The Annals of Statistics
2001, Vol. 29, No. 3, 815–838

UNIFORM CONVERGENCE OF SAMPLE SECOND MOMENTS OF FAMILIES OF TIME SERIES ARRAYS

BY DAVID F. FINDLEY, BENEDIKT M. PÖTSCHER AND CHING-ZONG WEI

Bureau of the Census, University of Vienna and Academia Sinica

We consider abstractly defined time series arrays $y_t(T), 1 \leq t \leq T$, requiring only that their sample lagged second moments converge and that their end values $y_{1+j}(T)$ and $y_{T-j}(T)$ be of order less than $T^{\frac{1}{2}}$ for each $j \geq 0$. We show that, under quite general assumptions, various types of arrays that arise naturally in time series analysis have these properties, including regression residuals from a time series regression, seasonal adjustments and infinite variance processes rescaled by their sample standard deviation. We establish a useful uniform convergence result, namely that these properties are preserved in a uniform way when relatively compact sets of absolutely summable filters are applied to the arrays. This result serves as the foundation for the proof, in a companion paper by Findley, Pötscher and Wei, of the consistency of parameter estimates specified to minimize the sample mean squared multistep-ahead forecast error when invertible short-memory models are fit to (short- or long-memory) time series or time series arrays.

1. Introduction. This article provides uniform convergence results for sample second moments of families of time series arrays under very weak conditions. By array we mean data $y_t(T), 1 \leq t \leq T$, that may change as the series length T increases. The motivation for considering such results is two-fold: (i) These results serve as the foundation for very general consistency results for parameter estimators obtained by minimizing sample mean squared p-step-ahead forecast errors, $p \geq 1$, that are presented in the companion article Findley, Pötscher and Wei (2000). Estimation based on multistep prediction criteria has received increasing attention in recent years, mainly because it can result in better forecasts than maximum likelihood estimation when the model is misspecified; see Tiao and Xu (1993) and Findley (1983). However, for the case $p > 1$, few results on the convergence of such estimators are available. (ii) In practice, observations to which standard models are fit are often not time series but rather time series arrays, a possibility the conventional asymptotic theory does not address. Regression or other model residuals, including forecast errors, are examples of arrays which are often further analyzed by time series methods. So are outputs of time varying filters like the seasonally adjusted major economic indicators. A further source of arrays is the data-dependent rescaling of infinite variance processes done to

Received October 1997; revised August 2000.

AMS 2000 *subject classifications*. Primary 62M10, 62M15, 62M20; secondary: 60G10, 62J05, 93E12.

Key words and phrases. Regression residuals, lacunary systems, infinite variance processes, long memory processes, seasonally adjusted series, locally stationary series, uniform laws of large numbers, consistency.

obtain arrays whose sample second moments converge; see Davis and Resnick (1985). Also, the locally stationary processes considered in Dahlhaus (1997) are arrays.

The precise description of general model classes to which we apply the results of this article and the resulting convergence theory for the parameter estimates are developed in a separate article [Findley, Pötscher and Wei (2000)]. Here we shall only sketch the estimation problem enough to motivate the formalism of the uniform convergence results of Section 2: Consider a class of time series models indexed by a parameter vector θ which completely specifies the model's autocovariance function, for example, autoregressive moving average models or the exponential model of Bloomfield (1973). Then the infinite-past and finite-past p-step-ahead linear forecast functions, which are mean square optimal for predicting a process x_t whose autocovariances coincide with the model's autocovariances, can be calculated as projections or, equivalently, as conditional expectations w.r.t. the zero-mean Gaussian distribution determined by the model's autocovariance function,

$$x_{t+p|t}(\theta) = E_\theta\left(x_{t+p}|x_s, -\infty < s \le t\right) = \sum_{j=0}^{\infty} \pi_j(p,\theta)\, x_{t-j} \tag{1.1}$$

and

$$\widetilde{x}_{t+p|t}(\theta) = E_\theta\left(x_{t+p}|x_s, 1 \le s \le t\right) = \sum_{j=0}^{t-1} \pi_{t,j}(p,\theta)\, x_{t-j}.$$

We wish to consider the case in which the linear forecast function in (1.1) is applied to a weakly stationary series y_t different from x_t and we want to be certain that the resulting infinite series $\sum_{j=0}^{\infty} \pi_j(p,\theta)\, y_{t-j}$ converges. Mean square convergence is guaranteed for every weakly stationary y_t if

$$\sum_{j=0}^{\infty} \left|\pi_j(p,\theta)\right| < \infty, \tag{1.2}$$

a property that holds for all models that are invertible (i.e., have a strictly positive spectral density) and that are of short-memory type in the sense that their autocovariance sequences are absolutely summable; see Section 3 of Findley, Pötscher and Wei (2000). Given only finitely many data $y_1, \dots, y_T$, or more generally any array $y_t(T)$, $1 \le t \le T$, one can define the model-based truncated infinite-past predictor for times $t+p$, $t \le T$, as

$$y_{t+p|t}(\theta, T) = \sum_{j=0}^{t-1} \pi_j(p,\theta)\, y_{t-j}(T),$$

and its finite-past predictor as

$$\tilde{y}_{t+p|t}(\theta, T) = \sum_{j=0}^{t-1} \pi_{t,j}(p,\theta)\, y_{t-j}(T).$$

It is of interest to establish the almost sure (in probability) convergence of parameter estimates $\hat{\theta}^T$ obtained by minimizing one of the associated sample mean squared p-step-ahead forecast errors,

$$s_{T,p}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left(y_t(T) - y_{t|t-p}(\theta, T)\right)^2$$

or

$$\tilde{s}_{T,p}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left(y_t(T) - \tilde{y}_{t|t-p}(\theta, T)\right)^2,$$

over a parameter set Θ. Convergence of such $\hat{\theta}^T$ as $T \to \infty$ can be established by standard arguments [see, e.g., Lemma 4.2 of Pötscher and Prucha (1997)] if the uniform convergence on Θ of $s_{T,p}(\theta)$ and $\tilde{s}_{T,p}(\theta)$ can be shown. For the case in which (1.2) holds for all $\theta \in \Theta$, the uniform convergence result given in Theorem 2.1 below provides the foundation for a proof of this uniform convergence since $s_{T,p}(\theta)$ and $\tilde{s}_{T,p}(\theta)$ are easily seen to be sample second moments of linear filters applied to the array $y_t(T)$. [For the uniform convergence of $\tilde{s}_{T,p}(\theta)$, a bound for $\sup_{\theta\in\Theta}\sum_{j=0}^{t-1}\left|\pi_{t,j}(p,\theta) - \pi_j(p,\theta)\right|$ that yields condition (2.15) of Theorem 2.1 below is also needed. This can be obtained from a simple variant of Baxter's inequality; see Findley (1991).]

The uniform convergence result is presented in Section 2. It only requires the underlying array $y_t(T)$ to have convergent sample lagged second moments (and satisfy natural negligibility conditions), thus covering a wide range of dependent processes and arrays. Sections 3 and 4 are devoted to verifying these properties for some important examples of arrays. Section 3 considers $y_t(T) = \widehat{u}_t(T)$ where

$$\widehat{u}_t(T) = Y_t - \hat{A}_T X_t, \qquad 1 \le t \le T,$$

are the residuals from the estimation of a regression model of the form

$$Y_t = AX_t + u_t,$$

in which the error process u_t is weakly stationary and has convergent sample lagged second moments. A substantial extension is obtained of a result by Gleser (1966) on the almost sure convergence of the sample mean of squared regression residuals to the regression error variance in the case of i.i.d. errors. Section 4 considers arrays

$$y_t(T) = y_t\left(T^{-1}\sum_{t=1}^{T} y_t^2\right)^{-1/2}$$

obtained from rescaling processes y_t with infinite variance. Implications, related results and extensions are discussed in the concluding Section 5. Most proofs have been placed in appendices.

2. The uniform convergence result. We start from an array of real-valued random variables $y_t(T), 1 \le t \le T$, defined on a common probability space $(\Omega, \mathfrak{A}, P)$. In the statistical applications discussed in Findley, Pötscher and Wei (2000) the array $y_t(T)$ has the interpretation of the (possibly pre-processed) data available at each sample size T which are to be modeled. We shall usually assume that their sequences of lag j sample second moments converge to real numbers γ_j almost surely or in probability as $T \to \infty$:

$$\text{(CVG)} \qquad \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-j} y_{t+j}(T) y_t(T) = \gamma_j \qquad \text{a.s. [i.p.]}, \; j = 0, 1, \ldots.$$

Concerning extensions to the case of stochastic γ_j, see Section 5.2. Condition CVG is clearly satisfied for many classes of long- and short-memory processes or arrays: As a simple example, strictly stationary and ergodic processes y_t with finite second moments satisfy CVG almost surely with $\gamma_j = E y_{t+j} y_t$. Other sets of conditions that yield CVG for stationary processes are described in Theorem IV.3.6 of Hannan [(1970), page 210]; cf. also Sections 3.1.1 and 3.2 below. In the context of not necessarily stationary processes/arrays, condition CVG also holds for the class of asymptotically stationary processes considered in Parzen (1962), or under standard mixing or near epoch dependence conditions. Locally stationary processes [Dahlhaus (1997)] also satisfy CVG under appropriate assumptions. Regression residuals and suitably rescaled infinite variance processes will be shown to obey CVG; see Sections 3 and 4 below.

We shall usually also need the property that finitely many data at one or both ends of the array can be neglected when calculating the limits in CVG. It is convenient to express these negligibility conditions as

$$\text{(N1)} \qquad \lim_{T\to\infty} \frac{y_{1+j}(T)}{T^{1/2}} = 0 \qquad \text{a.s. [i.p.]}, \; j = 0, 1, \ldots$$

and

$$\text{(N2)} \qquad \lim_{T\to\infty} \frac{y_{T-j}(T)}{T^{1/2}} = 0 \qquad \text{a.s. [i.p.]}, \; j = 0, 1, \ldots.$$

[In expressions like these, whenever an index value (e.g., $j = T$) produces a time index t outside $[1, T]$, set $y_t(T) = 0$. Similarly, any empty sum (e.g. $\sum_{t=1}^{0}$) is assigned the value zero.] The pair of negligibility conditions will be denoted by N. The combination of CVG and N is denoted by CVGN. We add subscripts to restrict consideration to a particular mode of convergence, for example, $\text{CVG}_{a.s.}$. We note that a simple moment condition like $\sup_{T\ge1} \sup_{1\le t\le T} E\left|y_t(T)\right|^{\alpha} < \infty$ implies $\text{N}_{i.p.}$ if $\alpha > 0$, and implies $\text{N}_{a.s.}$ if $\alpha > 2$. Throughout this article, a statement such as that CVG plus N2 imply a property P is to be interpreted as saying that $P_{a.s.}$ follows from $\text{CVG}_{a.s.}$ plus $\text{N2}_{a.s.}$ and $P_{i.p.}$ follows from $\text{CVG}_{i.p.}$ plus $\text{N2}_{i.p.}$.

REMARK 2.1 (The time series case). For a real-valued time series y_t, $t = 1, 2, \ldots,$ the negligibility conditions are automatic in the following sense:

$N1_{a.s.}$ (and hence $N1_{i.p.}$) always holds. If y_t satisfies CVG,

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-j} y_{t+j} y_t = \gamma_j \qquad \text{a.s. [i.p.]}, \quad j = 0, 1, \ldots,$$

then N2 also holds, that is, $\lim_{T\to\infty} y_{T-j}/T^{\frac{1}{2}} = 0$ a.s. [i.p.], $j = 0, 1, \ldots$. This follows from

$$\begin{aligned} \frac{|\, y_T \,|^2}{T} &= \frac{1}{T} \sum_{t=1}^{T} y_t^2 - \frac{1}{T} \sum_{t=1}^{T-1} y_t^2 \\ &= \frac{1}{T} \sum_{t=1}^{T} y_t^2 - \frac{T-1}{T} \frac{1}{T-1} \sum_{t=1}^{T-1} y_t^2 \to 0 \qquad \text{a.s. [i.p.]}. \end{aligned}$$

2.1. *Uniform analogues of* CVG *and N for families of arrays.* Let $z_t(\alpha, T)$, $1 \le t \le T$, $T = 1, 2, \ldots$, be arrays of real-valued random variables defined for each $\alpha \in A$, where A is some non-empty index set. We consider uniform versions of CVG, N1 and N2 for $z_t(\alpha, T)$ formulated as

$$\lim_{T\to\infty} \sup_{\alpha,\beta\in A} \left| \frac{1}{T} \sum_{t=1}^{T-j} z_{t+j}(\alpha, T)\, z_t(\beta, T) - \gamma_j(\alpha, \beta) \right| = 0 \qquad \text{a.s. [i.p.]} \tag{2.1}$$

for $j = 0, 1, \ldots$,

$$\lim_{T\to\infty} \sup_{\alpha\in A} \frac{\left| z_{1+j}(\alpha, T) \right|}{T^{\frac{1}{2}}} = 0 \qquad \text{a.s. [i.p.]}, j = 0, 1, \ldots \tag{2.2}$$

and

$$\lim_{T\to\infty} \sup_{\alpha\in A} \frac{\left| z_{T-j}(\alpha, T) \right|}{T^{\frac{1}{2}}} = 0 \qquad \textit{a.s.} \text{ [i.p.]}, j = 0, 1, \ldots, \tag{2.3}$$

with real numbers $\gamma_j(\alpha, \beta)$ in (2.1).

We shall make repeated use of the following elementary result, which provides conditions under which approximating arrays inherit properties like CVG, N, or (2.1)–(2.3). Proposition 2.1 is proved in Appendix A, where we also comment on measurability issues in Remark A.1.

PROPOSITION 2.1. *For each $\alpha \in A$, a non-empty index set, let $z_t(\alpha, T)$, $1 \le t \le T$, $T = 1, 2, \ldots$, be an array of real-valued random variables. Suppose $\hat{z}_t(\alpha, T)$ is a second family of arrays of real-valued random variables that approximates $z_t(\alpha, T)$ in the sense that*

$$\lim_{T\to\infty} \sup_{\alpha\in A} \frac{1}{T} \sum_{t=1}^{T} \left(z_t(\alpha, T) - \hat{z}_t(\alpha, T) \right)^2 = 0 \qquad \textit{a.s. [i.p.]} \tag{2.4}$$

holds. Then, if the family $z_t(\alpha, T)$ *satisfies* (2.2) *or* (2.3)*, so does* $\hat{z}_t(\alpha, T)$*. If* (2.1) *holds for* $j \in J$*, a set of nonnegative integers containing* 0*, and if* $\gamma_0(\alpha, \alpha)$ *is bounded on* A*, then*

$$\text{(2.5)} \qquad \lim_{T\to\infty} \sup_{\alpha,\beta\in A} \left| \frac{1}{T} \sum_{t=1}^{T-j} \hat{z}_{t+j}(\alpha, T)\, \hat{z}_t(\beta, T) - \gamma_j(\alpha, \beta) \right| = 0 \qquad \textit{a.s.}\ \text{[i.p.]}$$

holds for $j \in J$.

REMARK 2.2 (i) In the case of a single array, that is, when A contains only one element, Proposition 2.1 with J the set of all nonnegative integers shows that CVGN and its sample second moment limits are inherited by approximating arrays for which (2.4) holds.

(ii) Condition (2.4) implies via the Cauchy-Schwarz inequality that

$$\lim_{T\to\infty} \sup_{\alpha\in A} \left| \frac{1}{T} \sum_{t=1}^{T} z_t(\alpha, T) - \frac{1}{T} \sum_{t=1}^{T} \hat{z}_t(\alpha, T) \right| = 0 \qquad \text{a.s. [i.p.].}$$

Hence, if one of these sample means converges, the other converges to the same limit.

(iii) The proof of Proposition 2.1 in fact shows closeness of the sample second moments of $z_t(\alpha, T)$ and $\hat{z}_t(\alpha, T)$ uniformly in the lag [as well as in (α, β)], that is,

$$\lim_{T\to\infty} \max_{0\le j\le T-1} \sup_{\alpha,\beta\in A} \left| \frac{1}{T} \sum_{t=1}^{T-j} z_{t+j}(\alpha, T)\, z_t(\beta, T) - \frac{1}{T} \sum_{t=1}^{T-j} \hat{z}_{t+j}(\alpha, T)\, \hat{z}_t(\beta, T) \right| = 0 \qquad \text{a.s. [i.p.]}$$

under the assumptions used to prove (2.5). Inequality (A.1) in Appendix A then even shows that the rate of convergence to zero in the above display is not slower than the square root of the rate of convergence to zero in (2.4). This result is of interest as, for example, it allows one to establish closeness uniformly in the lag of the sample second moments of the errors and the residuals from a regression; cf. Section 3 and Pötscher (1998).

2.2. *A uniform law of large numbers for convolution-filter families.* Let Φ be a set of absolutely summable sequences $\phi = (\phi_0, \phi_1, \ldots)$ of real numbers ϕ_j satisfying

$$\text{(2.6)} \qquad \sup_{\phi\in\Phi} \sum_{j=k}^{\infty} |\phi_j| \le C_k, \qquad k = 0, 1, 2, \ldots,$$

for some decreasing sequence C_k of real numbers converging to 0, that is,

$$\text{(2.7)} \qquad C_k \searrow 0.$$

Conditions (2.6) and (2.7) characterize the relatively compact sets (i.e., sets with compact closure) in the normed space l^1 of absolutely summable sequences with norm $\|\phi\|_1 = \Sigma_{j=0}^{\infty} |\phi_j|$. More precisely, for any decreasing sequence of real numbers $(C_k)_{k\geq 0}$ satisfying (2.7), the set Φ_C of all sequences ϕ satisfying $\Sigma_{j=k}^{\infty} |\phi_j| \leq C_k$, $k = 0, 1, 2, \ldots$, is compact in l^1, and every compact set in the l^1-norm topology is a subset of some Φ_C; see Theorems IV.8.9 and IV.13.3 of Dunford and Schwartz (1957).

Elements $\phi, \psi \in \Phi$ define linear filters $\phi(B) = \Sigma_{j=0}^{\infty} \phi_j B^j$, $\psi(B) = \Sigma_{j=0}^{\infty} \psi_j B^j$, where B denotes the backshift operator. Being absolutely summable, such filters are applicable to any covariance stationary time series y_t; for example, $\phi(B) y_t = \Sigma_{j=0}^{\infty} \phi_j y_{t-j}$ (convergence in mean square). If $F(\lambda)$ denotes the spectral measure of y_t, the second moment between $\phi(B) y_{t+j}$ and $\psi(B) y_t$ has the frequency domain formula

$$E(\phi(B) y_{t+j} \psi(B) y_t) = \int_{-\pi}^{\pi} e^{-ij\lambda} \phi(e^{i\lambda}) \psi(e^{-i\lambda}) dF(\lambda),$$

with $\phi(e^{i\lambda}) = \Sigma_{j=0}^{\infty} \phi_j e^{ij\lambda}$, the frequency response function of $\phi(B)$. If the sample second moments of the filtered series converge a.s. [i.p.] to their population second moments, we obtain

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-j} \phi(B) y_{t+j} \psi(B) y_t = \int_{-\pi}^{\pi} e^{-ij\lambda} \phi(e^{i\lambda}) \psi(e^{-i\lambda}) dF(\lambda) \qquad \text{a.s. [i.p.]}.$$

In the theorem below, we obtain an analogous integral formula for the limits of the sample second moments, uniformly over Φ, for the more general situation in which truncated versions of the filters are applied to arrays that satisfy CVG and N or N2. The existence of a frequency domain formula for the limits rests on the fact that, since the sample second moment limits γ_j, $j = 0, 1, \ldots$, in CVG necessarily form a positive semidefinite sequence, there is an increasing function $G(\lambda)$, $-\pi \leq \lambda \leq \pi$, having the property

$$\gamma_j = \int_{-\pi}^{\pi} e^{-ij\lambda} dG(\lambda), \qquad j = 0, 1, \ldots, \tag{2.8}$$

by a well-known Theorem of Herglotz. We note that $G(\lambda)$ is non-constant if and only if $\gamma_0 > 0$.

The following theorem provides conditions under which (2.1)–(2.3) and, in particular, CVG and N hold for the output of families of convolution filters applied to $y_t(T)$. Its proof is given in Appendix A. Many objective functions defining parameter estimators in time series analysis, like the prediction error criteria introduced in Section 1 or the Gaussian (pseudo) likelihood function, can either be brought into the form of sample second moments of convolution filters applied to the data or are built up from such expressions. Hence, Theorem 2.1 forms the basis for consistency proofs of such estimators. In particular, the consistency results in Findley, Pötscher and Wei (2000) are built upon Theorem 2.1.

THEOREM 2.1. *Given an array* $y_t(T), 1 \le t \le T,\ T = 1, 2, \ldots,$ *let* Φ *denote a set of sequences* $\phi = (\phi_0, \ \phi_1, \ldots)$ *for which* (2.6) – (2.7) *hold for some sequence* C_k. *For each* $\phi \in \Phi$, *define*

$$z_t(\phi, T) = \sum_{j=0}^{t-1} \phi_j y_{t-j}(T), \qquad 1 \le t \le T,\ \ T = 1, 2, \ldots. \tag{2.9}$$

(a) *If the array* $y_t(T)$ *has the property* CVGN, *then* (a1) *and* (a2) *hold:*

(a1) *The arrays* $z_t(\phi, T)$ *satisfy* (2.2)–(2.3) *(with* $A = \Phi$*) and also property* (2.1) *in the form*

$$\lim_{T\to\infty} \sup_{\phi,\psi\in\Phi} \left| T^{-1} \sum_{t=1}^{T-j} z_{t+j}(\phi, T) z_t(\psi, T) - \int_{-\pi}^{\pi} e^{-ij\lambda} \phi(e^{i\lambda}) \psi(e^{-i\lambda}) dG(\lambda) \right| = 0 \qquad a.s.\ [i.p.] \tag{2.10}$$

for each $j \ge 0$, *with* $G(\lambda)$ *as in* (2.8). *The limit functions*

$$\gamma_j(\phi, \psi) = \int_{-\pi}^{\pi} e^{-ij\lambda} \phi(e^{i\lambda}) \psi(e^{-i\lambda}) dG(\lambda) \tag{2.11}$$

are bounded and jointly continuous on Φ *w.r.t. coordinatewise convergence (and hence w.r.t. the* l^1*-norm). Specifically,*

$$\sup_{\phi,\psi\in\Phi} \left| \gamma_j(\phi, \psi) \right| \le C_0^2 \gamma_0, \tag{2.12}$$

and, given sequences $\phi^N, \psi^N, N = 1, 2, \ldots,$ *in* Φ *that converge coordinatewise to limits* $\phi, \psi \in \Phi$, *we have*

$$\lim_{N\to\infty} \gamma_j(\phi^N, \psi^N) = \gamma_j(\phi, \psi). \tag{2.13}$$

(a2) *For each* $\phi \in \Phi$, *let* $(\phi_{t0}, \phi_{t1}, \ldots, \phi_{t,t-1}, 0, 0, \ldots), t = 1, 2, \ldots,$ *denote a sequence of finite-length filters (not necessarily in* Φ*) approximating* ϕ *in such a way that the following uniform boundedness and approximation properties hold:*

$$\sup_{\phi\in\Phi} |\phi_{tj}| < \infty, \qquad 0 \le j \le t-1,\ \ t = 1, 2, \ldots, \tag{2.14}$$

$$\lim_{k\to\infty} \sup_{\phi\in\Phi} \sum_{t=k}^{\infty} \left(\sum_{j=0}^{t-1} |\phi_{tj} - \phi_j| \right)^2 = 0. \tag{2.15}$$

Then the arrays $\hat{z}_t(\phi, T)$ *defined for each* $\phi \in \Phi$ *by*

$$\hat{z}_t(\phi, T) = \sum_{j=0}^{t-1} \phi_{tj} y_{t-j}(T), \qquad 1 \le t \le T,\ \ T = 1, 2, \ldots,$$

approximate the arrays $z_t(\phi, T)$ *in the sense of* (2.4). *That is,*

$$\lim_{T\to\infty} \sup_{\phi\in\Phi} \frac{1}{T} \sum_{t=1}^{T} (z_t(\phi, T) - \hat{z}_t(\phi, T))^2 = 0 \qquad a.s.\ [i.p.].$$

Consequently, the arrays $\hat{z}_t(\phi, T)$ *satisfy* (2.2) – (2.3) *and also* (2.1) *for all* $j \geq 0$, *with limits* $\gamma_j(\phi, \psi)$ *given by* (2.11).

(b) *If the array* $y_t(T)$ *only has properties* CVG *and* N2, *then the arrays* $z_t(\phi, T)$ *defined in* (2.9) *satisfy all of the conclusions of part* (a1) *except possibly* (2.2).

REMARK 2.3 (Extensions of Theorem 2.1). (i) Suppose the array $y_t(T)$ only satisfies N1. If Φ is any set of real-valued sequences (not necessarily absolutely summable) such that (2.14) holds, then $\hat{z}_t(\phi, T)$ satisfies (2.2) as is easily seen. From the special case $\phi_{tj} = \phi_j, 0 \leq j \leq t-1$, one obtains that $z_t(\phi, T)$ satisfies (2.2) if $\sup_{\phi \in \Phi} |\phi_j| < \infty$ for $j \geq 0$.

(ii) Suppose Φ is as in Theorem 2.1 and $y_t(T)$ satisfies CVG and N2. If $\sup_{\phi\in\Phi} \sum_{i=0}^{t-1} (\phi_{ti} - \phi_i)^2 \to 0$ as $t \to \infty$ [which is weaker than (2.15)], then $\hat{z}_t(\phi, T)$ satisfies (2.3).

(iii) Theorem 2.1(a2) also holds (with essentially the same proof) if the finite-length filter approximating ϕ is allowed to depend on T, that is, $\phi_{tj} = \phi_{tj}(T)$, provided (2.14) is changed to $\sup_{T\geq 1} \sup_{\phi\in\Phi} |\phi_{tj}(T)| < \infty$ for $t \geq 1$, $j \geq 0$ and a $\limsup_{T\to\infty}$-operator is inserted between the limit- and the supremum-operators in (2.15).

3. Convergence of sample lagged second moments of residuals. In this section we provide conditions under which the residuals from a linear regression satisfy CVGN. Such results are of importance, for example, in case a time series model is fit to the residuals rather than to the original data, and consistency of the estimators for the time series model is to be established; cf. Findley, Pötscher and Wei (2000).

Suppose that we observe data Y_t, $t = 1, \dots, T$ that conforms to a regression model of the form

$$Y_t = AX_t + u_t, \tag{3.1}$$

in which the regressors X_t are nonstochastic column vectors of fixed dimension $d_x \geq 1$. Any method of determining an estimate A_T^* of A from $Y_1, \dots, Y_T$ defines an array of regression residuals $u_t^*(T) = Y_T - A_T^* X_T$ for $1 \leq t \leq T$. If the error process u_t has the property CVG, that is,

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-j} u_{t+j} u_t = \gamma_j \qquad \text{a.s. [i.p.] for } j = 0, 1, \dots, \tag{3.2}$$

it then automatically has property N by Remark 2.1, and the residuals $u_t^*(T)$ will have the property CVGN with γ_j given by (3.2) in view of Proposition 2.1 and Remark 2.2(i), provided $u_t^*(T)$ approximates u_t in the sense of (2.4), that is,

$$\frac{1}{T}\sum_{t=1}^{T}(u_t^*(T)-u_t)^2 \tag{3.3}$$
$$= (A_T^* - A)\left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)(A_T^*-A)' \to 0 \qquad \text{a.s. [i.p.]}$$

as $T \to \infty$ ($'$ denotes transpose). Given (3.2) and (3.3), Remark 2.2(iii) moreover shows that the difference between the sample lagged second moments of the errors u_t and the residuals $u_t^*(T)$ converges to zero uniformly in the lag, that is,

$$\max_{0\le j\le T-1}\left|\frac{1}{T}\sum_{t=1}^{T-j} u_{t+j}^*(T)u_t^*(T) - \frac{1}{T}\sum_{t=1}^{T-j} u_{t+j}u_t\right| \tag{3.4}$$
$$= O\left(\left(\frac{1}{T}\sum_{t=1}^{T}(u_t^*(T)-u_t)^2\right)^{\frac{1}{2}}\right) \qquad \text{a.s. [i.p.].}$$

The crucial condition (3.3) for verifying CVGN for the residuals is satisfied, for example, if the sequence $T^{-1}\sum_{t=1}^{T} X_t X_t'$ is bounded and $A_T^* \to A$ a.s. [i.p.]. However, when the most familiar estimator of A is used, namely the ordinary least squares estimator, we shall show in Sections 3.1 and 3.2 that no assumptions concerning X_t, and therefore none concerning the consistency of the least squares estimator, are needed when u_t in (3.1) has certain quite general properties. [Inconsistency of the least squares estimator occurs, e.g., with additive outlier regressors and level shift regressors; see Findley et al. (1998).]

With Γ^+ denoting the Moore-Penrose inverse of a square matrix Γ, consider the ordinary least squares estimator of A in (3.1) given by

$$\widehat{A}_T = \sum_{t=1}^{T} Y_t X_t' \left(\sum_{t=1}^{T} X_t X_t'\right)^{+}, \tag{3.5}$$

and the associated regression residual array given for $1 \le t \le T$ by

$$\widehat{u}_t(T) = Y_t - \widehat{A}_T X_t\,. \tag{3.6}$$

Using the reflexivity of the Moore-Penrose inverse, $\Gamma^+\Gamma\Gamma^+ = \Gamma^+$, condition (3.3) now becomes

$$\frac{1}{T}\sum_{t=1}^{T}(\widehat{u}_t(T)-u_t)^2 = \frac{1}{T}\sum_{t=1}^{T} u_t X_t'\left(\sum_{t=1}^{T} X_t X_t'\right)^{+}\sum_{t=1}^{T} X_t u_t \to 0 \qquad \text{a.s. [i.p.]} \tag{3.7}$$

as $T \to \infty$. The ordinary least squares residuals satisfy

$$\frac{1}{T}\sum_{t=1}^{T}(\widehat{u}_t(T)-u_t)^2 = \frac{1}{T}\sum_{t=1}^{T} u_t^2 - \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t(T)^2$$

which makes it clear that, when (3.2) holds for u_t, condition (3.7) is in fact not only sufficient but also necessary for the residuals $\widehat{u}_t(T)$ to satisfy CVGN [with γ_j the same as in (3.2)].

3.1. *Obtaining* $\mathrm{CVGN}_{a.s.}$. The set of assumptions yielding the a.s. version of the crucial condition (3.7) under our least restrictive moment condition is:

ASSUMPTION A. *The series u_t has a linear representation,*

$$u_t = \sum_{j=-\infty}^{\infty} c_j e_{t-j}, \qquad \left(\sum_{j=-\infty}^{\infty} c_j^2 < \infty\right) \tag{3.8}$$

in which e_t is a martingale difference sequence (w.r.t. an increasing sequence of σ-fields F_t) having constant variance $\sigma_e^2 = Ee_t^2$ and also the property that, for some $r > 2$,

$$\sup_{-\infty<t<\infty} E \mid e_t \mid^r < \infty. \tag{3.9}$$

Further, the spectral density $f_u(\lambda) = \left(\sigma_e^2/2\pi\right) \mid \sum_{j=-\infty}^{\infty} c_j e^{ij\lambda} \mid^2$ of u_t is essentially bounded. That is, for some finite constant M,

$$f_u(\lambda) \le M \qquad (\lambda\ a.e.). \tag{3.10}$$

The following result is proved in Appendix B.

THEOREM 3.1. *Suppose that Y_t has the form (3.1) with nonstochastic X_t and with u_t satisfying Assumption A. Then $(3.7)_{a.s.}$ holds for the residual array $\widehat{u}_t(T)$ defined by (3.6). Therefore, if u_t also has the property $(3.2)_{a.s.}$, then the array $\widehat{u}_t(T)$ has the property $\mathrm{CVGN}_{a.s.}$ [with the same limits γ_j as in (3.2)].*

For the classical situation in which the regression errors u_t are independent and identically distributed with finite second moment $\gamma_0 = Eu_t^2$, Gleser (1966) proved the a.s. convergence of $T^{-1}\sum_{t=1}^{T}\widehat{u}_t(T)^2$ to γ_0, assuming the invertibility of the matrices $\sum_{t=1}^{T} X_t X_t'$. Theorem 3.1 shows that, when the u_t have bounded higher-than-second moments, then Gleser's i.i.d. assumption can be weakened substantially. [Schmidt (1976) extended Gleser's result to the case of i.i.d. vectors u_t and showed that Gleser's invertibility assumption on $\sum_{t=1}^{T} X_t X_t'$ could be avoided by use of the Moore-Penrose inverse. Theorem 3.1 can be extended to the multivariate case easily.]

3.1.1. *Variants of Assumption A.* We first describe, in (a) below, a strengthening of Assumption A that yields $(3.2)_{a.s.}$ as well as $(3.7)_{a.s.}$ and therefore $\mathrm{CVGN}_{a.s}$ for $\widehat{u}_t(T)$. In (b), we present an alternative set of conditions that yields the same result without requiring (3.10).

(a) To obtain $(3.2)_{a.s.}$ in the simplest case $u_t = e_t$, with e_t as in Assumption A, we must achieve $T^{-1}\sum_{t=1}^{T} e_t^2 \overset{a.s.}{\to} \gamma_0$, most naturally with $\gamma_0 = \sigma_e^2$. Under

(3.9), a necessary and sufficient condition for this is

$$\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E(e_t^2 \mid F_{t-1}) = \sigma_e^2 \qquad \text{a.s.} \tag{3.11}$$

[This follows from Corollary 2.8.5 of Stout (1974) and Kronecker's lemma.] This condition is therefore basic. In an unpublished note, T. Mikosch has shown how $(3.2)_{a.s.}$ can be obtained from (3.8), (3.9) and (3.11) under additional constraints on the coefficients in (3.8), namely

$$c_j = 0 \qquad \text{for } j < 0 \text{ and } \sum_{j=1}^{\infty} j c_j^2 < \infty. \tag{3.12}$$

[His proof is a modification of the arguments given in 3.7 and 3.9 of Phillips and Solo (1992) for the case of i.i.d. e_t.] Hence, Assumption A augmented by (3.11), (3.12) yields $\text{CVGN}_{a.s.}$ for u_t and $\widehat{u}_t(T)$ (with the same limits γ_j).

(b) Alternatively, if we assume that the e_t in (3.8) are independent (with zero mean and constant variance) and that $r > 4$ in (3.9), then $\text{CVGN}_{a.s.}$ holds for u_t and for $\widehat{u}_t(T)$ (with the same limits γ_j) if $f(\lambda)$ is square integrable, a weaker assumption than (3.10). This follows from Theorem 2.1 of Findley and Wei (1993) via the Borel-Cantelli Lemma.

REMARK 3.1. Under Assumption A or the assumptions of (b), it follows from Theorems 1.1, 1.2 and Lemma 2.1 of Lai and Wei (1983) that the sample means of u_t, $1 \le t \le T$, converge to zero a.s. By Remark 2.2 (ii) and the results of Theorem 3.1 or (b), it follows that the same is true of the sample means of $\widehat{u}_t(T)$, $1 \le t \le T$.

REMARK 3.2. (a) Inspection of the proof of Theorem 3.1 shows that the expression in (3.7) is in fact $o(T^{\frac{2}{r}-1}\alpha_T^2)$ a.s. for any sequence $\alpha_T > 0$ such that $\sum_{T=1}^{\infty} T^{-1}\alpha_T^{-r} < \infty$ [e.g., $\alpha_T^2 = (\log T)^{\frac{2}{r}}(\log\log T)^{\frac{2+\beta}{r}}$ for $\beta > 0$]. This implies a rate of convergence in (3.4) with $u_t^*(T) = \widehat{u}_t(T)$. Under stronger assumptions, including conditions on the regressors, Chen and Ni (1989) obtain the better rate $O(T^{-1}\log\log T)$ a.s. for the expression in (3.7).

(b) If u_t is only assumed to be a mean zero, weakly stationary process with essentially bounded spectral density, the property $(3.7)_{a.s.}$ can still be obtained under suitable assumptions on the regressors and the rate of convergence of $T^{-1}\sum_{t=1}^{T} u_t^2$; see Theorem 3 of Hannan (1978). This theorem also yields rates of convergence in (3.7).

3.2. *Obtaining* $\text{CVGN}_{i.p.}$. For a mean zero, covariance stationary time series u_t satisfying $(3.2)_{i.p.}$, the essential boundedness of the spectral density, that is, (3.10), is sufficient for $\text{CVGN}_{i.p.}$ to hold for the residuals $\widehat{u}_t(T)$: With M as in (3.10) and d_x denoting the dimension of X_t, it is shown in Appendix B that

$$E\left\{\frac{1}{T}\sum_{t=1}^{T}(\widehat{u}_t(T) - u_t)^2\right\} \le \frac{2\pi d_x M}{T}. \tag{3.13}$$

This yields $(3.7)_{i.p.}$, and therefore $\mathrm{CVGN}_{i.p.}$ by Proposition 2.1 [with γ_j as in $(3.2)_{i.p.}$]. When the series u_t is Gaussian and $\sum_{t=1}^{T} X_t X_t'$ is invertible for some T, much less is required. In this case, continuity of the spectral measure $F(\lambda)$ of u_t on $(-\pi, \pi]$ yields CVGN for $\widehat{u}_t(T)$ with mean square convergence and with $\gamma_j = Eu_{t+j}u_t$, by Theorem 2.2 of Štulajter (1991), Theorem IV.3.6 of Hannan (1970) and Theorem III.9.6 of Zygmund (1968). Also (3.2) holds with almost sure as well as mean square convergence and $\gamma_j = Eu_{t+j}u_t$.

In either case, the sample means of $u_t, 1 \le t \le T$, converge to zero in probability. Hence, by Remark 2.2 (ii) the sample means of $\widehat{u}_t(T), 1 \le t \le T$, have the same property. For further results and a discussion of rates of convergence in $(3.7)_{i.p.}$ see Pötscher (1998).

4. Linear processes with infinite variance noise. Suppose y_t is a linear process, that is,

$$y_t = \sum_{i=0}^{\infty} c_i e_{t-i}, \tag{4.1}$$

where e_t is i.i.d. If e_t has zero mean and finite variance and the coefficients c_i are square-summable, then $\mathrm{CVG}_{a.s.}$ holds for y_t as an immediate consequence of the ergodic theorem; furthermore, $\mathrm{N1}_{a.s.}$ and $\mathrm{N2}_{a.s.}$ hold as explained in Remark 2.1. By contrast, if y_t is a linear process with infinite variance noise, CVG will typically not hold. In this section, we show that a large class of such processes can still be brought into the realm of the theory developed in the preceding sections through application of an appropriate rescaling; cf. also Section 5.1 below.

In the following we assume that the i.i.d. process e_t has an infinite second moment and regularly varying tail probabilities. More specifically, we assume

$$P\left(\left|e_t\right| > x\right) = x^{-\alpha} L(x), \qquad 0 < \alpha < 2, \tag{4.2}$$

where $L(x)$ is a slowly varying function at ∞, and

$$P(e_t > x) / P\left(\left|e_t\right| > x\right) \to \rho, \qquad 0 \le \rho \le 1, \text{ as } x \to \infty. \tag{4.3}$$

We note that (4.2) and (4.3) describe the fact that the distribution of e_t belongs to the domain of attraction of a (non-normal) stable distribution [Feller (1966), Theorems IX.8.1a and XVII.5.1a]. The coefficients c_i are assumed to satisfy

$$0 < \sum_{i=0}^{\infty} \left|c_i\right|^{\delta} < \infty \qquad \text{for some } \delta < \alpha,\ \delta \le 1. \tag{4.4}$$

Under (4.2)–(4.4) almost sure convergence holds in (4.1); see Cline (1983).

Set $\lambda_T = (T^{-1}\sum_{t=1}^{T} y_t^2)^{-1/2}$ if the r.h.s. is finite, and set $\lambda_T = 1$ otherwise. Define the array $y_t(T),\ 1 \le t \le T$, via

$$y_t(T) = \lambda_T y_t, \tag{4.5}$$

which amounts to a stochastic rescaling of the data. Making use of results in Davis and Resnick (1985), we now show that this array satisfies $\mathrm{CVGN}_{i.p.}$.

PROPOSITION 4.1. *Under* (4.1)–(4.4) *the array* $y_t(T)$ *defined by* (4.5) *satisfies* $\mathrm{CVGN}_{i.p.}$ *with* $\gamma_j = \sum_{i=0}^{\infty} c_i c_{i+j} / \sum_{i=0}^{\infty} c_i^2$ *where the* c_i *are as in* (4.1).

PROOF. From Theorem 4.2(i) in Davis and Resnick (1985) and the remark following that theorem, we can conclude that $a_T^{-2} \sum_{t=1}^{T} y_t^2$ converges in distribution to a nonnegative stable random variable, where a_T is defined in (2.1) of Davis and Resnick (1985). Since the limiting distribution (being stable) cannot have an atom at zero, the event $\{T^{-1} \sum_{t=1}^{T} y_t^2 > 0\}$ has probability approaching one. Consequently, the continuous mapping theorem and Theorem 4.2(i) of Davis and Resnick (1985) now establish $\mathrm{CVG}_{i.p.}$.

To establish $\mathrm{N1}_{i.p.}$ and $\mathrm{N2}_{i.p.}$, we first note that $T^{-1/2} a_T \lambda_T$ converges in distribution to the reciprocal of the square-root of the stable random variable mentioned above. Now $T^{-1/2} a_T \to \infty$, since the sequence a_T is regularly varying of index $1/\alpha$ [i.e., $a_T = T^{1/\alpha} L_1(T)$, where L_1 is a slowly varying function] with $0 < \alpha < 2$. Hence, λ_T converges to zero in probability. Then for any $\varepsilon > 0$ and $1 \le s \le T$ we have

$$
\begin{aligned}
P\left(T^{-1/2} \left|y_s(T)\right| > \varepsilon\right) &= P\left(T^{-1/2} \left|y_s\right| \lambda_T > \varepsilon\right) \\
&\le P\left(T^{-1/2} \left|y_s\right| > \varepsilon\right) + P\left(\lambda_T > 1\right) \qquad (4.6) \\
&= P\left(T^{-1/2} \left|y_1\right| > \varepsilon\right) + P\left(\lambda_T > 1\right),
\end{aligned}
$$

where we have made use of strict stationarity of y_t. Clearly, the r.h.s. of (4.6) converges to zero for $T \to \infty$, thus establishing the in probability version of the negligibility conditions. □

5. Some implications and extensions.

5.1. *Implications.* (i) In Sections 3 and 4 the property CVGN has been established for regression residuals and infinite variance processes. Consequently, any result in time series analysis that relies on this property as the only assumption on the data (as is the case for many consistency results) can now be applied to regression residuals and infinite variance processes. In particular, in Pötscher (1987) a theory of (generalized) consistency results for maximum likelihood and Whittle likelihood estimation of possibly misspecified ARMA models was given that only relies on property CVG as the only assumption on the data. That paper considered time series, not arrays, hence CVG coincides with CVGN; cf. Remark 2.1. It is easy to see that the results in Pötscher (1987) continue to hold for arrays, if CVG is replaced by CVGN. Hence, Theorem 3.1 and Proposition 4.1 show that the (generalized) consistency results of Pötscher (1987) are applicable when ARMA models are fit to regression residuals or linear processes with infinite variance noise using maximum likelihood or Whittle likelihood estimation. The same is true for the general invertible, short-memory models considered in Findley, Pötscher and Wei (2000) with parameter estimators optimizing either likelihood or forecast performance criteria. (The results in both papers also cover the case where the model being fit is misspecified in the sense that it is not capable of generating the asymptotic autocovariance structure of the data.)

(ii) Sections 4 and 5 of Dahlhaus (1997) (cf. especially the proof of Theorem 5.1) give assumptions and results sufficient for a proof that $\mathrm{CVGN}_{i.p.}$ holds for a broad class of locally stationary processes with mean zero. This observation, combined with the preceding remark, shows that the (generalized) consistency results in Pötscher (1987) and Findley, Pötscher and Wei (2000) apply when (misspecified) constant-parameter models are fit to locally stationary processes.

5.2. *Extensions.* (i) In property CVG we assumed the limits γ_j to be nonrandom for the sake of simplicity. The results of the paper extend quite straightforwardly to the case where the limits γ_j are allowed to be random. Such an extension is useful when considering non-ergodic processes, since such processes – after suitable renormalization – sometimes satisfy such a more general version of CVG.

(ii) In Section 1, we mentioned seasonal adjustment as a source of array data. This is true in a limited sense when a time series of length T is input to the linear X-11 seasonal adjustment procedure described in Findley et al. (1998). Then seasonally adjusted values, say $a_t(T), 1 \le t \le T$, depend on T only for $1 \le t \le M$ and for $T - M < t \le T$ for fixed finite M, and otherwise are the output of a finite-length, time invariant convolution filter. In this case, the adjusted series (after differencing) will automatically have the property CVGN if the input series has the property CVG (after differencing). If, instead of the X-11 procedure, the ARIMA model-based signal-extraction procedure discussed in Burman (1980) is used (and if the model has a moving average component), then all of the $a_t(T)$ depend on T, and results establishing CVGN for the differenced adjusted series in adequate generality have not yet been obtained. (The Tunnicliffe-Wilson method described by Burman for calculating the adjusted series can be used in a rather straightforward way to establish CVGN for the very restricted situation wherein the input series follows an ARMA model whose coefficients are known.)

(iii) The results of Section 2 (and their proofs) require only an appropriate change of notation in order to cover the case of n-dimensional vector time series with $n > 1$.

APPENDIX A: PROOFS OF THEOREM 2.1 AND PROPOSITION 2.1

We shall make repeated use of the following elementary result.

LEMMA A.1. *Suppose that the sequence of nonnegative extended real-valued random variables $U_T, T = 1, 2, \ldots,$ has the property that for every $k \ge 1$ and $T \ge 1$,*

$$U_T \le \eta_k V_T + W_{k,T}$$

holds, where η_k are nonnegative real numbers, and V_T and $W_{k,T}$ are nonnegative extended real-valued random variables, having the following properties:

(a) *The sequence V_T satisfies $V_T = O(1)$ a.s. $[O_p(1)]$.*
(b) *For each k, the sequence $W_{k,T}$ converges to 0 a.s. [i.p.] as $T \to \infty$.*
(c) $\lim_{k\to\infty} \eta_k = 0$.

Then

$$U_T \to 0 \qquad a.s.\ [i.p.].$$

PROOF. Without loss of generality we may assume $\eta_k > 0$ for all k. The convergence in probability assertion follows from the fact that for every $\delta > 0$

$$P(U_T \geq \delta) \leq P\left(V_T \geq \frac{\delta}{2\eta_k}\right) + P\left(W_{k,T} \geq \frac{\delta}{2}\right).$$

Indeed, for every $\varepsilon > 0$, there exist a $k = k(\varepsilon)$ and a T_ε such that $P(V_T \geq \delta/2\eta_k) < \varepsilon/2$ for all $T \geq T_\varepsilon$, and there is a $T_k \geq T_\varepsilon$ such that $P\left(W_{k,T} \geq \delta/2\right) < \varepsilon/2$ for $T \geq T_k$. Therefore $P(U_T \geq \delta) < \varepsilon$ for all $T \geq T_k$. Hence, $U_T \to 0$ i.p. The a.s. convergence assertion follows similarly. □

We note that Lemma A.1 and its proof remain valid even for non-measurable U_T, V_T and $W_{k,T}$, if convergence and boundedness in probability are interpreted in terms of the induced outer probability.

PROOF OF PROPOSITION 2.1. To obtain (2.2) and (2.3) for $\hat{z}_t(\alpha, T)$, observe that for $s = 1 + j$ or $s = T - j$ we have

$$\begin{aligned}\sup_{\alpha\in A} \frac{|\,\hat{z}_s(\alpha, T)\,|}{T^{\frac{1}{2}}} &\leq \sup_{\alpha\in A} \frac{|\,z_s(\alpha, T)\,|}{T^{\frac{1}{2}}} + \sup_{\alpha\in A} \frac{|\,\hat{z}_s(\alpha, T) - z_s(\alpha, T)\,|}{T^{\frac{1}{2}}} \\ &\leq \sup_{\alpha\in A} \frac{|\,z_s(\alpha, T)\,|}{T^{\frac{1}{2}}} \\ &\quad + \left\{\sup_{\alpha\in A} \frac{1}{T}\sum_{t=1}^{T} (\hat{z}_t(\alpha, T) - z_t(\alpha, T))^2\right\}^{\frac{1}{2}},\end{aligned}$$

with the first term of the final bound converging to zero a.s. [i.p.] by the respective negligibility conditions for $z_t(\alpha, T)$ assumed in the proposition. The second term converges to zero a.s. [i.p.] by (2.4).

To establish (2.5), observe that for all $j \geq 0$,

$$\begin{aligned}&\left|\frac{1}{T}\sum_{t=1}^{T-j} \hat{z}_{t+j}(\alpha, T)\hat{z}_t(\beta, T) - \frac{1}{T}\sum_{t=1}^{T-j} z_{t+j}(\alpha, T)z_t(\beta, T)\right| \\ &\quad = \frac{1}{T}\left|\sum_{t=1}^{T-j} (\hat{z}_{t+j}(\alpha, T) - z_{t+j}(\alpha, T))(\hat{z}_t(\beta, T) - z_t(\beta, T))\right. \\ &\qquad + \sum_{t=1}^{T-j} (\hat{z}_{t+j}(\alpha, T) - z_{t+j}(\alpha, T))z_t(\beta, T)\end{aligned}$$

$$
\begin{aligned}
&\qquad + \sum_{t=1}^{T-j} z_{t+j}(\alpha, T)(\hat{z}_t(\beta, T) - z_t(\beta, T))\Bigg| \\
&\le \sqrt{\frac{1}{T}\sum_{t=1}^{T-j}(\hat{z}_{t+j}(\alpha, T) - z_{t+j}(\alpha, T))^2} \\
&\qquad \times \left(\sqrt{\frac{1}{T}\sum_{t=1}^{T-j}(\hat{z}_t(\beta, T) - z_t(\beta, T))^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T-j} z_t^2(\beta, T)}\right) \\
&\qquad + \sqrt{\frac{1}{T}\sum_{t=1}^{T-j} z_{t+j}^2(\alpha, T))}\sqrt{\frac{1}{T}\sum_{t=1}^{T-j}(\hat{z}_t(\beta, T) - z_t(\beta, T))^2} \\
&\le \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{z}_t(\alpha, T) - z_t(\alpha, T))^2} \\
&\qquad \times \left(\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{z}_t(\beta, T) - z_t(\beta, T))^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T} z_t^2(\beta, T)}\right) \\
&\qquad + \sqrt{\frac{1}{T}\sum_{t=1}^{T} z_t^2(\alpha, T))}\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{z}_t(\beta, T) - z_t(\beta, T))^2}\,.
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
&\sup_{\alpha,\beta\in A}\left|\frac{1}{T}\sum_{t=1}^{T-j}\hat{z}_{t+j}(\alpha, T)\hat{z}_t(\beta, T) - \frac{1}{T}\sum_{t=1}^{T-j} z_{t+j}(\alpha, T)z_t(\beta, T)\right| \\
&\le \sqrt{\sup_{\alpha\in A}\frac{1}{T}\sum_{t=1}^{T}(\hat{z}_t(\alpha, T) - z_t(\alpha, T))^2} \qquad \text{(A.1)} \\
&\qquad \times \left(\sqrt{\sup_{\alpha\in A}\frac{1}{T}\sum_{t=1}^{T}(\hat{z}_t(\alpha, T) - z_t(\alpha, T))^2} + 2\sqrt{\sup_{\alpha\in A}\frac{1}{T}\sum_{t=1}^{T} z_t^2(\alpha, T)}\right).
\end{aligned}
$$

The expression $\sup_{\alpha\in A}\frac{1}{T}\sum_{t=1}^{T} z_t^2(\alpha, T)$ is bounded a.s. [i.p.] because (2.1) holds for $j = 0$ and because $\gamma_0(\alpha, \alpha)$ is assumed to be bounded on A. Therefore, the final bound on the r.h.s. of (A.1) converges to zero a.s. [i.p.] in view of (2.4). Now (A.1) and (2.1) yield (2.5). □

PROOF OF THEOREM 2.1. We first prove part (b). To establish (2.10) consider first the case $j = 0$ in (2.10). Note that, for any $k \ge 0$, the supremum term in (2.10) with $j=0$ is bounded by the sum of the expressions (A.2)–(A.4) below, where we use the conventions that $y_t(T)=0$ when $t\le 0$ and $\Sigma_{i=a}^{b}=0$ if $a>b$:

$$\sup_{\phi,\psi\in\Phi}\left|\int_{-\pi}^{\pi}\left(\sum_{j=0}^{k}\phi_j e^{ij\lambda}\right)\left(\sum_{j=0}^{k}\psi_j e^{-ij\lambda}\right)dG(\lambda) - \int_{-\pi}^{\pi}\phi(e^{ij\lambda})\psi(e^{-ij\lambda})dG(\lambda)\right|, \tag{A.2}$$

$$\sup_{\phi,\psi\in\Phi}\left|\frac{1}{T}\sum_{t=1}^{T}\sum_{j=0}^{k}\phi_j y_{t-j}(T)\sum_{j=0}^{k}\psi_j y_{t-j}(T) - \int_{-\pi}^{\pi}\left(\sum_{j=0}^{k}\phi_j e^{ij\lambda}\right)\left(\sum_{j=0}^{k}\psi_j e^{-ij\lambda}\right)dG(\lambda)\right|, \tag{A.3}$$

$$\sup_{\phi,\psi\in\Phi}\left|\frac{1}{T}\sum_{t=1}^{T}\sum_{j=0}^{k}\phi_j y_{t-j}(T)\sum_{j=0}^{k}\psi_j y_{t-j}(T) - \frac{1}{T}\sum_{t=1}^{T}\sum_{j=0}^{t-1}\phi_j y_{t-j}(T)\sum_{j=0}^{t-1}\psi_j y_{t-j}(T)\right|. \tag{A.4}$$

Clearly (A.2) is equal to

$$\begin{aligned}
&\sup_{\phi,\psi\in\Phi}\left|\int_{-\pi}^{\pi}\left(\sum_{j=0}^{k}\phi_j e^{ij\lambda}\sum_{j=k+1}^{\infty}\psi_j e^{-ij\lambda}+\sum_{j=0}^{\infty}\psi_j e^{ij\lambda}\sum_{j=k+1}^{\infty}\phi_j e^{-ij\lambda}\right)dG(\lambda)\right| \\
&\le \gamma_0\cdot\sup_{\phi,\psi\in\Phi}\left\{\sum_{j=0}^{\infty}|\phi_j|\sum_{j=k+1}^{\infty}|\psi_j|+\sum_{j=0}^{\infty}|\psi_j|\sum_{j=k+1}^{\infty}|\phi_j|\right\} \\
&\le 2\gamma_0 C_0 C_{k+1}.
\end{aligned} \tag{A.5}$$

Furthermore, (A.3) is bounded by

$$\begin{aligned}
&\sup_{\phi,\psi\in\Phi}\sum_{i,j=0}^{k}|\phi_i||\psi_j|\left|\frac{1}{T}\sum_{t=1}^{T}y_{t-i}(T)y_{t-j}(T)-\gamma_{|j-i|}\right| \\
&\le C_0^2\max_{0\le i,\,j\le k}\left|\frac{1}{T}\sum_{t=1}^{T}y_{t-i}(T)y_{t-j}(T)-\gamma_{|j-i|}\right| \\
&= C_0^2\max_{0\le i,\,j\le k}\left|\frac{1}{T}\sum_{t=1}^{T-\max(i,j)}y_{t+|j-i|}(T)y_t(T)-\gamma_{|j-i|}\right|.
\end{aligned} \tag{A.6}$$

The expression inside the absolute value in (A.4) can be written as the sum of

$$D_{k,T}^{(1)}(\phi,\psi)=\frac{1}{T}\sum_{t=k+2}^{T}\left(\sum_{i=0}^{k}\phi_i y_{t-i}(T)-\sum_{i=0}^{t-1}\phi_i y_{t-i}(T)\right)\sum_{j=0}^{k}\psi_j y_{t-j}(T)$$

and

$$D_{k,T}^{(2)}(\phi,\psi)=\frac{1}{T}\sum_{t=k+2}^{T}\sum_{i=0}^{t-1}\phi_i y_{t-i}(T)\left(\sum_{j=0}^{k}\psi_j y_{t-j}(T)-\sum_{j=0}^{t-1}\psi_j y_{t-j}(T)\right).$$

Observe that

$$\begin{aligned}\sup_{\phi,\psi\in\Phi}|D_{k,T}^{(1)}(\phi,\psi)| &\le \sup_{\phi,\psi\in\Phi}\frac{1}{T}\sum_{t=k+2}^{T}\sum_{i=k+1}^{t-1}\sum_{j=0}^{k}|\phi_i||\psi_j||y_{t-i}(T)||y_{t-j}(T)|\\ &= \sup_{\phi,\psi\in\Phi}\frac{1}{T}\sum_{j=0}^{k}\sum_{i=k+1}^{T-1}|\phi_i||\psi_j|\sum_{t=i+1}^{T}|y_{t-i}(T)||y_{t-j}(T)|\\ &\le \sup_{\phi,\psi\in\Phi}\sum_{j=0}^{k}\sum_{i=k+1}^{T-1}|\phi_i||\psi_j|\left(\frac{1}{T}\sum_{t=i+1}^{T}y_{t-i}^2(T)\right)^{1/2}\\ &\qquad\times\left(\frac{1}{T}\sum_{t=i+1}^{T}y_{t-j}^2(T)\right)^{1/2}\\ &\le\left(\frac{1}{T}\sum_{t=1}^{T}y_t^2(T)\right)C_0C_{k+1}.\end{aligned}\tag{A.7}$$

A similar argument shows that also

$$\sup_{\phi,\psi\in\Phi}|\,D_{k,T}^{(2)}(\phi,\psi)\,|\le\left(\frac{1}{T}\sum_{t=1}^{T}y_t^2(T)\right)C_0C_{k+1}.\tag{A.8}$$

Thus, for the supremum term in (2.10), the bounds (A.5)–(A.8) yield

$$\begin{aligned}&\sup_{\phi,\psi\in\Phi}\left|\frac{1}{T}\sum_{t=1}^{T}z_t(\phi,T)z_t(\psi,T)-\int_{-\pi}^{\pi}\phi\left(e^{i\lambda}\right)\psi\left(e^{-i\lambda}\right)dG(\lambda)\right|\\ &\quad\le 2\left(\gamma_0+\frac{1}{T}\sum_{t=1}^{T}y_t^2(T)\right)C_0C_{k+1}\\ &\qquad+C_0^2\max_{0\le i,j\le k}\left|\frac{1}{T}\sum_{t=1}^{T-\max(i,j)}y_{t+|j-i|}(T)y_t(T)-\gamma_{|j-i|}\right|.\end{aligned}$$

Hence Lemma A.1 applies, because of (2.7), CVG and N2, to yield (2.10) with $j=0$. The validity of (2.10) for $j>0$ is now established as follows: For each $\phi=(\phi_0,\phi_1,\dots)\in\Phi$ and each $l>0$, define $\phi[l]=(0,\dots 0,\phi_0,\phi_1,\dots)$ (l leading zeroes). For $z_t(\phi,T)$ given by (2.9) and for $z_t(\phi[l],T)=\sum_{i=0}^{t-1}\phi[l]_i\,y_{t-i}(T)$, we have $z_t(\phi[l],T)=z_{t-l}(\phi,T)$ for $l+1\le t\le T$ and $z_t(\phi[l],T)=0$ for $1\le t\le l$. Therefore, for each $T=1,2,\dots,$ we have

$$T^{-1}\sum_{t=1}^{T-l}z_{t+l}(\phi,T)z_t(\psi,T)=T^{-1}\sum_{t=1}^{T}z_t(\phi,T)z_t(\psi[l],T).$$

Now enlarge Φ to include all the filters $\phi[l]$ for some fixed $l > 0$, then the enlarged set is still a relatively compact subset of l^1 w.r.t. the norm topology, because

$$\sup_{\phi\in\Phi}\sum_{i=k}^{\infty}\left|\phi[l]_i\right| = \sup_{\phi\in\Phi}\sum_{i=\max(0,k-l)}^{\infty}\left|\phi_i\right| \leq C_{\max(0,k-l)},$$

by (2.6). Thus applying the already established relation (2.10) with $j = 0$ to this enlarged set gives (2.10) for the original set Φ and with $j = l$. This completes the proof of (2.10). We next prove (2.3). Clearly, for any $k \geq 0$ and $j \geq 0$

$$\begin{aligned}\frac{\left|z_{T-j}(\phi, T)\right|}{T^{\frac{1}{2}}} &\leq \left|T^{-\frac{1}{2}}\sum_{i=0}^{k}\phi_i y_{T-j-i}(T)\right| + \left|T^{-\frac{1}{2}}\sum_{i=k+1}^{T-j-1}\phi_i y_{T-j-i}(T)\right| \\ &\leq T^{-\frac{1}{2}}\sum_{i=0}^{k}\left|\phi_i\right|\left|y_{T-j-i}(T)\right| + \left(\sum_{i=k+1}^{T-j-1}\phi_i^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{i=k+1}^{T-j-1}y_{T-j-i}^2(T)\right)^{\frac{1}{2}}.\end{aligned}$$

Hence,

$$\sup_{\phi\in\Phi}\frac{\left|z_{T-j}(\phi, T)\right|}{T^{\frac{1}{2}}} \leq C_0\max_{0\leq i\leq k}\frac{\left|y_{T-j-i}(T)\right|}{T^{\frac{1}{2}}} + C_{k+1}\left(\frac{1}{T}\sum_{t=1}^{T}y_t^2(T)\right)^{\frac{1}{2}}.$$

The first term on the r.h.s. goes to zero a.s. [i.p.] as $T \to \infty$ by N2 and the square root in the second term is $O(1)$ a.s. $\left[O_p(1)\right]$ because of CVG. Since $C_{k+1} \to 0$ as $k \to \infty$ by (2.7), Lemma A.1 now establishes (2.3). The bound (2.12) follows from

$$\sup_{\phi,\psi\in\Phi}\left|\gamma_j(\phi, \psi)\right| \leq \sup_{\phi\in\Phi}\int_{-\pi}^{\pi}\left|\phi\left(e^{i\lambda}\right)\right|^2 dG(\lambda) \leq C_0^2\gamma_0 .$$

Continuity w.r.t. coordinatewise convergence on Φ follows easily from the observation that under (2.6) – (2.7) coordinatewise convergence of $\phi^N \in \Phi$ to $\phi \in \Phi$ implies l^1-norm convergence and hence uniform convergence on $[-\pi, \pi]$ of the sequence of functions $\phi^N\left(e^{i\lambda}\right)$ to $\phi\left(e^{i\lambda}\right)$. This completes the proof of (b).

Part (a1), with its additional assertion that $z_t(\phi, T)$, has property (2.2) if N1 holds, follows from (2.6) – (2.7) and (2.9), since (2.6) implies $\sup_{\phi\in\Phi}\left|\phi_j\right| < \infty$ for $j \geq 0$.

It remains to prove part (a2). By Theorem 2.1(a1) and Proposition 2.1 it suffices to verify (2.4) for $z_t(\phi, T)$ and $\hat{z}_t(\phi, T)$, setting $A = \Phi$ and $J = \{0, 1, \ldots\}$ in the Proposition. Observe that (2.1) then reduces to (2.10) and that, by (2.12), the integrals in (2.10) are bounded.

Set $\delta_{ti}^{\phi} = \phi_{ti} - \phi_i, 0 \leq i \leq t-1$. Then

$$\hat{z}_t(\phi, T) - z_t(\phi, T) = \sum_{i=0}^{t-1}\delta_{ti}^{\phi}y_{t-i}(T).$$

To verify (2.4), we have therefore to show that

$$\text{(A.9)} \qquad \lim_{T\to\infty} \sup_{\phi\in\Phi} \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=0}^{t-1} \delta_{ti}^{\phi} y_{t-i}(T) \right)^2 = 0 \qquad \text{a.s. [i.p.].}$$

For fixed $k \geq 1$,

$$\text{(A.10)} \qquad \begin{aligned} &\sup_{\phi\in\Phi} \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=0}^{t-1} \delta_{ti}^{\phi} y_{t-i}(T) \right)^2 \\ &\quad \leq \sup_{\phi\in\Phi} \frac{1}{T} \sum_{t=1}^{k} \left(\sum_{i=0}^{t-1} \delta_{ti}^{\phi} y_{t-i}(T) \right)^2 \\ &\qquad + \sup_{\phi\in\Phi} \frac{1}{T} \sum_{t=k+1}^{T} \left(\sum_{i=0}^{t-1} \delta_{ti}^{\phi} y_{t-i}(T) \right)^2 \\ &\quad \leq \sup_{1\leq t\leq k} \frac{\mid y_t(T) \mid^2}{T} \cdot \sup_{\phi\in\Phi} \sum_{t=1}^{k} \left(\sum_{i=0}^{t-1} \mid \delta_{ti}^{\phi} \mid \right)^2 \\ &\qquad + \sup_{1\leq t\leq T} \frac{\mid y_t(T) \mid^2}{T} \cdot \sup_{\phi\in\Phi} \sum_{t=k+1}^{T} \left(\sum_{i=0}^{t-1} \mid \delta_{ti}^{\phi} \mid \right)^2 . \end{aligned}$$

The first factor of the first term of this final bound converges to 0 by N1, and its second factor is finite for fixed k, due to (2.6) and (2.14). For the second term, it follows from

$$\text{(A.11)} \qquad \sup_{1\leq t\leq T} \frac{\mid y_t(T) \mid^2}{T} \leq \frac{1}{T} \sum_{t=1}^{T} y_t(T)^2 \to \gamma_0 \qquad \text{a.s. [i.p.]}$$

that its first factor is a.s. [i.p.] bounded, whereas its second factor is bounded by the supremum in (2.15) which converges to 0 as $k \to \infty$. Thus Lemma A.1 applies to (A.10) and yields (A.9). This completes the proof of Theorem 2.1. □

REMARK A.1. The suprema of families of random variables in parts (a1) and (b) of Theorem 2.1 (and its proof) are clearly measurable, since they are of the form $\sup_{\phi,\psi\in\Phi} f(\omega, \phi, \psi)$, where $f(\omega, ., .)$ is continuous and $f(., \phi, \psi)$ is $\mathfrak{A}$-measurable, and since Φ is separable. The corresponding suprema in part (a2) (and its proof) are all of the form $\sup_{\phi,\psi\in\Phi} \left| Y' F(\phi, \psi) Y - h(\phi, \psi) \right|$ or $\sup_{\phi\in\Phi} \left| f'(\phi) Y \right|$, where $F(\phi, \psi)$ is a $T \times T$ matrix valued function, f is a $T \times 1$ vector valued function, and h is a real-valued function. The $T \times 1$ vector Y is given by $(y_1(T), \dots, y_T(T))'$. Define $M = \{(F(\phi, \psi), h(\phi, \psi)) : \phi, \psi \in \Phi\}$ which is (isomorphic to) a subset of an Euclidean space of appropriate dimension. Consequently, M is separable. Let $N \subseteq M$ be a countable dense subset. Then

$$\sup_{\phi,\psi\in\Phi} \left| Y' F(\phi, \psi) Y - h(\phi, \psi) \right| = \sup_{(C,d)\in M} \left| Y' C Y - d \right| = \sup_{(C,d)\in N} \left| Y' C Y - d \right|,$$

the latter equality being true since $|Y'CY - d|$ is continuous in (C, d) and since N is dense in M. This establishes measurability of the supremum, since N is countable. The argument for $\sup_{\phi\in\Phi}|f'(\phi)Y|$ is analogous.

In Proposition 2.1 the analogous suprema need not be measurable in general. In the case of non-measurability, convergence in probability is to be understood w.r.t. the induced outer probability.

APPENDIX B: DERIVATIONS OF THEOREM 3.1 AND (3.13)

PROOF OF THEOREM 3.1. It suffices to verify $(3.7)_{a.s.}$. Let $S(T)$ denote a matrix of order d_x such that $S(T)'S(T) = (\sum_{t=1}^{T} X_t X_t')^+$ and define $c(t, T) = S(T)X_t,\ 1 \le t \le T$. Then

$$\text{(B.1)} \qquad \begin{aligned} \sum_{t=1}^{T} u_t X_t' \left(\sum_{t=1}^{T} X_t X_t'\right)^+ \sum_{t=1}^{T} X_t u_t &= \left(\sum_{t=1}^{T} u_t c(t, T)\right)' \left(\sum_{t=1}^{T} u_t c(t, T)\right) \\ &= \sum_{j=1}^{d_x} \left(\sum_{t=1}^{T} c_j(t, T) u_t\right)^2, \end{aligned}$$

where $c_j(t, T)$ is the jth coordinate of $c(t, T)$. Thus it suffices to verify

$$\text{(B.2)} \qquad T^{-\frac{1}{2}} \left(\sum_{t=1}^{T} c_j(t, T) u_t\right) \to 0 \qquad \text{a.s.}$$

for each $1 \le j \le d_x$. By Lemma 2.1 and Theorem 1.2 of Lai and Wei (1983), a time series u_t that satisfies Assumption A is an S_r-system. That is, there is a finite constant C_r such that

$$\text{(B.3)} \qquad E\left|\sum_{t=1}^{T} c_j(t, T) u_t\right|^r \le C_r \left(\sum_{t=1}^{T} c_j^2(t, T)\right)^{\frac{r}{2}},$$

with r as in (3.9). Observe that

$$\text{(B.4)} \qquad \begin{aligned} \sum_{t=1}^{T} c_j^2(t, T) \le \sum_{t=1}^{T} \sum_{j=1}^{d_x} c_j^2(t, T) &= \sum_{t=1}^{T} c(t, T)'c(t, T) = \sum_{t=1}^{T} X_t' S(T)'S(T) X_t \\ &= \operatorname{trace}\left\{\left(\sum_{t=1}^{T} X_t X_t'\right)\left(\sum_{t=1}^{T} X_t X_t'\right)^+\right\} \le d_x. \end{aligned}$$

Applying (B.4) to (B.3), we obtain

$$E\left|\sum_{t=1}^{T} c_j(t, T) u_t\right|^r \le C_r d_x^{\frac{r}{2}}.$$

Because $r > 2$, a simple application of the Borel-Cantelli lemma yields (B.2). □

PROOF OF (3.13). Observe from (B.1) and (B.4) that

$$
\begin{aligned}
E\left\{\sum_{t=1}^{T} u_t X_t' \left(\sum_{t=1}^{T} X_t X_t'\right)^{+} \sum_{t=1}^{T} X_t u_t\right\} &= \sum_{j=1}^{d_x} E\left\{\sum_{t=1}^{T} c_j(t, T) u_t\right\}^2 \\
&= \sum_{j=1}^{d_x} \int_{-\pi}^{\pi} \left|\sum_{t=1}^{T} c_j(t, T) e^{it\lambda}\right|^2 f_u(\lambda)\, d\lambda \\
&\le 2\pi M \sum_{j=1}^{d_x} \sum_{t=1}^{T} c_j^2(t, T) \le 2\pi M d_x,
\end{aligned}
$$

an inequality equivalent to (3.13). □

Acknowledgments. The material in this and the companion paper [Findley, Pötscher and Wei (2000)] was circulated earlier in a working paper entitled "Convergence Results for the Modeling of Time Series Arrays by Multistep Prediction or Likelihood Methods." The authors thank Daming Xu for his careful reading of an early draft of this paper and Martina Edl and Birgit Ewald for skillfully typing extensive revisions.

REFERENCES

BLOOMFIELD, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* **60** 217–226.

BURMAN, J. P. (1980). Seasonal adjustment by signal extraction. *J. Roy. Statist. Soc. Ser. A* **143** 321–337.

CHEN, Z.-G. and NI, J.-Y. (1989). Subset regression time series and its modeling procedures. *J. Multivariate Anal.* **31** 266–288.

CLINE, D. (1983). Infinite series of random variables with regularly varying tails. Technical Report 83-24, Institute of Applied Mathematics and Statistics, Univ. British Columbia.

DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37.

DAVIS, R. A. and RESNICK, S. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Probab.* **13** 179–195.

DUNFORD, N. and SCHWARTZ, J. T. (1957). *Linear Operators*, *Part I*. Wiley, New York.

FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications* **2.** Wiley, New York.

FINDLEY, D. F. (1983). On the use of multiple models for multi-period forecasting. *Proc. Business and Economic Statistics Section* 528–531. Amer. Statist. Assoc., Alexandria, VA.

FINDLEY, D. F. (1991). Convergence of finite multistep predictors from incorrect models and its role in model selection. *Note Mat.* **11** 145–155.

FINDLEY, D. F., MONSELL, B. C., BELL, W. R., OTTO, M. C. and CHEN, B.-C. (1998). New capabilities and methods of the X-12-ARIMA seasonal adjustment program (with discussion). *J. Bus. Econom. Statist.* **16** 127–177.

FINDLEY, D. F., PÖTSCHER, B. M. and WEI, C. Z. (2000). Modeling of time series arrays by multistep prediction or likelihood methods. *J. Econometrics*. To appear.

FINDLEY, D. F. and WEI, C.-Z. (1993). Moment bounds for deriving time series CLT's and model selection procedures. *Statist. Sinica* **3** 453–480.

GLESER, L. J. (1966). Correction to "On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters." *Ann. Math. Statist.* **37** 1053–1054.

HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York.

HANNAN, E. J. (1978). Rates of convergence for time series regression. *Adv. Appl. Probab.* **10** 740–743.

LAI, T. L. and WEI, C.-Z. (1983). Lacunary systems and generalized linear processes. *Stochastic Proces. Appl.* **14** 187–199.

PARZEN, E. (1962). Spectral analysis of asymptotically stationary time series. *Bull. Inst. Internat. Statist.* **39** 87–103.

PHILLIPS, P. C. B. and SOLO, V. (1992). Asymptotics for linear processes. *Ann. Statist.* **20** 971–1001.

PÖTSCHER, B. M. (1987). Convergence results for maximum likelihood type estimators in multivariable ARMA models. *J. Multivariate Anal.* **21** 29–52.

PÖTSCHER, B. M. (1998). Solution to Problem 97.4.2: Asymptotic properties of the least-squares estimator of the variance in a linear model. *Econometric Theory* **14** 527–533.

PÖTSCHER, B. M. and PRUCHA, I. R. (1997). *Dynamic Nonlinear Econometric Models: Asymptotic Theory.* Springer, Berlin.

SCHMIDT, W. H. (1976). Strong consistency of variance estimation and asymptotic theory for tests of the linear hypothesis in multivariate linear models. *Statistics* **7** 701–705.

STOUT, W. F. (1974). *Almost Sure Convergence.* Academic Press, New York.

STULAJTER, F. (1991). Consistency of linear and quadratic least squares estimators in regression models with covariance stationary errors. *Appl. Math.* **36** 149–155.

TIAO, G. C. and XU, D. (1993) Robustness of maximum likelihood estimates for multistep predictions: the exponential smoothing case. *Biometrika* **80** 623–641.

ZYGMUND, A. (1968). *Trigonometric Series, Vols. I and II.* Cambridge Univ. Press.

D. F. FINDLEY
STATISTICAL RESEARCH DIVISION
BUREAU OF THE CENSUS
WAHINGTON, DC
E-MAIL: david.f.findley@census.gov

B. M. PÖTSCHER
DEPARTMENT OF STATISTICS
UNIVERSITY OF VIENNA
UNIVERSITÄTSSTRASSE 5
A-1010 VIENNA
AUSTRIA
E-MAIL: benedikt.poetscher@univie.ac.at

C.-Z. WEI
INSTITUTE OF STATISTICAL SCIENCE
ACADEMIA SINICA
TAIPEI
TAIWAN, R.O.C.
E-MAIL: czw@stat.sinica.edu.tw