). Invariance principles for the law of the iterated logarithm for martingales and processes with stationary increments. *Ann. Probab.* **1** 428–436.

[11] JOHNSON, N. and KOTZ, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions* 1. Wiley, New York.

[12] LJUNG, L. and CAINES, P. (1979). Asymptotic normality of prediction error estimators for approximate system models. *Stochastics* **3** 29–46.

[13] PLOBERGER, W. (1981). Prediction error Schätzung in linearen Systemen. Doctoral thesis, Dept. of Econometrics and Operations Research, University of Technology Vienna.

[14] PLOBERGER, W. (1982). On the prediction error estimation of linear rational models I. Research Report No. 13, Dept. of Econometrics and Operations Research, Univ. of Technology Vienna.
[15] POSKITT, D. S. and TREMAYNE, A. R. (1981a). An approach to testing linear time series models. *Ann. Statist.* **9** 974–986.
[16] POSKITT, D. S. and TREMAYNE, A. R. (1980). Testing the specification of a fitted autoregressive moving average model. *Biometrika* **67** 2, 359–363.
[17] POSKITT, D. S. and TREMAYNE, A. R. (1982). Diagnostic tests for multiple time series models. *Ann. Statist.* **10** 114–120.
[18] POSKITT, D. S. and TREMAYNE, A. R. (1981b). Testing misspecification in vector time series models with exogenous variables. Univ. of York.
[19] PÖTSCHER, B. M. (1982). The behaviour of the Lagrangian multiplier test in testing the orders of an ARMA-model. Research Report No. 16, Dept. of Econometrics and Operations Research, Univ. of Technology Vienna.
[20] RAO, C. R. and MITRA, S. K. (1971). *Generalized Inverse of Matrices and its Applications.* Wiley, New York.
[21] SHIBATA, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.* **8** 1, 147–164.
[22] STOUT, W. F. (1970). The Hartman-Winter law of the iterated logarithm for martingales. *Ann. Math. Statist.* **41** 6, 2158–2160.

INSTITUT FÜR ÖKONOMETRIE
TECHNISCHE UNIVERSITÄT WIEN
ARGENTINIERSTRASSE 8
A-1040 WIEN
AUSTRIA