Moving-Average Representation of Autoregressive Approximations

By

Peter Bühlmann

Technical Report No. 423
January 1995

Department of Statistics
University of California
Berkeley, California

Moving-Average Representation of Autoregressive Approximations

Peter Bühlmann *
University of California Berkeley

January 1995

Abstract

We study the properties of an MA(∞)-representation of an autoregressive approximation for a stationary, real-valued process. In doing so we give an extension of Wiener's Theorem in the deterministic approximation set-up. When dealing with data, we can use this new key result to obtain insight into the structure of MA(∞)-representations of fitted autoregressive models where the order increases with the sample size. In particular, we show strong consistency of the MA(∞)-transfer function via autoregressive approximation.

Key words and phrases. AR(∞), Causal, Complex analysis, Invertible, Linear process, MA(∞), Mixing, Time series, Transfer function, Stationary process.

Short title: MA(∞) via Autoregressive Approximation

*Supported by the Swiss National Science Foundation.

1 Introduction

By Wold's decomposition theorem, every stationary, purely nondeterministic, real-valued process $\{X_t\}_{t\in\mathbb{Z}}$ with $E[X_t] = 0$ can be represented as an infinite order moving-average (MA(∞)) $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ ($\psi_0 = 1$) with uncorrelated innovations $\{\varepsilon_t\}_{t\in\mathbb{Z}}$. Under some regularity conditions it is possible to invert this MA(∞)-process and represent $\{X_t\}_{t\in\mathbb{Z}}$ as an infinite order autoregressive process (AR(∞)), $\sum_{j=0}^{\infty} \phi_j X_{t-j} = \varepsilon_t$ ($\phi_0 = 1$). For approximating the process we could therefore rely on a moving-average or on an autoregressive approximation. When dealing with data, autoregressive approximation, as a linear method, is more popular and computationally much faster and easier. It is known that this approximation is often useful and leads to good results in practice, cf. Durbin (1960), Kromer (1970), Berk (1974), An et al. (1982), Hannan (1987).

On the other hand, the MA(∞)-representation of a stationary process is extremely useful in analyzing structural properties. Parts of the probability structure of a stationary process can be discovered in an easy way via the MA(∞)-representation, whereas they become very complex or even intractable in the AR(∞)-representation. The prime example is the autocovariance function $R(k) = Cov(X_0, X_k)$. In the MA(∞)- representation, $R(k) = \sum_{j=0}^{\infty} \psi_j \psi_{j+k}$ is the convolution of the ψ_j's, whereas in the AR(∞)-representation, $\sum_{j=0}^{\infty} \phi_j R(k-j) = 0$ ($k > 0$) describes an 'infinite recursion'. Another example is that of establishing mixing properties of stationary processes, a tractable approach is given via MA(∞)- representations (cf. Doukhan (1994), Withers (1981)).

To better understand the structural properties of autoregressive approximations it would therefore be very useful to know more about the corresponding MA(∞)-representation.

In section 2 we consider deterministic approximations $\{\phi_{j,n}; j \in \mathbb{N}_0\}_{n\in\mathbb{N}}$ for $\{\phi_j\}_{j\in\mathbb{N}_0}$. If $\Phi_n(z) = \sum_{j=0}^{\infty} \phi_{j,n} z^j$ ($z \in \mathbb{C}$) is the approximating AR-transfer function we study the behavior of $\Psi_n(z) = 1/\Phi_n(z) = \sum_{j=0}^{\infty} \psi_{j,n} z^j$ ($z \in \mathbb{C}$) which can be seen as an approximation for the MA(∞)-transfer function$\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j$ ($z \in \mathbb{C}$). We will give the following key result: roughly speaking, if $\sum_{j=0}^{\infty} j^r |\phi_j| < \infty$ ($r \in \mathbb{N}_0$) there exists n_0 such that

$$\sup_{n\geq n_0} \sum_{j=0}^{\infty} j^r |\psi_{j,n}| < \infty. \tag{1.1}$$

For $r = 0$ this is a result in the spirit of Wiener, who has given such an implication in the non-approximation set-up (cf. Wiener (1933), Zygmund (1959)). By representing this new key result in a deterministic context we believe that it can serve as an excellent tool in many different areas.

In section 3 we consider the autoregressive approximations which are estimated by observed data. More precisely, consider a sample $X_1, \ldots, X_n$ from $\{X_t\}_{t\in\mathbb{Z}}$. Let $\hat{\phi}_{1,n}, \ldots, \hat{\phi}_{p,n}$ be the estimated coefficients for an approximating AR(p)- process, where $p = p(n) \to \infty$ ($n \to \infty$) with $p(n) = o(n)$. Denote the fitted autoregressive process by $\hat{X}_t$, i.e., $\sum_{j=0}^{p} \hat{\phi}_{j,n} \hat{X}_{t-j} = \varepsilon_t$, $\hat{\phi}_{0,n} = 1$. If $\sup_t E|\varepsilon_t| < \infty$, it is known that one can usually invert $\{\hat{X}_t\}_{t\in\mathbb{Z}}$ and represent it as an MA(∞), i.e., $\hat{X}_t = \sum_{j=0}^{\infty} \hat{\psi}_{j,n} \varepsilon_{t-j}$, $\hat{\psi}_{0,n} = 1$. However, little is known about the behavior of the MA(∞)-coefficients $\{\hat{\psi}_{j,n}; j \in \mathbb{N}_0\}_{n\in\mathbb{N}}$ and the corresponding transfer function $\sum_{j=0}^{\infty} \hat{\psi}_{j,n} z^j$ ($z \in \mathbb{C}$, $|z| \leq 1$) which are derived via autoregressive approximation. Our result (1.1) translates then to the case where estimation through data is performed. Roughly speaking we obtain under the same condition as for (1.2) about the summability of the coefficients $\{\phi_j\}_{j\in\mathbb{N}_0}$: there exists a random variable n_1 such that

$$\sup_{n\geq n_1} \sum_{j=0}^{\infty} j^r |\hat{\psi}_{j,n}| < \infty \text{ almost surely.} \tag{1.2}$$

Our result (1.2) opens now the door for answering structural questions of interest of (stochastic) autoregressive approximations. It is very useful from a theoretical point of view and can serve as a key result in analyzing statistical problems in the domain of this approximation theory. An almost immediate consequence of (1.2) will be the strong consistency of the MA(∞)-transfer function via autoregressive approximation $\sum_{j=0}^{\infty} \hat{\psi}_{j,n} z^j$ for $\sum_{j=0}^{\infty} \psi_j z^j$ ($|z| \leq 1$). Bhansali (1989) has shown consistency and even asymptotic normality of $\hat{\psi}_{j,n}$ for ψ_j for a finite collection of j's. His results do not handle infinite collections, as in the case of the transfer function $\sum_{j=0}^{\infty} \hat{\psi}_{j,n} z^j$. Another example where a non-finite collection of j's come into play is the h-step prediction mean square error $\sigma^2 \sum_{j=0}^{h-1} \psi_j^2$ ($h \geq 1$), where $h = h(n) \to \infty$ ($n \to \infty$) and $h(n) = o(n)$ (see Bhansali (1989), formula (2.3)). Our result (1.2) together with the consistency of the MA(∞)-transfer function are new contributions in the field of autoregressive approximation.

There are situations in practice where MA-representations arise naturally, as an example we mention here oil investigation (cf. Silvia and Robinson (1979)). If one uses autoregressive approximation for estimation in these situations, one would finally like to transform the results back in the MA-representation for obvious reasons of interpretation.

2 Inverse of a deterministically approximated AR-transfer function

A real-valued AR(∞)-process with mean zero

$$\sum_{j=0}^{\infty} \phi_j X_{t-j} = \varepsilon_t, \ \phi_0 = 1 \ (t \in \mathbb{Z}) \tag{2.1}$$

is determined by the AR(∞)-transfer function $\Phi(z) = \sum_{j=0}^{\infty} \phi_i z^j$ ($z \in \mathbb{C}$) and the sequence $\{\varepsilon_t\}_{t\in\mathbb{Z}}$. Often $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is an i.i.d. or martingale-difference sequence with $E[\varepsilon_t] = 0$, $E|\varepsilon_t|^2 < \infty$. The structure of the model (2.1) can be better exploited for some purposes, if the model is representable as a real-valued MA(∞)-process, i.e.,

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \ \psi_0 = 1 \ (t \in \mathbb{Z}). \tag{2.2}$$

For modelling processes with arbitrary mean we would replace X_t by $X_t - E[X_t]$, $t \in \mathbb{Z}$.

In the sequel we denote by $\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j$ ($z \in \mathbb{C}$) the MA(∞)-transfer function. The equivalence of models (2.1) and (2.2) is known if $E[\varepsilon_t] = 0$, $\sup_t E|\varepsilon_t| < \infty$ and either of the following conditions hold (cf. Berk (1974)):

(i) $\Phi(.) \neq 0$ for $|z| \leq 1$, $\sum_{j=0}^{\infty} |\phi_j| < \infty$.

(ii) $\Psi(.) \neq 0$ for $|z| \leq 1$, $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

Then $\Psi(z) = 1/\Phi(z)$ ($|z| \leq 1$). The decays of the coefficients $\{\phi_j\}_{j\in\mathbb{N}_0}$ and $\{\psi_j\}_{j\in\mathbb{N}_0}$ are related, for completeness we restate a known result (cf. Brillinger (1975), chapter 3.8).

Lemma 2.1 *The following statements (i) and (ii) are equivalent:*

(i) $\Phi(.) \neq 0$ *for* $|z| \leq 1$ *and* $\sum_{j=0}^{\infty} j^r|\phi_j| < \infty$, $r \geq 0$ $(r \in \mathbb{R})$,

(ii) $\Psi(.) \neq 0$ *for* $|z| \leq 1$ *and* $\sum_{j=0}^{\infty} j^r|\psi_j| < \infty$, $r \geq 0$ $(r \in \mathbb{R})$.

Remark: The case $r = 0$ is essentially due to Wiener (1933), $r > 0$ follows by results of Gelfand et al. (1964).

Let us consider now a sequence $\{\Phi_n(.)\}_{n\in\mathbb{N}}$ of approximations for the AR(∞)-transfer function $\Phi(.)$ and study the behavior of the inverted $\Phi_n(.)$'s as approximations for the MA(∞)-transfer function $\Psi(.)$. Let $\{\phi_{j,n}; j \in \mathbb{N}_0\}_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}^{\mathbb{N}_0}$ with $\phi_{0,n} = 1$ for all $n \in \mathbb{N}$. Denote by

$$\Phi_n(z) = \sum_{j=0}^{\infty} \phi_{j,n} z^j \ (z \in \mathbb{C})$$

and (at least formally)

$$1/\Phi_n(z) = \Psi_n(z) = \sum_{j=0}^{\infty} \psi_{j,n} z^j \ (z \in \mathbb{C})$$

(the validity of the expansion of $\Psi_n(z)$ is given below). The following Lemma gives insight into the structure of an inverted approximation of an AR(∞)-transfer function. The results are crucial in the estimation stage for AR(∞)-processes. For striving generality we formulate them in a deterministic set-up, since we believe that they could be useful in various other contexts.

Lemma 2.2 *Assume that* $\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j$ *is bounded away from zero for* $|z| \leq 1$. *Suppose that* $\sum_{j=0}^{\infty} j^r|\psi_j| < \infty$ *and* $\sum_{j=0}^{\infty} j^r|\phi_{j,n} - \phi_j| = o(1)$ $(n \to \infty)$ *for some* $r \in \mathbb{N}_0$. *Then the following holds:*

(i) There exists $n_0 \in \mathbb{N}$ *such that*

$$\sup_{n\geq n_0} \sum_{j=0}^{\infty} j^r|\phi_{j,n}| < \infty,$$

and

$$\inf_{n\geq n_0} \inf_{|z|\leq 1} |\Phi_n(z)| > 0.$$

(ii) There exists n_1 *such that for* $n \geq n_1$, $\Psi_n(z)$ *converges absolutely for* $|z| \leq 1$ *and*

$$\sup_{n\geq n_1} \sum_{j=0}^{\infty} j^r|\psi_{j,n}| < \infty.$$

Proof:

(i) The first statement follows immediately by the assumptions about $\{\phi_{j,n}; j \in \mathbb{N}_0\}_{n\in\mathbb{N}}$ and Lemma 2.1. For the second statement we use that $\Phi(z) = 1/\Psi(z)$ is bounded away from zero for $|z| \leq 1$.

(ii) We first prove the case $r = 0$, which needs a modification of Wiener's Theorem. We closely follow Zygmund (1959), Theorem 5.2. For a Fourier series $g(x) = \sum_{j=0}^{\infty} a_j e^{ijx}$ ($a_j \in \mathbb{R}$) we denote by $\|g\| = \sum_{j=0}^{\infty} |a_j|$. Let $f(x) = \Phi(e^{ix}) = \sum_{j=0}^{\infty} \phi_j e^{ijx}$ $(0 \leq x \leq 2\pi)$. Denote by $f_n(x) = \Phi_n(e^{ix})$.

By (i) there exists $\kappa > 0$ and an n_0 such that

$$\inf_{n\geq n_0} \inf_{x\in[0,2\pi]} |f_n(x)| \geq \kappa.$$

Since $\xi(z) = 1/z$ is analytic for $\{z \in \mathbb{C}; |z| \geq \kappa'\}$ $(\kappa' > 0)$ there exists a $\rho > 0$ such that $\xi(.)$ is regular in $\{z \in \mathbb{C}; |z - f_n(x)| \leq 2\rho,\ x \in [0, 2\pi],\ n \geq n_0\}$. (We use here the notation $\xi(.)$, whereas Zygmund uses $\phi(.)$). Here it is crucial that ρ is universal for all $n \geq n_0$.

Let $s(.)$ be a partial sum of $f(.)$ such that $\|s - f\| \leq \rho/4$. Then $s(.)$ is also a good approximation for $f_n(.)$ if n is sufficiently large. More precisely, we show now that there exists an $\tilde{n}_0 = \tilde{n}_0(\rho)$ such that

$$\max_{0\leq x\leq 2\pi} |s(x) - f_n(x)| \leq \|s - f_n\| \leq \rho/2 \text{ for } n \geq \tilde{n}_0. \tag{2.3}$$

We write $\|s - f_n\| \leq \|s - f\| + \|f_n - f\| \leq \rho/4 + \|f_n - f\|$. By the assumptions about the $\phi_{j,n}$'s we know that there exists an $\tilde{n}_0 = \tilde{n}_0(\rho)$ such that $\|f_n - f\| \leq \rho/4$ for $n \geq \tilde{n}_0$, which proves (2.3). Note again that $s(.)$ is universal for all $n \geq \tilde{n}_0$.

By Cauchy's formula we have

$$\xi(f_n(x)) = \frac{1}{2\pi} \int_0^{2\pi} \frac{\xi[s(x) + \rho e^{i\theta}]}{s(x) + \rho e^{i\theta} - f_n(x)} \rho e^{i\theta} d\theta.$$

The universal bounds for $\|\xi[s(x) + \rho e^{i\theta}]\|$ and $\|s(x) + \rho e^{i\theta} - f_n(x)\|$ follow as in Zygmund (1959), for the latter we use the inequality in (2.3). Then we complete the proof for the case $r = 0$ as in Zygmund (1959), proof of Theorem 5.2.

The case $r > 0$, $r \in \mathbb{N}$ can be handled via differentiation. Denote by ${}^{(r)}$ the r-th derivative. We know that for n sufficiently large

$$\Psi_n^{(r)}(z) = \sum_{j=r}^{\infty} j!/(j-r)!\psi_{j,n} z^{j-r}$$

has the same radius of convergence as $\Psi_n(z)$, which is ≥ 1. Let us now calculate $\Psi_n^{(r)}(z)$ via the identity $\Psi_n(z) = 1/\Phi_n(z)$. Then

$$\Psi_n^{(r)}(z) = \sum_{finite} A_i \frac{1}{B_i},$$

where A_i involves finite products of $\Phi_n^{(s_1)}(z)$ and $(\Phi_n(z))^{s_2}$ with positive or negative signs, $s_1, s_2 \in \{0, \ldots, r\}$ and B_i involves $(\Phi_n(z))^{s_3}$, $s_3 \in \{2, \ldots, 2r\}$.

In particular, $\frac{1}{B_i}$ can be expressed in terms of $(\Psi_n(z))^{s_3}$. By the definition of $\|.\|$ we have

$$\sum_{j=r}^{\infty} j!/(j-r)!|\psi_{j,n}| = \|\Psi_n^{(r)}(.)\| \leq \sum_{finite} \|A_i\| \|\frac{1}{B_i}\|.$$

(For properties of $\|.\|$ see also Zygmund (1959), page 245).
From the assumptions about the first statement in (i) we know that

$$\|A_i\| \leq M_1 < \infty \text{ for all } n \geq n_0.$$

From the case $r = 0$ we know that

$$\|\frac{1}{B_i}\| \leq \|\Psi_n(.)\|^{s_3} \leq M_2 < \infty \text{ for all } n \geq \tilde{n}_0.$$

Hence there exists $n_1 = \max\{n_0, \tilde{n}_0\}$ such that

$$\sup_{n \geq n_1} \sum_{j=0}^{\infty} j!/(j-r)!|\psi_{j,n}| < \infty. \tag{2.4}$$

We write

$$\begin{aligned}\sum_{j=0}^{\infty} j^r |\psi_{j,n}| &\leq \sum_{j=0}^{r-1} j^r |\psi_{j,n}| + \sum_{j=r}^{\infty} j^r |\psi_{j,n}| \\ &\leq (r-1)^r \sum_{j=0}^{\infty} |\psi_{j,n}| + 2^r \sum_{j=r}^{\infty} (j-(r-1))^r |\psi_{j,n}| + 2^r (r-1)^r \sum_{j=r}^{\infty} |\psi_{j,n}|.\end{aligned}$$

Hence we complete the proof by using the fact that r is finite, part (ii) with $r = 0$ and (2.4). $\square$

3 Inverse of an estimated AR-transfer function

In this section we demonstrate the use of Lemma 2.2 when the approximation for the AR(∞)-transfer function is estimated from the data. Let $X_1, \ldots, X_n$ be a sample from $\{X_t\}_{t \in \mathbb{Z}}$. We make the following assumptions:

(A) Model (2.2) holds with $\sum_{j=0}^{\infty} j^r |\psi_j| < \infty$, $r \in \mathbb{N}_0$, strengthening this for the case $r = 0$ to $\sum_{j=0}^{\infty} j^{1/2} |\psi_j| < \infty$, and $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ is stationary ergodic with

$$E[\varepsilon_t | \mathcal{F}_{t-1}] \equiv 0, \; E[\varepsilon_t^2 | \mathcal{F}_{t-1}] \equiv \sigma^2 < \infty, \; E|\varepsilon_t|^4 < \infty,$$

where $\mathcal{F}_t = \sigma(\{\varepsilon_s; s \leq t\})$ denotes the σ-field generated by $\varepsilon_s, \; s \leq t$.

The estimation of $\Phi(z) = \sum_{j=0}^{\infty} \phi_j z^j$ is based on a usual autoregressive approximation, cf. Berk (1974), An et al. (1982), Hannan (1987). Let $p = p(n) \to \infty$ $(n \to \infty)$ with $p(n) = o(n)$. Then we estimate $(\phi_1, \ldots, \phi_p)^T$ by the Yule-Walker estimates $\hat{\phi}_p = (\hat{\phi}_{1,n}, \ldots, \hat{\phi}_{p,n})^T$, defined by

$$\hat{\Gamma}_p \hat{\phi}_p = -\hat{\gamma}_p,$$

where $\hat{\Gamma}_p = [\hat{R}(i-j)]_{i,j=1,\ldots,p}$, $\hat{\gamma}_p = (\hat{R}(1),\ldots,\hat{R}(p))^T$, $\hat{R}(j) = n^{-1}\sum_{t=1}^{n-|j|} X_t X_{t+|j|}$.
Here we have neglected mean correction which would be needed in practice. But all the results presented below still hold if $E[X_t] \neq 0$ and the sample X_t is adjusted by $X_t - \bar{X}$, $t = 1,\ldots,n$.
We set $\hat{\Phi}_n(z) = \sum_{j=0}^{p} \hat{\phi}_{j,n} z^j$, $\hat{\phi}_{0,n} = 1$. If $\sup_t E|\varepsilon_t| < \infty$, the fitted autoregressive process

$$\sum_{j=0}^{p} \hat{\phi}_{j,n} \hat{X}_{t-j} = \varepsilon_t$$

is always causal (cf. Brockwell and Davis (1987), page 233), i.e.,

$$X_t = \sum_{j=0}^{\infty} \hat{\psi}_{j,n} \varepsilon_{t-j} \text{ with } \sum_{j=0}^{\infty} |\hat{\psi}_{j,n}| < \infty,$$

$$1/\hat{\Phi}_n(z) = \hat{\Psi}_n(z) = \sum_{j=0}^{\infty} \hat{\psi}_{j,n} z^j \ (|z| \leq 1).$$

The next result clarifies about the behavior of $\sum_{j=0}^{\infty} j^r |\hat{\psi}_{j,n}|$, $r \in \mathbb{N}_0$.

Theorem 3.1 *Suppose assumption (A) holds with $r \in \mathbb{N}_0$ and $p(n) = o((n/log(n))^{1/(2r+2)})$. Then there exists a random variable n_1 such that*

$$\sup_{n \geq n_1} \sum_{j=0}^{\infty} j^r |\hat{\psi}_{j,n}| < \infty \textit{ almost surely.}$$

Proof: Let $\phi_{j,n}$ be defined by the (theoretical) Yule-Walker equations

$$\Gamma_p \phi_p = -\gamma_p,$$

where $\phi_p = (\phi_{1,n},\ldots,\phi_{p,n})^T$, $\Gamma_p = [R(i-j)]_{i,j=1,\ldots,p}$, $\gamma_p = (R(1),\ldots,R(p))^T$, $R(j) = Cov(X_0, X_j)$. For ease of notation set $\hat{\phi}_{j,n} = \phi_{j,n} = 0$ for $j > p$ and $\phi_{0,n} = 1$. Then

$$\begin{aligned} \sum_{j=0}^{\infty} j^r |\hat{\phi}_{j,n} - \phi_j| &\leq \sum_{j=0}^{p} j^r |\hat{\phi}_{j,n} - \phi_{j,n}| + \sum_{j=0}^{p} j^r |\phi_{j,n} - \phi_j| + \sum_{j=p+1}^{\infty} j^r |\phi_j| \\ &= I + II + III. \end{aligned} \tag{3.1}$$

A first rough bound leads to

$$I \leq p^{r+1} \max_{1 \leq j \leq p} |\hat{\phi}_{j,n} - \phi_{j,n}|.$$

In Hannan and Kavalieris (1986), Theorem 2.1 it is shown that

$$\max_{1 \leq j \leq p} |\hat{\phi}_{j,n} - \phi_{j,n}| = O((log(n)/n)^{1/2}) \text{ almost surely.}$$

Therefore by the assumption about p we have that $I = o(1)$ $(n \to \infty)$ almost surely.
Expression II can be bounded by the extended Baxter inequality (cf. Deistler and Hannan

(1988), Theorem 6.6.12 and page 271; note that this result also holds for their $\lambda = 0$). Then, for n sufficiently large

$$II \leq c \sum_{j=p+1}^{\infty} (2^r + j^r)|\phi_j|,$$

where c is a constant depending on the true structure. Hence $II = o(1)$ $(n \to \infty)$.
Finally $III = o(1)$ $(n \to \infty)$ by assumption (A) and Lemma 2.1. Note that the terms II and III are deterministic.
Therefore by (3.1)

$$\sum_{j=0}^{\infty} j^r |\hat{\phi}_{j,n} - \phi_{j,n}| = o(1) \ (n \to \infty) \text{ almost surely,}$$

and the proof is completed by Lemma 2.2. □

Theorem 3.1 also holds under the weaker assumption in (A) that $E[\varepsilon_t^2|\mathcal{F}_{-\infty}] = \sigma^2$ instead of $E[\varepsilon_t^2|\mathcal{F}_{t-1}] = \sigma^2$.

The above Theorem is useful for establishing bounds, but says nothing about the convergence of $\hat{\Psi}_n(z) = \sum_{j=0}^{\infty} \hat{\psi}_{j,n} z^j$ to $\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j$ $(|z| \leq 1)$. The following Theorem establishes the consistency of estimating $\Psi(.)$ via autoregressive approximation.

Theorem 3.2 *Suppose assumption (A) holds with $r = 1$ and $p(n) = o((n/log(n))^{1/4})$. Then*

$$\hat{\Psi}_n(z) = \Psi(z) + o(1) \ (n \to \infty) \textit{ almost surely,}$$

the convergence being uniform in z for $|z| \leq 1$.

Remark: Similar results could be derived for the derivatives of $\hat{\Psi}_n(.)$ and $\Psi(.)$.

Proof: It is sufficient to show

$$\sum_{j=a(n)+1}^{\infty} |\hat{\psi}_{j,n}| = o(1) \text{ almost surely,} \tag{3.2}$$

$$\sum_{j=1}^{a(n)} |\hat{\psi}_{j,n} - \psi_j| = o(1) \text{ almost surely,} \tag{3.3}$$

where $a(n) \to \infty$ suitably $(n \to \infty)$. We choose $a(n) = o(p(n))$ and $a(n) \leq p(n)$.
Formula (3.2) follows by Theorem 3.1, i.e., there exists a random variable n_1 such that for $n \geq n_1$

$$\sum_{j=a(n)+1}^{\infty} |\hat{\psi}_{j,n}| \leq a(n)^{-1} \sup_{n \geq n_1} \sum_{j=0}^{\infty} j |\hat{\psi}_{j,n}| = o(1) \text{ almost surely.}$$

We show (3.3) by constructing a fitted AR(p)-process $\{Y_t\}_{t \in \mathbb{Z}}$ independent of the observation process $\{X_t\}_{t \in \mathbb{Z}}$; then the proof is elegant by using known results about autoregressive approximation. Let

$$\sum_{j=0}^{p} \hat{\phi}_{j,n} Y_{t-j} = \eta_t, \tag{3.4}$$

where η_t is i.i.d. with $E[\eta_t] = E[\varepsilon_t] = 0,\ E|\eta_t|^2 = E|\varepsilon_t|^2 = \sigma^2$ and the η_t being independent of $Y_s,\ s \leq t$. Then, since every fitted AR(p)-process is causal (cf. Brockwell and Davis (1987), p.233),

$$Y_t = \sum_{j=0}^{\infty} \hat{\psi}_{j,n} \eta_{t-j}. \tag{3.5}$$

Denote by E^Y and by Cov^Y the conditional expectation and covariance with respect to Y given the sample $X_1, \ldots, X_n$. By (3.4)-(3.5) we get for $u \geq 0$:

$$E[Y_{t+u}\eta_t] = \sigma^2 \hat{\psi}_{u,n} = \sum_{j=0}^{p} \hat{\phi}_{j,n} R_Y(u+j),$$

where $R_Y(k) = Cov(Y_0, Y_k) = \sigma^2 \sum_{j=0}^{\infty} \hat{\psi}_{j,n} \hat{\psi}_{j+|k|,n}$. Note that by construction via the Yule-Walker estimates, $R_Y(k) = \hat{R}(k)$ $(|k| \leq p)$, where $\hat{R}(k) = n^{-1} \sum_{t=1}^{n-|k|} X_t X_{t+|k|}$.
On the other hand,

$$\sigma^2 \psi_u = \sum_{j=0}^{\infty} \phi_j R(u+j).$$

Now

$$\begin{aligned}
& \sum_{j=1}^{a(n)} |\hat{\psi}_{j,n} - \psi_j| \\
\leq \ & \sigma^{-2} \sum_{j=1}^{a(n)} \sum_{i=0}^{\infty} |\hat{\phi}_{i,n} - \phi_i| |R_Y(j+i)| + \sigma^{-2} \sum_{j=1}^{a(n)} \sum_{i=0}^{\infty} |\phi_i| |R_Y(j+i) - R(j+i)| \\
\leq \ & \sigma^{-2} a(n) |R_Y(0)| \sum_{i=0}^{\infty} |\hat{\phi}_{i,n} - \phi_i| + \sigma^{-2} \sum_{j=1}^{a(n)} \sum_{i=0}^{\infty} |\phi_i| |R_Y(j+i) - R(j+i)| \\
= \ & I + II.
\end{aligned}$$

By choosing $a(n)$ growing sufficiently slow, the term I goes to zero almost surely by the proof of Theorem 3.1, see formula (3.1).
For expression II we write:

$$\begin{aligned}
II \ \leq \ & \sigma^{-2} a(n) \sum_{i=0}^{p-a(n)} |\phi_i| \max_{0 \leq j \leq p} |\hat{R}(j) - R(j)| \\
+ \ & a(n) \sum_{i=p-a(n)+1}^{\infty} |\phi_i| \Big(\sum_{k=p-a(n)+2}^{\infty} \sum_{j=0}^{\infty} |\hat{\psi}_{j,n} \hat{\psi}_{j+k,n}| + \sum_{k=p-a(n)+2}^{\infty} \sum_{j=0}^{\infty} |\psi_{j,n} \psi_{j+k,n}| \Big).
\end{aligned}$$

By Theorem 3.1 and $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and $\sum_{i=0}^{\infty} |\phi_i| < \infty$ we know that the second term on the right hand side is going to zero almost surely (choose $a(n)$ sufficiently slowly growing). For bounding the first term on the right hand side we use

$$\max_{0 \leq j \leq p} |\hat{R}(j) - R(j)| = o((log(n)/n)^{1/2}) \text{ almost surely}$$

(cf. An et al. (1982), Theorem 3).
Therefore:

$$II \leq a(n)o((log(n)/n)^{1/2}) + o(1) \text{ almost surely.}$$

Since $a(n)$ is growing slowly, we get $II = o(1)$ almost surely. This completes the proof of (3.3) and hence of Theorem 3.2. $\square$

Theorem 3.1 might be extended to parameters r with values in $\mathbb{R}^+ \cup \{0\}$ by using fractional derivatives. Such a stronger result would yield weaker assumptions for Theorem 3.2, in particular (A) with some $r > 0$. To see this, note that for proving Theorem 3.2 we have only used $r = 1$ in (3.2) to bound the term $\sum_{j=a(n)+1}^{\infty} |\hat{\psi}_{j,n}|$. These mathematical generalizations are not the scope of this paper.

Acknowledgments: I would like to thank David Brillinger for helpful comments and suggestions. I appreciate also many interesting discussions with Ritei Shibata.

References

[1] An,H.-Z., Chen,Z.-G. and Hannan,E.J. (1982). Autocorrelation, autoregression and autoregressive approximation. Ann.Statist. **10** 926-936 (Corr: **11** p1018).

[2] Berk,K.N. (1974). Consistent autoregressive spectral estimates. Ann.Statist. **2** 489-502.

[3] Bhansali,R.J. (1989). Estimation of the moving-average representation of a stationary process by autoregressive model fitting. J. Time Series Anal. **10** 215-232.

[4] Brillinger,D.R. (1975). Time Series. Data Analysis and Theory. Holt, Rinehart and Winston, New York.

[5] Brockwell,P.J. and Davis,R.A. (1987). Time Series: Theory and Methods. Springer, New York.

[6] Deistler,M. and Hannan,E.J. (1988). The Statistical Theory of Linear Systems. Wiley, New York.

[7] Doukhan,P. (1994). Mixing. Properties and Examples. Lect. Notes in Stat. **85**. Springer, New York.

[8] Durbin,J. (1960). The fitting of time-series models. Rev. Intern. Statist. Inst. **28** 233-244.

[9] Gelfand,I., Raikov,D. and Shilov, G. (1964). Commutative Normed Rings. Chelsea, New York.

[10] Hannan,E.J. (1987). Rational transfer function approximation. Stat.Science **5** 105-138.

[11] Hannan,E.J. and Kavalieris,L. (1986). Regression, autoregression models. J. Time Series Anal. **7**, 27-49.

[12] Kromer,R.E. (1970). Asymptotic properties of the autoregressive spectral estimator. Ph.D. thesis, Dept. Statistics, Stanford University.

[13] Silvia,M.T. and Robinson,E.A. (1979). Deconvolution of Geophysical Time Series in the Exploration for Oil and Natural Gas. Elsevier, Amsterdam.

[14] Wiener,N. (1933). The Fourier Integral and Certain of its Applications. Cambridge Univ. Press, Cambridge.

[15] Withers,C.S. (1981). Central limit theorems for dependent variables I. Z. Wahrsch. verw. Gebiete **57** 509-534 (Corr: **63** p555).

[16] Zygmund,A. (1959). Trigonometric Series, Vol.1. Cambridge Univ. Press, Cambridge.

Department of Statistics
Evans Hall
University of California
Berkeley, CA 94720
USA