Statistica Sinica 3(1993), 453-480

MOMENT BOUNDS FOR DERIVING TIME SERIES CLT'S AND MODEL SELECTION PROCEDURES

David F. Findley and Ching-Zong Wei

Bureau of the Census and Academia Sinica

Abstract: We establish a $2p$th moment bound ($p \geq 1$) for quadratic forms, and a pth moment bound for an important ratio of such forms, for vector time series having estimated regression mean functions and stationary, square summable autocovariances. The mean-centered series are assumed to have a joint linear representation in terms of a martingale difference sequence with bounded conditional $4p$th moments. Applications discussed include a quite general Central Limit Theorem for sample covariances and the derivation of several model selection methods for stochastic regression, among them methods concerned with multi-step-ahead forecasting performance.

Key words and phrases: Sample covariances, stationary regressor selection, Vuong's test, multi-step-ahead forecast-model comparisons, quadratic forms, nonstationary means.

1. Introduction

Let $X^*(t)$ and $Y^*(t)$ be vector-valued time series whose mean functions are linear functions of vector sequences $Z_X(t)$ and $Z_Y(t)$ respectively,

$$\mathcal{E}X^*(t) = A_X Z_X(t), \quad \mathcal{E}Y^*(t) = A_Y Z_Y(t). \tag{1.1}$$

We assume that the mean zero series

$$X(t) \equiv X^*(t) - A_X Z_X(t), \quad Y(t) \equiv Y^*(t) - A_Y Z_Y(t), \tag{1.2}$$

are jointly covariance stationary. If the coefficients in (1.1) are unknown, we shall be concerned only with the situation in which they are estimated by ordinary least squares, perhaps as a prelude to investigating the strength of a linear relation between the unobserved series $X(t)$ and $Y(t)$, such as the relation

$$Y(t) = DX(t) + E(t), \tag{1.3}$$

with $D \equiv \Gamma_{YX}\Gamma_{XX}^{-1}$, where $\Gamma_{YX} \equiv \mathcal{E}Y(t)X(t)'$ and $\Gamma_{XX} \equiv \mathcal{E}X(t)X(t)'$, assuming

$$\det \Gamma_{XX} > 0. \tag{1.4}$$

Such least squares estimation gives rise to quadratic forms $\sum_{s,t=1}^{T} x(s)q(s,t)y(t)$ in entries $x(t)$, $y(t)$ of $X(t)$ and $Y(t)$ with nonrandom $q(s,t)$. In this paper we derive and apply $2p$th absolute moment bounds ($p \geq 1$) such as

$$\mathcal{E}\left|\sum_{s,t=1}^{T}\left\{x(s)q(s,t)y(t)-\mathcal{E}x(s)q(s,t)y(t)\right\}\right|^{2p} \leq M_p\left\{\mathcal{E}\left|\sum_{s,t=1}^{T}\tilde{x}(s)q(s,t)\tilde{y}(t)\right|^{2}\right\}^{p},$$

in which M_p is a constant not dependent on T or on the coefficients $q(s,t)$, and $\tilde{x}(t)$ and $\tilde{y}(t)$ denote *mutually independent* time series having the same mean and autocovariances as $x(t)$ and $y(t)$ respectively. This result, which is reformulated as (2.6) below, makes it possible to obtain a moment bound like (1.10) and a Central Limit Theorem like Theorem 3.1 for covariance statistics of regression residuals.

Throughout the paper, Γ' denotes the transpose of a matrix Γ and, when Γ is square, Γ^{+} denotes its Moore-Penrose inverse. Define, for $T = 1, 2, \ldots$, the ordinary least squares A_X-estimator

$$\hat{A}_X(T) \equiv \sum_{t=1}^{T} X^{*}(t)Z_X(t)'\left[\sum_{t=1}^{T} Z_X(t)Z_X(t)'\right]^{+} \tag{1.5}$$

and the associated estimator of $X(t)$,

$$\hat{X}(t;T) \equiv X^{*}(t) - \hat{A}_X(T)Z_X(t),$$

for $1 \leq t \leq T$. As no confusion is likely to result, we shall abbreviate $\hat{X}(t;T)$ by $\hat{X}(t)$. Define $\hat{A}_Y(T)$ and $\hat{Y}(t) \equiv \hat{Y}(t;T)$ similarly, and also the estimates

$$\hat{D}(T) \equiv \sum_{t=1}^{T} \hat{Y}(t)\hat{X}(t)'\left[\sum_{t=1}^{T} \hat{X}(t)\hat{X}(t)'\right]^{+} \tag{1.6}$$

and

$$\hat{E}(t;T) \equiv \hat{Y}(t) - \hat{D}(T)\hat{X}(t),$$

the latter to be abbreviated $\hat{E}(t)$.

Let $x(t)$ and $y(t)$ denote coordinate entries of $X(t)$ and $Y(t)$, and let $\varepsilon(t)$ and $\eta(t)$ denote entries of the weakly stationary process $E(t)$. Also, let $\hat{x}(t)$, $\hat{y}(t)$, Γ_{xy}, $\Gamma_{\varepsilon\eta}$, $\hat{\varepsilon}(t)$, $\hat{\eta}(t)$ denote the corresponding entries of $\hat{X}(t)$, $\hat{Y}(t)$, Γ'_{YX}, $\Gamma_{EE} \equiv \mathcal{E}E(t)E(t)'$, $\hat{E}(t)$, and let $\hat{A}_x(T)$ denote the row of $\hat{A}_X(T)$ corresponding to $x(t)$, etc. Thus, for example, $\hat{x}(t) = x(t) - \hat{A}_x(T)Z_X(t)$. For a variety of purposes, including model selection theory, it is useful to have information about the boundedness in T of absolute moments of the covariance statistics

$$\hat{\tau}_{xy}(T) \equiv T^{-1/2}\sum_{t=1}^{T}\left(\hat{x}(t)\hat{y}(t) - \Gamma_{xy}\right) \tag{1.7}$$

and

$$\hat{r}_{\varepsilon\eta}(T) \equiv T^{-1/2} \sum_{t=1}^{T} \left(\hat{\varepsilon}(t)\hat{\eta}(t) - \Gamma_{\varepsilon\eta}\right). \tag{1.8}$$

For our basic results, such as (1.10) and (1.11) below, we only require the spectral density functions of the coordinate processes to be square integrable,

$$\int_{-\pi}^{\pi} f_{xx}^2(\lambda)d\lambda < \infty, \quad \int_{-\pi}^{\pi} f_{yy}^2(\lambda)d\lambda < \infty, \tag{1.9}$$

or, equivalently, the autocovariance sequences $\Gamma_{xx}(j)$, $\Gamma_{yy}(j)$, $j = 0, \pm 1, \ldots$, to be square summable. Thus, some long-memory processes are covered. For any real number $p > 0$, let $\|\cdot\|_p$ denote the pth absolute moment norm, defined for every random variate v by $\|v\|_p \equiv \{\mathcal{E}|v|^p\}^{1/p}$. We shall assume $x(t)$ and $y(t)$ have linear representations in terms of a conditional-covariance stationary martingale difference sequence with bounded conditional $4p$th moments (see (M1–3) below), and prove two general moment-bound results, presented in Sections 2 and 4. We use these to establish the bound

$$\sup_{T\geq 1} \|\hat{r}_{xy}(T)\|_{2p} \leq M_p \left\{ \int_{-\pi}^{\pi} f_{xx}^2(\lambda)d\lambda \cdot \int_{-\pi}^{\pi} f_{yy}^2(\lambda)d\lambda \right\}^{1/4} \tag{1.10}$$

for some finite constant M_p, and also that

$$\sup_{T\geq 1} \|\hat{r}_{\varepsilon\eta}(T)\|_p < \infty, \tag{1.11}$$

holds, both with *no restrictions* on the mean-regressor functions $Z_X(t)$ and $Z_Y(t)$ in (1.1).

One consequence of (1.11) is the weak consistency of a natural estimator of Γ_{EE},

$$T^{-1} \sum_{t=1}^{T} \hat{E}(t)\hat{E}(t)' \xrightarrow{\text{pr}} \Gamma_{EE}. \tag{1.12}$$

This provides the foundation for a model selection procedure for comparing $\hat{X}(t)$ with other regressors for $\hat{Y}(t)$, as we explain in Section 6. A generalization of the linear regression version of the hypotheses testing procedure of Vuong (1989) to the case of non-i.i.d. data is also presented there, as a corollary of a Central Limit Theorem for $\hat{r}_{\varepsilon\eta}$. This CLT follows from a quite general CLT for sample serial covariances which is derived in Section 3, using results related to (1.10) and (1.11). Section 5 considers how (1.10) is affected by linear filtering, especially by multistep-ahead-forecasting with ARMA models, an analysis which helps to justify the forecast-model selection procedure utilized in Findley (1991a).

The stochastic nature of the generalized inverse factor in $\hat{D}(T)$ is what causes the moment-order to be lower in (1.11) than in (1.10) (p versus $2p$) and makes it difficult to find an upper bound with a simple expression. For convenience of later reference, we close this section by stating, in general form, the basic least squares formula from which the role of these inverse terms becomes apparent. Given vectors $u(t)$, $v(t)$, $w(t)$, $t = 1, \dots, T$, define the least squares regression coefficients

$$\hat{C}_u(T) \equiv \sum_{t=1}^{T} u(t)w(t)'\Big[\sum_{t=1}^{T} w(t)w(t)'\Big]^{+}$$

and

$$\hat{C}_v(T) \equiv \sum_{t=1}^{T} v(t)w(t)'\Big[\sum_{t=1}^{T} w(t)w(t)'\Big]^{+}.$$

Then, for any matrices C_u and C_v having the same dimensions as these, we have the algebraic identities

$$\begin{aligned}
&\sum_{t=1}^{T}\Big(u(t) - C_u w(t)\Big)\Big(v(t) - C_v w(t)\Big)' \\
&\quad - \sum_{t=1}^{T}\Big(u(t) - \hat{C}_u(t)w(t)\Big)\Big(v(t) - \hat{C}_v(T)w(t)\Big)' \\
&= \Big(\hat{C}_u(T) - C_u\Big)\sum_{t=1}^{T} w(t)w(t)'\Big(\hat{C}_v(T) - C_v\Big)' \\
&= \Big\{\sum_{t=1}^{T}\Big(u(t)-C_u w(t)\Big)w(t)'\Big\}\Big[\sum_{t=1}^{T} w(t)w(t)'\Big]^{+}\Big\{\sum_{t=1}^{T} w(t)\Big(v(t)-C_v w(t)\Big)'\Big\}, \qquad (1.13)
\end{aligned}$$

the last equality arising from the reflexive property, $\Gamma^{+}\Gamma\Gamma^{+} = \Gamma^{+}$, of Moore-Penrose inverses (see Rao (1973, p.26)).

2. The First Moment Bound Theorem and the Proof of (1.10)

We shall always assume that all entries $x(t)$ and $y(t)$ satisfy (1.9) and have linear representations

$$\begin{aligned}
x(t) &= \sum_{j=-\infty}^{\infty} a(j)\alpha(t-j) \\
y(t) &= \sum_{j=-\infty}^{\infty} b(j)\beta(t-j)
\end{aligned} \qquad (2.1)$$

in terms of the same F_t-measurable random vectors $\alpha(t)$ and $\beta(t)$ of dimensions $d(\alpha)$ and $d(\beta)$, respectively, where F_t, $-\infty < t < \infty$ is an increasing sequence of

σ-fields of events. Further, we assume $\alpha(t)$ and $\beta(t)$ have the martingale difference property (M1) below as well as the conditional moment properties (M2)–(M3), all with probability 1:

(M1). $\mathcal{E}\{\alpha(t)|F_{t-1}\} = 0$; $\mathcal{E}\{\beta(t)|F_{t-1}\} = 0$.

(M2i). *There is a constant* $d(\alpha) \times d(\beta)$-*matrix* $\Sigma = (\Sigma_{ij})$ *such that*

$$\mathcal{E}\Big\{\alpha(t)\beta(t)'|F_{t-1}\Big\} = \Sigma.$$

(M2ii). *The conditional variance matrices* $E\{\alpha(t)\alpha(t)'|F_{t-1}\}$ *and* $E\{\beta(t)\beta(t)'|F_{t-1}\}$ *are the identity matrices* I_α *and* I_β *of their respective orders,*

$$\begin{aligned}\mathcal{E}\Big\{\alpha(t)\alpha(t)'|F_{t-1}\Big\} &= I_\alpha \\ \mathcal{E}\Big\{\beta(t)\beta(t)'|F_{t-1}\Big\} &= I_\beta.\end{aligned} \tag{2.2}$$

(M3). *There is a finite constant* C_p *such that* (2.3) *holds for some* $p \geq 1$,

$$\begin{aligned}\sup_{-\infty<t<\infty} \mathcal{E}\Big\{(\alpha'(t)\alpha(t))^{2p}|F_{t-1}\Big\} &\leq C_p \\ \sup_{-\infty<t<\infty} \mathcal{E}\Big\{(\beta'(t)\beta(t))^{2p}|F_{t-1}\Big\} &\leq C_p.\end{aligned} \tag{2.3}$$

Remark. The essential part of (M2) is the assumption that the conditional covariances are constant. As Hannan (1987) has discussed, this is not an entirely natural assumption, unless $\alpha(t)$, $\beta(t)$ are independent of $\alpha(s)$, $\beta(s)$ for $s \neq t$, but it is commonly made, because it helps to ensure that sample second moments have a limiting normal distribution whose covariance matrix has a relatively simple form. Assumption (M3) permits random conditional $2p$th moments, but its boundedness requirement is somewhat restrictive.

When $p > 1$, the assumptions (M1)–(M3) will be shown to imply that upper bounds of *higher-than-second* moments of $\hat{r}_{xy}(T)$ can be obtained from *second* moments calculated under the simplifying assumption that the series $x(t)$ and $y(t)$ are *independent*. We first present such results for general quadratic forms.

Consider a constant-coefficient quadratic form in $x(t)$, $y(t)$, $t = 1, \cdots, T$,

$$Q \equiv \sum_{s,t=1}^{T} x(s)q(s,t)y(t).$$

If we set $\Gamma_{xy}(j) \equiv \mathcal{E}x(t)y(t-j)$, $j = 0, \pm 1, \ldots$, then the mean of Q is given by

$$\mathcal{E}Q = \sum_{s,t=1}^{T} q(s,t)\Gamma_{xy}(s-t). \tag{2.4}$$

Observe that the $2p$th moment norm of Q satisfies

$$\|Q\|_{2p} \le |\mathcal{E}Q| + \|Q - \mathcal{E}Q\|_{2p}. \tag{2.5}$$

Let $\tilde{Q}$ denote the variate obtained by replacing $x(t)$ and $y(t)$ in Q with series $\tilde{x}(t)$ and $\tilde{y}(t)$ which have the same spectral densities as these series but which are *independent* of one another. Our first theorem establishes the existence of a constant K_p such that

$$\|Q - \mathcal{E}Q\|_{2p} \le K_p \|\tilde{Q}\|_2. \tag{2.6}$$

Bounds for $\|Q\|_{2p}$ then follow from (2.4–6) and are given in the following theorem, whose proof can be found in the Appendix.

First Moment Bound Theorem. *Suppose the processes $\alpha(t)$ and $\beta(t)$ satisfy (M1) - (M3), with $p \ge 1$ in (2.3). Then there exists a constant K_p, depending only on $p, d(\alpha), d(\beta), \Sigma$ and C_p, such that (2.6) holds. Consequently, for all time series $x(t)$, $y(t)$ with representations (2.1), and for all choices of non-random coefficients $q(s,t)$, $1 \le s,\ t \le T$, there is a constant K_p such that*

$$\Big\| \sum_{s,t=1}^{T} q(s,t)\Big\{x(s)y(t) - \Gamma_{xy}(s-t)\Big\}\Big\|_{2p}$$
$$\le K_p \Big\{ \sum_{s,t,u,v=1}^{T} q(s,t)q(u,v)\Gamma_{xx}(s-u)\Gamma_{yy}(t-v)\Big\}^{1/2}, \tag{2.7}$$

and

$$\Big\| \sum_{s,t=1}^{T} q(s,t)x(s)y(t)\Big\|_{2p}$$
$$\le \Big| \sum_{s,t=1}^{T} q(s,t)\Gamma_{xy}(s-t)\Big| + K_p \Big\{ \sum_{s,t,u,v=1}^{T} q(s,t)q(u,v)\Gamma_{xx}(s-u)\Gamma_{yy}(t-v)\Big\}^{1/2}. \tag{2.8}$$

2.1. Useful special cases

As a first application of (2.7) and a step toward (1.10), we note that for

$$\tau_{xy}(T) \equiv T^{-1/2} \sum_{t=1}^{T} \Big\{x(t)y(t) - \Gamma_{xy}\Big\}, \tag{2.9}$$

it follows from (2.7) with $q(s,t) = T^{-1/2}\delta_{st}$ (Kronecker's delta) that

$$\|\tau_{xy}(T)\|_{2p} \le K_p \Big\{ T^{-1} \sum_{t,u=1}^{T} \Gamma_{xx}(t-u)\Gamma_{yy}(t-u)\Big\}^{1/2}$$

$$= K_p\left\{\sum_{k=-(T-1)}^{T-1}(1-|k|/T)\Gamma_{xx}(k)\Gamma_{yy}(k)\right\}^{1/2}$$
$$\le K_p\left\{\sum_{k=-\infty}^{\infty}\Gamma_{xx}^2(k)\cdot\sum_{k=-\infty}^{\infty}\Gamma_{yy}^2(k)\right\}^{1/4}$$
$$= (2\pi)^{1/2}K_p\left\{\int_{-\pi}^{\pi}f_{xx}^2(\lambda)d\lambda\cdot\int_{-\pi}^{\pi}f_{yy}^2(\lambda)d\lambda\right\}^{1/4}, \qquad (2.10)$$

by Cauchy-Schwarz and Parseval. This is (1.10) for the special case in which $Z_X(t)=0$, $Z_Y(t)=0$ for all t. It can be used to verify some of the uniform integrability conditions required for a rigorous derivation of the bias properties of Akaike's AIC criterion, see Findley (1985) and Findley and Wei (1993).

One can apply (2.10) with $p=1$ within a straightforward adaptation of the subsequence argument on pp. 184–185 of Hall and Heyde (1980) to show that the variate $T^{-1/2}\tau_{xy}(T)=T^{-1}\sum_{t=1}^{T}x(t)y(t)-\Gamma_{xy}$ converges to 0 with probability one. (Strong Law)

Our applications of (2.8) come by way of the following corollary, in which we use $\|c\|^2$ to denote the inner product $c'c$ of a vector with itself.

Corollary. *For each* $1\le T<\infty$, *let* $c(t;T)$ *and* $d(t;T)$, $1\le t\le T$ *be vector sequences such that*

$$\sum_{t=1}^{T}\|c(t;T)\|^2\le K_c,\quad \sum_{t=1}^{T}\|d(t;T)\|^2\le K_d \qquad (2.11)$$

hold, for some constants K_c *and* K_d. *Then under the assumptions of the theorem, the inequality*

$$\left\|T^{-1/2}\left[\sum_{s=1}^{T}c(s;T)x(s)\right]'\left[\sum_{t=1}^{T}d(t;T)y(t)\right]\right\|_{2p}$$
$$\le (2\pi K_cK_d)^{1/2}(K_p+1)\left\{\int_{-\pi}^{\pi}f_{xx}^2(\lambda)d\lambda\cdot\int_{-\pi}^{\pi}f_{yy}^2(\lambda)d\lambda\right\}^{1/4} \qquad (2.12)$$

is valid for all T, *where* K_p *is the constant occurring in* (2.7).

Alternatively, if the spectral density functions are essentially bounded,

$$f_{xx}(\lambda)\le M_x,\quad f_{yy}(\lambda)\le M_y\quad(\lambda\text{-a.e.}) \qquad (2.13)$$

then the multiplier $T^{-1/2}$ *in* (2.12) *can be dispensed with, in the sense that* (2.14) *holds:*

$$\left\|\left[\sum_{s=1}^{T}c(s;T)X(s)\right]'\left[\sum_{t=1}^{T}d(t;T)y(t)\right]\right\|_{2p}\le 2\pi(K_cK_dM_xM_y)^{1/2}(K_p+1). \qquad (2.14)$$

Proof. By Cauchy-Schwarz, the quantity on the left in (2.12) is bounded above by

$$\left\| T^{-1/2}\left\| \sum_{t=1}^{T} c(t;T)x(t)\right\|^2 \right\|_{2p}^{1/2} \cdot \left\| T^{-1/2}\left\| \sum_{t=1}^{T} d(t;T)y(t)\right\|^2 \right\|_{2p}^{1/2}.$$

By the symmetry of the roles of $x(t)$ and $y(t)$, we can establish (2.12) by verifying

$$\left\| T^{-1/2}\left\| \sum_{t=1}^{T} c(t;T)x(t)\right\|^2 \right\|_{2p} \leq (2\pi)^{1/2}K_c(K_p+1)\left\{ \int_{-\pi}^{\pi} f_{xx}^2(\lambda)d\lambda \right\}^{1/2}. \quad (2.15)$$

This follows by setting $y(s) = x(s)$ and $q(r,s) \equiv T^{-1/2}c(s;T)'c(t;T)$ in (2.8) and then observing, via Cauchy-Schwarz and Parseval, that

$$\begin{aligned} &\left| T^{-1/2} \sum_{s,t=1}^{T} c(s;T)'c(t;T)\Gamma_{xx}(s-t) \right| \\ \leq\ & \left[\sum_{t=1}^{T} \|c(t;T)\|^2 \right] \left\{ T^{-1} \sum_{s,t=1}^{T} \Gamma_{xx}^2(s-t) \right\}^{1/2} \\ \leq\ & (2\pi)^{1/2}K_c \left\{ \int_{-\pi}^{\pi} f_{xx}^2(\lambda)d\lambda \right\}^{1/2}. \end{aligned}$$

Setting $d \equiv \dim c(t,T)$, the final assertion (2.14) follows similarly with the aid of

$$\begin{aligned} &\left| \sum_{s,t=1}^{T} c(s;T)'c(t;T)\Gamma_{xx}(s-t) \right| \\ =\ & \int_{-\pi}^{\pi} \sum_{j=1}^{d} \left| \sum_{t=1}^{T} c_j(t;T)e^{it\lambda} \right|^2 f_{xx}(\lambda)d\lambda \\ \leq\ & 2\pi M_x \sum_{t=1}^{T} \|c(t;T)\|^2 \leq 2\pi K_c M_x. \end{aligned}$$

2.2. Proof of (1.10)

Conditions under which (1.10) is valid are given in the following theorem.

Theorem 2.1. *Under the conditions* (1.4), (1.9) *and* (M1)–(M3) *on the coordinate entries of the processes* $X(t)$ *and* $Y(t)$, *the bound* (1.10) *holds for all of the entries* $\hat{r}_{xy}(T)$ *of* $T^{-1/2}\sum_{t=1}^{T}(\hat{X}(t)\hat{Y}(t)' - \Gamma_{XY})$, *and for all* $1 \leq T < \infty$.

Proof. For an arbitrary vector sequence $Z(t)$ of dimension $d(Z)$, let the matrices $S(T)$ be such that

$$S(T)'S(T) \equiv \left[\sum_{t=1}^{T} Z(t)Z(t)' \right]^{+}$$

and define

$$c(t;T) \equiv S(T)Z(t), \quad (1 \le t \le T).$$

Then, with tr denoting trace, we have

$$\begin{aligned}\sum_{t=1}^{T} c(t;T)'c(t;T) &= \sum_{t=1}^{T} Z'(t)S(T)'S(T)Z(t) \\ &= \operatorname{tr}\left\{\left[\sum_{t=1}^{T} Z(t)Z(t)'\right]\left[\sum_{t=1}^{T} Z(t)Z(t)'\right]^{+}\right\} \le d(Z), \quad (2.16)\end{aligned}$$

so (2.11) is satisfied, with $K_c = d(Z)$. From (2.12), we conclude that the variates

$$\begin{aligned}R_{xy:Z}(T) &\equiv \left[\sum_{t=1}^{T} x(t)Z(t)'\right]\left[\sum_{t=1}^{T} Z(t)Z(t)'\right]^{+}\left[\sum_{t=1}^{T} Z(t)y(t)\right] \\ &= \left[\sum_{s=1}^{T} c(s;T)x(s)\right]'\left[\sum_{t=1}^{T} c(t;T)y(t)\right] \quad (2.17)\end{aligned}$$

satisfy

$$\begin{aligned}&\sup_{T\ge1}\left\|T^{-1/2}R_{xy:Z}(T)\right\|_{2p} \\ &\le (2\pi)^{1/2}d(Z)(K_p+1)\left\{\int_{-\pi}^{\pi} f_{xx}^2(\lambda)d\lambda \cdot \int_{-\pi}^{\pi} f_{yy}^2(\lambda)d\lambda\right\}^{1/4}. \quad (2.18)\end{aligned}$$

For later reference, we note that, if (2.13) holds, then (2.14) yields

$$\sup_{T\ge1}\|R_{xy:Z}(T)\|_{2p} \le 2\pi d(Z)(M_xM_y)^{1/2}(K_p+1). \quad (2.19)$$

Now (1.10) is within easy reach. Since

$$\left\|\hat{\tau}_{xy}(T)\right\|_{2p} \le \left\|\tau_{xy}(T)\right\|_{2p} + \left\|\hat{\tau}_{xy}(T) - \tau_{xy}(T)\right\|_{2p},$$

it follows from (2.10) and (2.18) that we have only to verify

$$\begin{aligned}&T^{1/2}\left\|\hat{\tau}_{xy}(T) - \tau_{xy}(T)\right\|_{2p} \\ &\le \left\|R_{xx:Z_X}(T)\right\|_{2p}^{1/2}\left\|R_{yy:Z_Y}(T)\right\|_{2p}^{1/2} + \left\|R_{xy:Z_X}(T)\right\|_{2p} + \left\|R_{xy:Z_Y}(T)\right\|_{2p}, \quad (2.20)\end{aligned}$$

the R-variates on the right being defined as in (2.17).

Using the identity

$$\begin{aligned}&\hat{x}(t)\hat{y}(t) - x(t)y(t) \\ &= (\hat{x}(t) - x(t))(\hat{y}(t) - y(t)) + x(t)(\hat{y}(t) - y(t)) + (\hat{x}(t) - x(t))y(t),\end{aligned}$$

we obtain via special cases of (1.13) that

$$
\begin{aligned}
&T^{1/2}\Big\{\hat{\tau}_{xy}(T)-\tau_{xy}(T)\Big\}\\
&=\sum_{t=1}^{T}\Big(A_x-\hat{A}_x(T)\Big)Z_X(t)Z_Y(t)'\Big(A_y-\hat{A}_y(T)\Big)'\\
&\qquad+\sum_{t=1}^{T}x(t)Z_Y(t)'\Big(A_y-\hat{A}_y(T)\Big)'+\Big(A_x-\hat{A}_x(T)\Big)\sum_{t=1}^{T}Z_X(t)y(t)\\
&=\sum_{t=1}^{T}\Big(A_x-\hat{A}_x(T)\Big)Z_X(t)Z_Y(t)'\Big(A_y-\hat{A}_y(T)\Big)'-R_{xy:Z_Y}(T)-R_{xy:Z_X}(T). \qquad (2.21)
\end{aligned}
$$

By Cauchy-Schwarz, first for vector inner products and then for moment norms, the $2p$-norm of the summation term in this last expression is bounded above by

$$\Big\|R_{xx:Z_X}^{1/2}(T)R_{yy:Z_Y}^{1/2}(T)\Big\|_{2p}\le\Big\|R_{xx:Z_X}(T)\Big\|_{2p}^{1/2}\cdot\Big\|R_{yy:Z_Y}(T)\Big\|_{2p}^{1/2}. \qquad (2.22)$$

(Note that, because A_x and $\hat{A}_x(T)$ are row vectors and $Z_X(t)$ is a column vector, $R_{xx:Z_X}$ is a nonnegative real number, and the same is true of $R_{yy:Z_Y}$.) Therefore (2.20) follows from (2.21) and (2.22). This completes the proof of (1.10).

3. Joint Representation and Central Limit Theorems

The use of differently designated processes $\alpha(t)$ and $\beta(t)$ in (2.1) was helpful for the formulation of the First Moment Bound Theorem, but for other purposes it is useful to note that there is no loss of generality in assuming that $\alpha(t)=\beta(t)$. More specifically, under (M1–3), $(\alpha(t)'\ \beta(t)')'$ is also a covariance stationary, F_t-adapted martingale difference sequence, and one can always find non-random matrices D_α, D_β and $D_{\alpha\beta}$ with the properties that the process $\delta(t)\equiv D_{\alpha\beta}(\alpha(t)'\ \beta(t)')'$ is such that $\alpha(t)=D_\alpha\delta(t)$, $\beta(t)=D_\beta\delta(t)$ and also such that, with probability one, $\delta(t)$ satisfies

$$E\Big\{\delta(t)|F_{t-1}\Big\}=0, \qquad (3.1)$$

and

$$E\Big\{\delta(t)\delta(t)'|F_{t-1}\Big\}=I_\delta, \qquad (3.2)$$

where I_δ denotes the identity matrix of order $d(\delta)\equiv\dim\delta(t)$. Clearly, with

$A(j) \equiv a(j)D_\alpha$ and $B(j) \equiv b(j)D_\beta$, we can write

$$x(t) = \sum_{j=-\infty}^{\infty} A(j)\delta(t-j)$$
$$y(t) = \sum_{j=-\infty}^{\infty} B(j)\delta(t-j). \tag{3.3}$$

Further, with probability one,

$$\begin{aligned} E\Big\{(\delta(t)'\delta(t))^{2p}|F_{t-1}\Big\} &\le \lambda_{\max}^{2p}\Big(D'_{\alpha\beta}D_{\alpha\beta}\Big)E\Big\{(\alpha'(t)\alpha(t)+\beta'(t)\beta(t))^{2p}|F_{t-1}\Big\} \\ &\le \Big(2\lambda_{\max}(D'_{\alpha\beta}D_{\alpha\beta})\Big)^{2p}C_p. \end{aligned}$$

In other words, there is a finite constant $C_{\delta,p}$ such that

$$E\Big\{(\delta(t)'\delta(t))^{2p}|F_{t-1}\Big\} \le C_{\delta,p} \tag{3.4}$$

holds for all t with probability one.

In the jointly Gaussian case, a representation with these properties exists if the (square integrable) spectral density matrix of the joint process $(X(t)\ Y(t))$ has the same (non-zero) rank at almost all frequencies. Then the representing process $\delta(t)$ can be chosen to have dimension equal to this rank (see Rozanov (1967, pp. 41–2)).

Suppose that $p > 1$ and that $\delta(t)$ satisfying (3.1–4) is also *fourth-order stationary.* In the following subsections, CLT's for the statistics under consideration will be obtained with the only additional requirement that the coordinate entries of $\delta(t)$ satisfy the following fourth-moment "mixing" condition: *There is a summable, positive sequence a_m such that*

$$\mathcal{E}|\mathcal{E}\Big\{\delta_i(t_1)\delta_j(t_2)\delta_k(t_3)\delta_\ell(t_4)|F_{t_1-m}\Big\} - \mathcal{E}\delta_i(t_1)\delta_j(t_2)\delta_k(t_3)\delta_\ell(t_4)| = O(a_m) \tag{3.5}$$

holds uniformly in t_1, where $t_1 \le t_2 \le t_3 \le t_4$, for $m = 1, 2, \ldots$.

3.1. CLT's for $\hat{\tau}_{xy}(T)$ and $\hat{\tau}_{\varepsilon\eta}(T)$

In this subsection, we shall explain how (1.10) can be applied to verify a Lindeberg condition which leads to a corrected and generalized version of the CLT Theorem 2.2 of Hosoya and Taniguchi (1982) for the serial covariance statistics $\tau_{xy}(T)$ under our assumptions on $\delta(t)$. Then CLT's for $\hat{\tau}_{xy}(T)$ and $\hat{\tau}_{\varepsilon\eta}(T)$ follow from conditions which imply, respectively,

$$\hat{\tau}_{xy}(T) - \tau_{xy}(T) \xrightarrow{\text{pr}} 0, \tag{3.6}$$

and

$$\hat{\tau}_{\varepsilon\eta}(T) - \tau_{\varepsilon\eta}(T) \xrightarrow{\text{pr}} 0, \tag{3.7}$$

conditions established below in Theorem 3.2. Hannan (1976) appears to have been the first to demonstrate that square integrability is the only condition required of the spectral density functions in order to obtain the classical Gaussian limiting distribution for $\tau_{xy}(T)$ (and $\hat{\tau}_{xy}(T)$, when $Z_X(T)$ and $Z_Y(T)$ are constant). Hosoya and Taniguchi (1982) avoided the strict stationarity and ergodicity assumptions used by Hannan and endeavored to obtain a CLT under weaker assumptions on the representing process. Theorem 3.1 below achieves this, apart from our use of $p > 1$ in (3.4) in place of Hannan's $p = 1$, although not in quite the manner Hosoya and Taniguchi envisioned. (Our moment bound results permit $p = 1$, but other conditions than those we use below are required to obtain a CLT for this case.) Our main use for Theorems 3.1 and 3.2 is the derivation in Section 6 of a test statistic for regressor selection, comparing two non-nested regressor processes $\hat{X}(t)$ without the requirement that one of the processes $X(t)$ be correct. However, we also show in subsection 3.2 that these theorems yield a CLT for the regression coefficient estimator $\hat{D}(T)$.

The variance matrices in the limiting distributions referred to in the following theorem are described by the usual formulas, (see Hannan (1976) for example), which will not be repeated here. It suffices for our purposes to mention that, in the case of the $\tau_{xy}(T)$ for example, the joint variance matrix is 2π times the spectral density matrix of the covariance stationary process $\text{vec}(X(t)Y(t)' - \Gamma_{xy})$ evaluated at the frequency $\lambda = 0$.

Theorem 3.1. *If the coordinate entries of $X(t)$ and $Y(t)$ have linear representations* (3.3) *in terms of a sequence $\delta(t)$ satisfying* (3.1–2), (3.4) *with $p > 1$, and* (3.5), *then the coordinate covariance statistics $\tau_{xy}(T)$ defined by* (2.9) *converge jointly in distribution as $T \to \infty$ to a zero-mean Gaussian variate, as do the analogous statistics $\tau_{\varepsilon\eta}(T)$ associated with $E(t)$. Under either of the additional assumptions* (i) *or* (ii) *of Theorem* 3.2 *below, the sequence $\hat{\tau}_{xy}(T)$ (respectively $\hat{\tau}_{\varepsilon\eta}(T)$) has the same limiting distribution as $\tau_{xy}(T)$ (respectively $\tau_{\varepsilon\eta}(T)$).*

Proof. The only part of the proof of Hannan (1976) that needs to be modified to yield this theorem is that which concerns the CLT for the special case of $\tau_{xy}(T)$ obtained when

$$x(t) \equiv \sum_{j=-M}^{M} A(j)\delta(t-j)$$

and

$$y(t) \equiv \sum_{j=-M}^{M} B(j)\delta(t-j)$$

for *fixed, finite* $M > 0$, which we now consider. To show joint asymptotic normality of the $\tau_{xy}(T)$'s, it suffices to establish the asymptotic normality of linear combinations $\sum_{x,y} c_{xy}\tau_{xy}(T)$. These can all be written as $T^{-1/2}\sum_{t=1}^{T} v(t+M)$, for $v(t)$ of the form

$$v(t) = \sum_{i,j=-M}^{M} \sum_{k,\ell=1}^{d(\delta)} c_{ijk\ell}\Big\{\delta_k(t-M-i)\delta_\ell(t-M-j) - \delta_{ij}\delta_{k\ell}\Big\}$$

where $\delta_{ij} = 1$ or 0, according to whether $i = j$ or $i \neq j$, etc. The variates $v(t)$ so defined satisfy the hypotheses of the CLT Lemma below, with (i) following from (1.10). This lemma yields the joint asymptotic normality of the $\tau_{xy}(T)$ in the special case $M < \infty$ being considered. Thus it completes the proof of the theorem.

We shall have occasion to use the largest-singular-value matrix-norm, $\|\Gamma\| \equiv \lambda_{\max}^{1/2}(\Gamma'\Gamma)$. If Γ is random, we define $\|\Gamma\|_p \equiv \{\mathcal{E}\|\Gamma\|^p\}^{1/p}$ for any $p > 0$.

CLT Lemma. *Suppose that $v(t)$ is a d-variate, covariance stationary, zero-mean time series whose autocovariance sequence $\Gamma(j) \equiv \mathcal{E}v(t)v(t-j)'$, $j = 0, \pm 1 \ldots$ is summable. Assume that for some $p > 1$,*
(i) $\sup_{t_0,T} \|T^{-1/2}\sum_{t=1}^{T} v(t_0+t)\|_{2p} < \infty$ *holds.*
Assume further that there is an increasing sequence of σ-fields $\{F_t\}$ such that $v(t)$ is F_t-measurable, and
(ii) $\sup_t \mathcal{E}\|\mathcal{E}\{v(t+k)|F_t\}\| = O(k^{-1/2-\varepsilon})$ *for all $k \geq 1$, for some $\varepsilon > 0$.*
(iii) *For all $1 \leq j \leq k$,*

$$\mathcal{E}\Big\|\mathcal{E}\{v(t+j)v(t+k)'|F_t\} - \mathcal{E}v(t+j)v(t+k)'\Big\| = O(a_j),$$

where a_j, $j > 1$ is a summable, positive sequence.

Then the sequence $\xi_T \equiv T^{-1/2}\sum_{t=1}^{T} v(t)$ is asymptotically normal, with mean zero and covariance matrix $\sum_{j=-\infty}^{\infty}\Gamma(j)$.

Proof. By the usual argument, it suffices to consider the case $d = 1$. We establish the asserted result by verifying the hypotheses of Theorem 4.1 of Serfling (1968). Set $S(t,T) \equiv \sum_{k=1}^{T} v(t+k)$. Under our assumptions, we only have to show (3.8) and (3.9):

$$\sup_t \mathcal{E}|\mathcal{E}\{S(t,T)|F_t\}| = O(T^{1/2-\delta}) \text{ for some } \delta > 0, \tag{3.8}$$

$$\sup_t \mathcal{E}|\mathcal{E}\{S^2(t,T)|F_t\} - \mathcal{E}S^2(t,T)| = o(T). \tag{3.9}$$

By (ii), there is a constant $K > 0$ such that

$$\mathcal{E}|\mathcal{E}\{S(t,T)|F_t\}| \;\leq\; \sum_{k=1}^{T}\mathcal{E}|\mathcal{E}\{v(t+k)|F_t\}|$$

$$\leq K\sum_{k=1}^{T} k^{-1/2-\varepsilon} = O(T^{1/2-\varepsilon}).$$

This proves (3.8). For (3.9), let $m = m(T) \leq T$ be a sequence such that $m(T) \to \infty$ and $m(T) = o(T)$. We consider individually the terms of the decomposition of $S^2(t,T)$ obtained by squaring the identity $S(t,T) = S(t,m) + S(t+m, T-m)$,

$$S^2(t,T) = S^2(t,m) + 2S(t,m)S(t+m,T-m) + S^2(t+m,T-m). \quad (3.10)$$

First, note that,

$$\sup_t \mathcal{E}(\mathcal{E}\{S^2(t,m)|F_t\}) = \sup_t \mathcal{E}S^2(t,m)$$
$$= m \sum_{j=-(m-1)}^{m-1} (1-|j|/m)\Gamma(j) = O(m) = o(T). \quad (3.11)$$

Next, using Cauchy-Schwarz, observe that

$$\mathcal{E}|\mathcal{E}\{S(t,m)S(t+m,T-m)|F_t\}| + \mathcal{E}|S(t,m)S(t+m,T-m)|$$
$$\leq 2\{\mathcal{E}S^2(t,m)\}^{1/2}\{\mathcal{E}S^2(t+m,T-m)\}^{1/2} = O(m^{1/2})O(T^{1/2}) = o(T). \quad (3.12)$$

Because of (3.10–12), it remains only to verify (3.9) for $S(t+m,T-m)$. For this variate, there exists, by (iii), a constant C such that the left-hand side of (3.9) is bounded above by

$$\sup_t \Bigg[\sum_{j=m+1}^{T} \mathcal{E}|\mathcal{E}\{v^2(t+j)|F_t\} - \mathcal{E}v^2(t)|$$
$$+ 2\sum_{k=m+2}^{T}\sum_{j=m+1}^{k-1} \mathcal{E}|\mathcal{E}\{v(t+j)v(t+k)|F_t\} - \mathcal{E}v(t+j)v(t+k)| \Bigg]$$
$$\leq C\sum_{j=m+1}^{T} a_j + 2C\sum_{k=m+2}^{T}\sum_{j=m+1}^{k-1} a_j$$
$$\leq (2T+1)\sum_{j=m+1}^{\infty} a_j = o(T).$$

This completes the proof.

Remark 1. The above Lemma is a corrected and generalized version of Theorem 2.1 of Hosoya and Taniguchi (1982), having the additional hypothesis (i). Our hypotheses (ii) and (iii) are weaker than their hypotheses
(ii)′ $\sup_t \|\mathcal{E}\{v(t+k)|F_t\}\|_2 = O(k^{-1-\varepsilon})$ for all $k \geq 1$, for some $\varepsilon > 0$,

and
(iii)′ Set $a_j = j^{-1-\varepsilon}$ in (iii).
In fact, if (iii) in the Lemma is replaced by (iii)′, then (i) can be replaced by a weaker assumption
(i)′ $\sup_t \|v(t)\|_{2p} < \infty$.
The proof uses Theorem 5.1 of Serfling (1968) and the same arguments we used above.

Remark 2. To see that an additional assumption such as (i) or (i)′ is required, consider the sequence of independent variates $w(t)$ in which the tth variate takes only the values 0 and $\pm(|t|+1)^{1/2}$, each of the latter pair occurring with probability $(2|t|+2)^{-1}$. These variates have mean zero and variance one, and they fulfill the hypotheses of Hosoya and Taniguchi's Theorem 2.1, but since

$$T^{-1}\sum_{t=1}^{T}\mathcal{E}\left\{w(t)^2 I[2^{1/2}|w(t)| > T^{1/2}]\right\} \to 1/2,$$

the Lindeberg condition is not satisfied. (We use $I[G]$ to denote the indicator function of the event G.) Hence $T^{-1/2}\sum_{t=1}^{T} w(t)$ does not have a Gaussian limiting distribution (see Loève (1977, p.292)), which contradicts the assertion of their theorem. (In their proof, Inequality (6.18) is erroneous.)

Theorem 3.1 above can play the role of Hosoya and Taniguchi's incorrect Theorem 2.2 in the proofs of the later CLT's for general time series models in their paper. That is, these CLT's are justified if their hypotheses concerning the representing white noise process are modified to include the assumptions of our Theorem 3.1.

Now we turn to (3.6) and (3.7).

Theorem 3.2. *Suppose that $X(t)$ and $Y(t)$ satisfy the assumptions of Theorem* 3.1 *with the possible exception of* (3.5). *Then* (3.6) *and* (3.7) *hold if either of the conditions below is satisfied:*
(i) *The mean function regressors $Z_X(t)$ and $Z_Y(t)$ in* (1.1) *are constant.*
(ii) *All coordinate-entry spectral densities $f_{xx}(\lambda)$ and $f_{yy}(\lambda)$ are essentially bounded* (*condition* (2.13)).

Proof. Hannan (1976, p.397) proves that (3.6) follows from (i), and (2.19–20) show that it is also a consequence of (ii). The assertion (3.7) is a consequence of (3.6) and the result (4.5) established below.

3.2. CLT's for estimators of D

The preceding theorems yield CLT's for $\hat{D}(T)$. Consider first the simpler mean-zero-case estimator, $D(T) \equiv \sum_{t=1}^{T} Y(t)X(t)'(\sum_{t=1}^{T} X(t)X(t)')^{+}$. Observe that

$$T^{1/2}(D(T)-D)=\Big\{T^{-1/2}\sum_{t=1}^{T}E(t)X(t)'\Big\}\Big\{T^{-1}\sum_{t=1}^{T}X(t)X(t)'\Big\}^{+}$$
$$-T^{1/2}D\left[I_{d(X)}-\Big(\sum_{t=1}^{T}X(t)X(t)'\Big)\Big(\sum_{t=1}^{T}X(t)X(t)'\Big)^{+}\right]. \quad (3.13)$$

Let $\Delta(T)$ denote the matrix in square brackets in this last expression. This is an idempotent, positive semidefinite matrix, and we now show that it is zero with probability tending to 1,

$$\lim_{T\to\infty} P(\{\Delta(T)=0\})=1. \quad (3.14)$$

Indeed, since $\mathrm{tr}\Delta(T)=\mathrm{rank}(\Delta(T))$ takes on only integer values in $[0,d(X)]$, (3.14) will follow from $\Delta(T)\xrightarrow{\text{pr}}0$, which is a consequence of

$$\operatorname*{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T}X(t)X(t)'=\Gamma_{XX}, \quad (3.15)$$

the latter being an implication of (1.10). From (3.14) we obtain the negligibility of the final term in (3.13),

$$T^{1/2}D\Delta(T)\xrightarrow{\text{pr}}0. \quad (3.16)$$

For the other term, under the assumptions of Theorem 3.1, we have

$$T^{-1/2}\mathrm{vec}\Big(\sum_{t=1}^{T}E(t)X(t)'\Big)\xrightarrow{\text{dist}}\mathcal{N}(0,V), \quad (3.17)$$

for some V. Using the identity $\mathrm{vec}(AB)=(B\otimes I)\,\mathrm{vec}A$ (where I is the identity matrix whose order is the row dimension of A), we conclude from (3.13) and (3.16–17) that

$$T^{1/2}\mathrm{vec}(D(T)-D)\xrightarrow{\text{dist}}\mathcal{N}\Big(0,\big(\Gamma_{XX}^{-1}\otimes I_{d(X)}\big)V\big(\Gamma_{XX}^{-1}\otimes I_{d(X)}\big)\Big). \quad (3.18)$$

The argument that $T^{1/2}(\hat{D}(T)-D)$ has the same limiting distribution is completely analogous, with (3.17) replaced by

$$T^{-1/2}\mathrm{vec}\Big(\sum_{t=1}^{T}\bar{E}(t)\hat{X}(t)'\Big)\xrightarrow{\text{dist}}\mathcal{N}(0,V), \quad (3.19)$$

where $\bar{E}(t) \equiv \hat{Y}(t) - D\hat{X}(t)$. To verify (3.19), one uses the fact that under (i) or (ii) of Theorem 3.2,

$$T^{-1/2}\Big\{\sum_{t=1}^{T}\bar{E}(t)\hat{X}(t)' - \sum_{t=1}^{T}E(t)X(t)'\Big\} \xrightarrow{\text{pr}} 0.$$

In summary, we have

Theorem 3.3. *Under the assumptions of Theorem* 3.1, *the limiting result* (3.18) *holds for* $T^{1/2}(D(T) - D)$. *If, further, either* (i) *or* (ii) *of Theorem* 3.2 *is valid, then* $T^{-1/2}(\hat{D}(T) - D)$ *has the same limiting distribution.*

4. Second Moment Bound Theorem and the Proof of (1.11)

Our other general theorem concerns expressions like the final one in (1.13). Its proof is given in the Appendix.

Second Moment Bound Theorem. *For each* $T \geq 1$, *let* $u(t) = u(t;T)$ *and* $v(t) = v(t;T)$ *be scalar variates, and let* $w(t) = w(t;T)$ *be vector variates, for* $1 \leq t \leq T$, *such that a finite constant* M, *not depending on* T, *and a nonsingular matrix* Γ_{ww} *exist for which the following conditions are satisfied for some* $p \geq 1$:

$$\Big\|T^{-1}\sum_{t=1}^{T}u(t)^2\Big\|_{2p} + \Big\|T^{-1}\sum_{t=1}^{T}v(t)^2\Big\|_{2p} \leq M \tag{4.1}$$

$$\Big\|T^{-1}\sum_{s,t=1}^{T}v(s)v(t)w'(s)w(t)\Big\|_{p} + \Big\|T^{-1}\sum_{s,t=1}^{T}u(s)u(t)w'(s)w(t)\Big\|_{p} \leq M \tag{4.2}$$

$$\Big\|T^{-1/2}\sum_{t=1}^{T}\Big(w(t)w(t)' - \Gamma_{ww}\Big)\Big\|_{2p} \leq M. \tag{4.3}$$

Then there is a finite constant $\bar{M}_p$, *not depending on* T, *such that* (4.4) *holds for all* T, *and for every* ρ *satisfying* $p/2 \leq \rho \leq p$:

$$T^{\frac{1}{2}(p/\rho-2)}\Big\|\Big[\sum_{t=1}^{T}u(t)w(t)'\Big]\Big[\sum_{t=1}^{T}w(t)w(t)'\Big]^{+}\Big[\sum_{t=1}^{T}w(t)v(t)\Big]\Big\|_{\rho} \leq \bar{M}_p. \tag{4.4}$$

This result, together with (1.10), enables us to prove a theorem affirming (1.11).

Theorem 4.1. *Under the conditions* (1.4), (1.9) *and* (M1)–(M3) *on the coordinate entries of the processes* $X(t)$ *and* $Y(t)$, *the assertion* (1.11) *holds for the entries* $\hat{r}_{\epsilon\eta}(T)$ *of* $T^{-1/2}\sum_{t=1}^{T}(\hat{E}(t)\hat{E}(t)' - \Gamma_{EE})$.

Proof. Set $\bar{E}(t) \equiv \hat{Y}(t) - D\hat{X}(t)$, let $\bar{\varepsilon}(t)$ and $\bar{\eta}(t)$ be the entries corresponding to $\hat{\varepsilon}(t)$ and $\hat{\eta}(t)$, and define $\bar{\tau}_{\varepsilon\eta}(T) \equiv T^{-1/2}\sum_{t=1}^{T}(\bar{\varepsilon}(t)\bar{\eta}(t) - \Gamma_{\varepsilon\eta})$. We have

$$\hat{\tau}_{\varepsilon\eta}(T) = \bar{\tau}_{\varepsilon\eta}(T) + \left\{\hat{\tau}_{\varepsilon\eta}(T) - \bar{\tau}_{\varepsilon\eta}(T)\right\},$$

and it follows readily from (1.10) that

$$\sup_{T\geq 1} \|\bar{\tau}_{\varepsilon\eta}(T)\|_q < \infty,$$

for all $q \leq 2p$. Thus, to prove the boundedness of $\|\hat{\tau}_{\varepsilon\eta}(T)\|_p$, it suffices to demonstrate (4.5) below, a result which also establishes (3.7) as a consequence of (3.6) applied to $\bar{\tau}_{\varepsilon\eta}(T) - \tau_{\varepsilon\eta}(T)$.

$$\sup_{T\geq 1, p/2\leq \rho\leq p} T^{\frac{1}{2}(p/\rho-1)}\left\|\hat{\tau}_{\varepsilon\eta}(T) - \bar{\tau}_{\varepsilon\eta}(T)\right\|_\rho < \infty. \tag{4.5}$$

To verify (4.5), observe via (1.13) that

$$\hat{\tau}_{\varepsilon\eta}(T) - \bar{\tau}_{\varepsilon\eta}(T) = T^{-1/2}\left[\sum_{t=1}^{T}\bar{\varepsilon}(t)\hat{X}(t)'\right]\left[\sum_{t=1}^{T}\hat{X}(t)\hat{X}(t)'\right]^{+}\left[\sum_{t=1}^{T}\hat{X}(t)\bar{\eta}(t)\right]. \tag{4.6}$$

Therefore, by the Second Moment Bound Theorem and the interchangeability of $\bar{\varepsilon}(t)$ and $\bar{\eta}(t)$, (4.5) will follow from (4.7–9):

$$\sup_{T\geq 1}\left\|T^{-1}\sum_{t=1}^{T}\bar{\varepsilon}(t)^2\right\|_{2p} < \infty, \tag{4.7}$$

$$\sup_{T\geq 1}\left\|T^{-1}\sum_{s,t=1}^{T}\bar{\varepsilon}(s)\bar{\varepsilon}(t)\hat{X}'(t)\hat{X}(s)\right\|_{p} < \infty, \tag{4.8}$$

and

$$\sup_{T\geq 1}\left\|T^{-1/2}\sum_{t=1}^{T}\left(\hat{X}(t)\hat{X}(t)' - \Gamma_{XX}\right)\right\|_{2p} < \infty. \tag{4.9}$$

The condition (4.7) is verified by noting that the bracketed expression in

$$T^{-1}\sum_{t=1}^{T}\bar{E}'(t)\bar{E}(t) = T^{-1/2}\left\{T^{-1/2}\sum_{t=1}^{T}\left(\bar{E}'(t)\bar{E}(t) - \mathrm{tr}\Gamma_{EE}\right)\right\} + \mathrm{tr}\Gamma_{EE}$$

has bounded $2p$th moments, by (1.10). Next, since $\mathcal{E}E(t)X(t)' = 0$, the quantity under consideration in (4.8) coincides with the sum of $d(X) \equiv \dim X(t)$ products of pairs of variates of the form $\hat{\tau}_{yx_j} - \sum_{i=1}^{d(X)} D_{ij}\hat{\tau}_{x_i x_j}$. Finally, (4.9) concerns the $\hat{\tau}_{xx}(T)$. Thus (4.7–9) all follow from the appropriate instances of (1.10).

5. Filtering and Forecasting

5.1. General results

One of the attractive properties of the linear representation assumptions (3.1–4) is that filtered versions of $x(t)$ and $y(t)$,

$$x^{\dagger}(t) = \sum_{j=-\infty}^{\infty} g_x(j)x(t-j), \quad y^{\dagger}(t) = \sum_{j=-\infty}^{\infty} g_y(j)y(t-j)$$

also satisfy these assumptions. The spectral density $f^{\dagger}_{xx}(\lambda)$ of $x^{\dagger}(t)$ is obtained from the frequency response function $G_x(e^{i\lambda}) \equiv \sum_{j=-\infty}^{\infty} g_x(j)e^{ij\lambda}$ by mean of $f^{\dagger}_{xx}(\lambda) = |G_x(e^{i\lambda})|^2 f_{xx}(\lambda)$. Hence $f^{\dagger}_{xx}(\lambda)$ will be square integrable under

$$\int_{-\pi}^{\pi} \left|G_x(e^{i\lambda})\right|^4 d\lambda < \infty \tag{5.1}$$

because of (1.9), and will be essentially bounded under (2.12) if $G_x(e^{i\lambda})$ is essentially bounded,

$$\left|G_x(e^{i\lambda})\right| \le M < \infty \quad (\lambda - \text{a.e.}). \tag{5.2}$$

Similar remarks apply concerning the spectral density $f^{\dagger}_{yy}(\lambda)$ of $y^{\dagger}(t)$. These observations lead to filtered-series analogues of the results established earlier for $\tau_{xy}(T)$, $\hat{\tau}_{xy}(T)$ and $\hat{\tau}_{\varepsilon\eta}(T)$, provided that, in the estimated means case, the series defining the filtered mean functions

$$Z^{\dagger}_X(t) \equiv \sum_{j=-\infty}^{\infty} g_x(j)Z_X(t-j), \quad Z^{\dagger}_Y(t) \equiv \sum_{j=-\infty}^{\infty} g_y(j)Z_Y(t-j)$$

are defined for all t. (Note that these are constant in t if $Z_X(t)$ and $Z_Y(t)$ are constant, assuming that $\sum_{j=-\infty}^{\infty} g_x(j)$ and $\sum_{j=-\infty}^{\infty} g_y(j)$ are convergent.)

5.2. Uniform convergence of moment estimates of multistep-ahead forecast error

We now present an application of the preceding observations which makes explicit use of the bound in (1.10) and how it is transformed by a family of filters having the same essential bound in (5.2) for their squared gain functions. Our application is concerned with m-step-ahead forecasting ($m \ge 1$) of a (for simplicity) univariate, mean zero series $y(t)$ for which (3.1)–(3.4) hold. We consider forecasts obtained from not necessarily correct, invertible ARMA(q, r) models for $y(t)$, with $q \le q_0$ and $r \le r_0$, whose autoregressive and moving average polynomials, $\phi_q(z)$ and $\theta_r(z)$, with $\phi_q(0) = \theta_r(0) = 1$, define innovation functions

$\Psi(z) = \theta_r(z)/\phi_q(z)$ whose zeros and poles belong to $\{|z| \geq 1 + \nu_0\}$ for some $\nu_0 > 0$. It is easy to see that, for fixed q_0, r_0, ν_0, every sequence $\Psi_N(z)$ of such innovation functions has a subsequence which converges uniformly on each of the disks $\{|z| \leq 1 + \nu\}$, $0 \leq \nu < \nu_0$. It follows from this that for each $0 \leq \nu < \nu_0$, there exists a positive constant M_ν such that

$$M_\nu^{-1} \leq |\Psi(z)| \leq M_\nu \quad (|z| \leq 1 + \nu) \tag{5.3}$$

holds for all such innovation functions. It is well known (see Theorem 7.3 of Rozanov (1967)) that if a model has innovation function

$$\Psi(z) = 1 + \sum_{j=1}^{\infty} \psi(j) z^j, \tag{5.4}$$

then its m-step-ahead prediction filter for estimating $y(t)$ from $y(s)$, $-\infty < s \leq t-m$, has the frequency response function $(\sum_{j=m}^{\infty} \psi(j) e^{ij\lambda})/\Psi(e^{i\lambda})$. Thus the *m-step-ahead prediction error filter* has the frequency response function $G[\Psi; m](e^{i\lambda})$ given by

$$G[\Psi; m](z) \equiv \sum_{j=0}^{m-1} \psi(j) z^j / \Psi(z) \quad (\psi(0) = 1). \tag{5.5}$$

Let $F(M_\nu)$ denote the set of functions of the form (5.4) which are analytic in $\{|z| \leq 1+\nu\}$ and satisfy (5.3). It follows from Cauchy's inequality (see Titchmarsh (1939, p.84)) that the coefficients $\psi(j)$ satisfy

$$\sup_{\Psi \in F(M_\nu)} |\psi(j)| \leq M_\nu (1+\nu)^{-j}. \tag{5.6}$$

Since $\sum_{j=0}^{\infty}(1+\nu)^{-j} = \nu^{-1}(1+\nu)$, it is a consequence of (5.3) and (5.6) that the squared gain functions $|G[\Psi; m](e^{i\lambda})|^2$ have the bound $M(\nu) \equiv M_\nu^4 \nu^{-2}(1+\nu)^2$,

$$\sup_{\Psi \in F(M_\nu)} \left| G[\Psi; m](e^{i\lambda}) \right|^2 \leq M(\nu).$$

Letting B denote the backshift operator, we conclude that the spectral density functions $f[\Psi; m](\lambda)$ of the m-step-ahead prediction error process associated with $\Psi \in F(M_\nu)$,

$$e[\Psi](t|t-m) \equiv G[\Psi; m](B) y(t),$$

are dominated by the spectral density $f_{yy}(\lambda)$ of $y(t)$,

$$f[\Psi; m](\lambda) = \left| G[\Psi; m](e^{i\lambda}) \right|^2 f_{yy}(\lambda) \leq M(\nu) f_{yy}(\lambda). \tag{5.7}$$

Setting $\sigma^2[\Psi;m] \equiv \mathcal{E}\{e^2[\Psi](t|t-m)\}$, we conclude from (5.7) and (1.10) that there is a constant $K(\nu)$ such that

$$\sup_{T>m,\Psi\in F(M_\nu)} (T-m)^{1/2}\left\| \frac{1}{T-m}\sum_{t=m+1}^{T} e^2[\Psi](t|t-m) - \sigma^2[\Psi;m]\right\|_{2p}$$
$$\leq K(\nu)\left\{\int_{-\pi}^{\pi} f_{yy}^2(\lambda)d\lambda\right\}^{1/2} \tag{5.8}$$

This is a uniform convergence result, but one that is often not of direct interest, because the errors $e[\Psi](t|t-m)$ are unobservable when they are functions of infinitely many $y(s)$, $s \leq t-m$. For $t \geq m+1$, set $\bar{e}[\Psi](t|t-m) \equiv y(t) - \mathcal{E}_\Psi\{y(t)|y(t-m),\ldots,y(1)\}$, where $\mathcal{E}_\Psi\{\cdot|\cdot\}$ denotes conditional expectation for a zero-mean *Gaussian* process whose spectral density is proportional to $|\Psi(e^{i\lambda})|^2$. Using results and arguments from Findley (1991b), it can be shown that there is a constant $K(\nu,m)$ such that

$$\sup_{T>m,\Psi\in F(M_\nu)} (T-m)^{-1/2}\left\| \sum_{t=m+1}^{T}\left\{\bar{e}^2[\Psi](t|t-m) - e^2[\Psi](t|t-m)\right\}\right\|_{2p} \leq K(\nu,m). \tag{5.9}$$

Combining (5.8) and (5.9), we obtain the *uniform convergence in 2p-norm* of the *observable* moment estimator $(T-m)^{-1}\sum_{t=m+1}^{T}\bar{e}^2[\Psi](t|t-m)$ to the mean square forecast error $\sigma^2[\Psi;m]$. This fact is used in Findley (1991a) to motivate a forecast-model selection procedure whose choices depend on the lead m. Rather than describe this procedure here, we describe a similar procedure for regression model selection in the next Section.

6. Two Regressor Selection Procedures

Suppose that there are two competing, possibly non-nested and incorrect, regressor processes, $X_1^*(t)$ and $X_2^*(t)$, for $Y^*(t)$, whose mean functions can be expressed as a linear function of the vector functions $Z_1(t)$ and $Z_2(t)$ respectively,

$$\mathcal{E}X_i^*(t) = A_iZ_i(t) \quad (i=1,2).$$

Suppose also that the $X_i(t) \equiv X_i^*(t) - A_iZ_i(t)$ have square summable autocovariance matrices $\Gamma_i^X(j)$, $j = 0,\pm1,\ldots$, with $\Gamma_i^X(0)$ nonsingular (a different notation from Section 1), and, together with $Y(t)$ (defined as in Section 1), have linear representations in terms of a process $\delta(t)$ satisfying (3.1) and (3.3–4) with $p \geq 1$.

For $T \geq 1$ and $1 \leq t \leq T$ and $i = 1,2$, define $D_i \equiv \mathcal{E}Y(t)X_i(t)'\Gamma_i^X(0)^{-1}$, $E_i(t) \equiv Y(t) - D_iX_i(t)$, $\Gamma_i^E \equiv \mathcal{E}E_i(t)E_i(t)'$ and, further,

$$\hat{A}_i(T) \equiv \sum_{t=1}^{T} X_i^*(t)Z_i(t)'\left[\sum_{t=1}^{T} Z_i(t)Z_i(t)'\right]^+,$$

$$\hat{X}_i(t) \equiv X_i^*(t) - \hat{A}_i(T)Z_i(t),$$

$$\hat{D}_i(T) \equiv \sum_{t=1}^{T} \hat{Y}(t)\hat{X}_i(t)\left[\sum_{t=1}^{T} \hat{X}_i(t)\hat{X}_i(t)'\right]^+,$$

$$\hat{E}_i(t) \equiv \hat{Y}(t) - \hat{D}_i(T)\hat{X}_i(t),$$

and, finally,

$$S^{(1,2)}(T) \equiv \sum_{t=1}^{T} \left\{\hat{E}_1(t)'\hat{E}_1(t) - \hat{E}_2(t)'\hat{E}_2(t)\right\}.$$

It follows from (1.12) that

$$\operatorname*{plim}_{T\to\infty} T^{-1}S^{(1,2)}(T) = \mathrm{tr}\Gamma_1^E - \mathrm{tr}\Gamma_2^E. \tag{6.1}$$

We will use $\mathrm{tr}\Gamma_i^E$ as a *measure of misfit* for the regressor process $X_i^*(t)$. In the case in which the regressors $X_1^*(t)$ and $X_2^*(t)$ do not have the same (mis)fit, that is, when

$$\mathrm{tr}\Gamma_1^E \neq \mathrm{tr}\Gamma_2^E, \tag{6.2}$$

then we wish to determine which regressor process has the better fit (smaller misfit). The result (6.1) suggests that this can be accomplished, when T is large enough, by plotting

$$S^{(1,2)}(\bar{T}), \quad T/2 \leq \bar{T} \leq T \tag{6.3}$$

against $\bar{T}$, and looking for the (from (6.1)) expected approximately linear movement of $S^{(1,2)}(\bar{T})$, which, if it has positive slope, reveals that $\mathrm{tr}\Gamma_1^E > \mathrm{tr}\Gamma_2^E$, and hence that $\hat{X}_2(t)$ is the preferred regressor, etc. This procedure is a regression analogue of the likelihood-ratio-based graphical model selection procedure of Findley (1990). It includes the autoregressive case of the mean square m-step ahead forecast-error-based procedure of Findley (1991a), because the regressors $X_i(t)$ and $Z_i(t)$ could consist of values measured at, respectively designed for, an earlier time $t - m$, in which case the $\hat{E}_i(t)$ are m-step ahead forecast errors. The reader is referred to these references for examples of graphs corresponding to (6.3) from competing models. The theoretical results discussed in these references in support of their graphical procedures do not cover the case of models with *estimated mean functions*, even though all of the examples given utilize mean functions initialized by OLS estimates. Thus (6.1) helps to fill a gap in this theory.

We will strengthen our assumptions to derive a hypothesis testing procedure for testing

$$\mathrm{tr}\Gamma_1^E = \mathrm{tr}\Gamma_2^E \tag{6.4}$$

against the various alternatives associated with (6.2).

6.2. A hypothesis test for (6.4)

We assume now that $Y(t)$, $X_1(t)$ and $X_2(t)$ all have representations in terms of a fourth-order-stationary martingale difference sequence satisfying (3.1–2), (3.4) with $p > 1$, and (3.5). Then the same is true of the error processes $E_1(t)$ and $E_2(t)$, and the related processes $E_1(t) + E_2(t)$ and $E_1(t) - E_2(t)$ (see Section 5). Under (6.4), the process

$$\delta^{(1,2)}(t) \equiv E_1(t)'E_1(t) - E_2(t)'E_2(t) = (E_1(t) + E_2(t))'(E_1(t) - E_2(t))$$

has mean zero. In any case, it follows from Theorem 3.1 that

$$\tau^{(1,2)}(T) \equiv T^{-1/2} \sum_{t=1}^{T} \left\{\delta^{(1,2)}(t) - (\mathrm{tr}\Gamma_1^E - \mathrm{tr}\Gamma_2^E)\right\}$$

has a limiting normal distribution with mean 0 and with variance $v^{(1,2)}$ given by 2π times the value at frequency $\lambda = 0$ of the spectral density of the process $\delta^{(1,2)}(t) - (\mathrm{tr}\Gamma_1^E - \mathrm{tr}\Gamma_2^E)$. If we define

$$\hat{\tau}^{(1,2)}(T) \equiv T^{-1/2} \sum_{t=1}^{T} \left\{\hat{E}_1(t)'\hat{E}_1(t) - \hat{E}_2(t)'\hat{E}_2(t)\right\},$$

then, under (6.4) and the hypotheses of Theorem 3.2, we have $\hat{\tau}^{(1,2)}(T) - \tau^{(1,2)}(T) \to 0$. Thus, if $\hat{v}^{(1,2)}(T)$ is any weakly consistent estimator of $v^{(1,2)}$, then when $v^{(1,2)} \neq 0$, we have

$$\hat{v}^{(1,2)}(T)^{-1/2}\hat{\tau}^{(1,2)}(T) \xrightarrow{\text{dist}} \mathcal{N}(0,1) \tag{6.5}$$

under the null hypotheses (6.4). In general,

$$\hat{v}^{(1,2)}(T)^{-1/2}\hat{\tau}^{(1,2)}(T) = T^{1/2}\hat{v}^{(1,2)}(T)^{-1/2}\left\{\mathrm{tr}\Gamma_1^E - \mathrm{tr}\Gamma_2^E\right\} + O_p(1).$$

Hence, under (6.2) (and assuming $v^{(1,2)} \neq 0$) the sign of this statistic will ultimately be that of $\mathrm{tr}\Gamma_1^E - \mathrm{tr}\Gamma_2^E$, and its magnitude will be larger than is plausible under (6.5). In this way it will reveal the better-fitting model. The construction of a provably consistent estimator $\hat{v}^{(1,2)}(T)$ is a topic for further research. However, the examples presented for an analogous hypothesis testing procedure in Findley (1990) suggest that a familiar robust spectrum estimator applied to $\hat{\delta}^{(1,2)}(T) \equiv \hat{E}_1(t)'\hat{E}_1(t) - \hat{E}_2'(t)\hat{E}_2(t)$ usually yields reliable results. The hypothesis testing procedure based on (6.5) is motivated by the somewhat analogous procedure of Vuong (1989), who considered the case of nonlinear regression when the variates $(Y(t)'\ X_1(t)'\ X_2(t)')'$, $t = 1, 2, \ldots$ are independent and identically distributed.

Remark. It $X_1(t)$ is nested in $X_2(t)$, then $\Gamma_2^E \leq \Gamma_1^E$. Therefore, under (6.4), we must have $\Gamma_2^E = \Gamma_1^E$, hence also $E_1(t) = E_2(t)$ (a.s.), so that $\delta^{(1,2)}(t) = 0$ (a.s.), and $v^{(1,2)} = 0$. Thus the test based on (6.5) is not available for nested comparisons.

Acknowledgement

Much of the research reported here was done while the first-named author was a Visiting Research Fellow at Academia Sinica in Taipei. He wishes to express his gratitude to the Institute of Statistical Science and its Fellows and staff for support and hospitality during this visit. Both authors wish to thank the referee for reading the paper very carefully and making suggestions.

Appendix

A.1. Proof of Moment Bound Theorem 1

Let $\alpha_i(t)$, $1 \leq i \leq d(\alpha)$ and $\beta_j(t)$, $1 \leq j \leq d(\beta)$ denote the entries of $\alpha(t)$ and $\beta(t)$. The mean-centered quadratic form $Q - \mathcal{E}Q$ is a variate of the form

$$s = \sum_{i=1}^{d(\alpha)} \sum_{j=1}^{d(\beta)} s_{ij}, \tag{A.1}$$

in which each s_{ij} can be written as

$$\begin{aligned} s_{ij} &= \sum_{k=-\infty}^{\infty} c_{ij}^{(1)}(k)\{\alpha_i(k)\beta_j(k) - \Sigma_{ij}\} + \sum_{k=-\infty}^{\infty} \Big[\sum_{\ell=-\infty}^{k-1} c_{ij}^{(2)}(k,\ell)\beta_j(\ell)\Big]\alpha_i(k) \\ &\quad + \sum_{k=-\infty}^{\infty} \Big[\sum_{\ell=-\infty}^{k-1} c_{ij}^{(3)}(k,\ell)\alpha_i(\ell)\Big]\beta_j(k), \end{aligned} \tag{A.2}$$

where the c_{ij}-coefficients are nonstochastic. We will use $\tilde{s}$ to denote a variate defined by the same coefficients in (A.2) but with $\Sigma_{ij} = 0$ and with $\alpha_i(k)$, $\beta_j(k)$ replaced by variates $\tilde{\alpha}_i(k)$, $\tilde{\beta}_j(k)$ (having mean zero and variance one) from series $\{\tilde{\alpha}_i(k)\}_{-\infty<k<\infty}$, $1 \leq i \leq d(\alpha)$, $\{\tilde{\beta}_j(k)\}_{-\infty<k<\infty}$, $1 \leq j \leq d(\beta)$, which are *statistically independent* of one another. Let us denote the sums on the right in (A.2) by $s_{ij}^{(1)}, s_{ij}^{(2)}$ and $s_{ij}^{(3)}$, according to the superscripts of the coefficients. The sum s in (A.1) clearly satisfies

$$\|s\|_{2p} \leq \sum_{i=1}^{d(\alpha)} \sum_{j=1}^{d(\beta)} \sum_{n=1}^{3} \left\|s_{ij}^{(n)}\right\|_{2p}. \tag{A.3}$$

If we can demonstrate the existence of constants $K_p^{(n)}$, $1 \leq n \leq 3$, depending only on p, C_p and Σ, such that (A.4) holds for all i, j and n,

$$\left\| s_{ij}^{(n)} \right\|_{2p} \leq K_p^{(n)} \left\| \tilde{s}_{ij}^{(n)} \right\|_2, \tag{A.4}$$

then, setting $K_p^* = \max\{K_p^{(1)}, K_p^{(2)}, K_p^{(3)}\}$, it follows from Cauchy-Schwarz, (A.3), and the independence of the $\tilde{s}_{ij}^{(n)}$ that

$$\begin{aligned} \|s\|_{2p} &\leq K_p^* \sum_{i=1}^{d(\alpha)} \sum_{j=1}^{d(\beta)} \sum_{n=1}^{3} \left\| \tilde{s}_{ij}^{(n)} \right\|_2 \\ &\leq (3d(\alpha)d(\beta))^{1/2} K_p^* \left[\sum_{i=1}^{d(\alpha)} \sum_{j=1}^{d(\beta)} \sum_{n=1}^{3} \left\| \tilde{s}_{ij}^{(n)} \right\|_2^2 \right]^{1/2} \\ &= (3d(\alpha)d(\beta))^{1/2} K_p^* \|\tilde{s}\|_2, \end{aligned}$$

the last equality being a consequence of the fact the $\tilde{s}_{ij}^{(n)}$ are uncorrelated and have mean zero. To establish (2.6), it therefore suffices to verify the special cases (A.4).

We start with several observations, beginning with the formulas

$$\begin{aligned} \left\| \tilde{s}_{ij}^{(1)} \right\|_2^2 &= \sum_{k=-\infty}^{\infty} \left| c_{ij}^{(1)}(k) \right|^2 \\ \left\| \tilde{s}_{ij}^{(n)} \right\|_2^2 &= \sum_{k=-\infty}^{\infty} \sum_{\ell=-\infty}^{k-1} \left| c_{ij}^{(n)}(k, \ell) \right|^2 \quad (n = 2, 3). \end{aligned} \tag{A.5}$$

Now consider the case $n = 1$, noting that each process $\{\alpha_i(t)\beta_j(t) - \Sigma_{ij}\}_{-\infty<t<\infty}$ is an F_t-adapated martingale difference sequence. By the convexity of the $2p$th power function and by (2.3), the $2p$th conditional moments of this process are almost surely bounded:

$$\begin{aligned} &\mathcal{E}\left\{ \left(\alpha_i(t)\beta_j(t) - \Sigma_{ij} \right)^{2p} \middle| F_{t-1} \right\} \leq 2^{2p} \left[\mathcal{E}\left\{ (\alpha_i(t)\beta_j(t))^{2p} | F_{t-1} \right\} + \Sigma_{ij}^{2p} \right] \\ &\leq 2^{2p} \left[\mathcal{E}\left\{ \alpha_i^{4p}(t) | F_{t-1} \right\} E\left\{ \beta_j^{4p}(k) | F_{t-1} \right\} \right]^{1/2} + 2^{2p} \Sigma_{ij}^{2p} \leq 2^{2p} \left\{ C_p + \Sigma_{ij}^{2p} \right\} \end{aligned}$$

holds for all t with probability one. Therefore, Lemma 2 of Wei (1987) and (A.3) yield (A.4) for this case.

For the case $n = 2$, set $h_{ij}^{(2)}(k) = \sum_{\ell=-\infty}^{k-1} c_{ij}(k, \ell)\beta_j(\ell)$, so that

$$s_{ij}^{(2)} = \sum_{k=-\infty}^{\infty} h_{ij}^{(2)}(k)\alpha_i(k).$$

Since $h_{ij}^{(2)}(k)$ is F_{k-1}-measurable, the proof of Lemma 2 of Wei (1987, pp.1677-8) shows that there is a constant B_p depending only on p such that

$$\begin{aligned}\left\|s_{ij}^{(2)}\right\|_{2p}^{2} &\leq (C_pB_p)^{1/p}\left\|\sum_{k=-\infty}^{\infty}\left(h_{ij}^{(2)}(k)\right)^2\right\|_p \\ &\leq (C_pB_p)^{1/p}\sum_{k=-\infty}^{\infty}\left\|h_{ij}^{(2)}(k)\right\|_{2p}^{2},\end{aligned}$$

because $\|(h_{ij}^{(2)}(k))^2\|_p = \|h_{ij}^{(2)}(k)\|_{2p}^2$. By the same reasoning

$$\left\|h_{ij}^{(2)}(k)\right\|_{2p}^{2} \leq (C_pB_p)^{1/p}\sum_{\ell=-\infty}^{k-1}\left|c_{ij}^{(2)}(k,\ell)\right|^2,$$

leading to (A.4) via (A.5) again. The same reasoning applies to the case $n = 3$, so the proof of the theorem is complete.

A.2. Proof of the Second Moment Bound Theorem

In conformity with (2.17), we denote the variate under investigation in (4.4) by $R_{uv:w}(T)$. By Cauchy-Schwarz, as in the derivation of (2.22), we have

$$\|R_{uv:w}(T)\|_\rho \leq \|R_{uu:w}(T)\|_\rho^{1/2}\|R_{vv:w}(T)\|_\rho^{1/2}.$$

Thus it suffices, by the symmetry of the roles of u and v, to verify

$$\sup_{T\geq 1,\, p/2\leq\rho\leq p} T^{p/2-\rho}\mathcal{E}R_{uu:w}^{\rho}(T) < \infty. \tag{A.6}$$

Since $\rho \leq p$ and $\|\cdot\|_{2\rho} \leq \|\cdot\|_{2p}$, it follows from (1.13) and (4.1), assuming $M \geq 1$, that

$$\mathcal{E}R_{uu:w}^{2\rho}(T) \leq T^{2\rho}M^{2p} \tag{A.7}$$

holds for all T. For some $0 < \delta < \lambda_{\min}(\Gamma_{ww})$, consider the event

$$G(T) \equiv \left\{\left\|T^{-1}\sum_{t=1}^{T}\left(w(t)w(t)' - \Gamma_{ww}\right)\right\| > \delta\right\}.$$

By Chebyshev's inequality, this has probability satisfying

$$P(G(T)) \leq \frac{\mathcal{E}\left\{\left\|T^{-1/2}\sum_{t=1}^{T}\left(w(t)w(t)' - \Gamma_{ww}\right)\right\|^{2p}\right\}}{T^p\delta^{2p}} \leq M^{2p}(\delta^2T)^{-p}. \tag{A.8}$$

On the complement $\bar{G}(T)$, clearly

$$\lambda_{\min}\left[\sum_{t=1}^{T} w_t w_t'\right] \geq T\left(\lambda_{\min}(\Gamma_{ww}) - \delta\right). \tag{A.9}$$

We have, by (4.2), (A.7–9), and Cauchy-Schwarz,

$$\begin{aligned}
\mathcal{E}\left\{R^{\rho}_{uu:w}(T)\right\} &= \mathcal{E}\left\{R^{\rho}_{uu:w}(T)\cdot I[G(T)]\right\} + \mathcal{E}\left\{R^{\rho}_{uu:w}(T)\cdot I[\bar{G}(T)]\right\} \\
&\leq \left\{\mathcal{E} R^{2\rho}_{uu:w}(T)\right\}^{1/2} P(G(T))^{1/2} + \left\{M/(\lambda_{\min}(\Gamma_{ww}) - \delta)\right\}^{\rho} \\
&\leq T^{\rho-p/2} M^{2p}\delta^{-p} + \left\{M/(\lambda_{\min}(\Gamma_{ww}) - \delta)\right\}^{\rho}
\end{aligned}$$

for all T, from which (A.6) follows. This completes the proof.

References

Findley, D. F. (1985). On the unbiasedness property of AIC for exact or approximating linear stochastic time series models. *J. Time Ser. Anal.* **6**, 229–252.

Findley, D. F. (1990). Making difficult model comparisons. Statistical Research Division Research Report RR-90/11, Bureau of the Census, Washington, D.C.

Findley, D. F. (1991a). Model selection for multistep-ahead forecasting. *Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, 243–247.

Findley, D. F. (1991b). Convergence of finite multistep predictors from incorrect models and its role in model selection. *Note di Matematica* **11** (to appear).

Findley, D. F. and Wei, C.-Z. (1993). Vector autoregression, AIC and moments of the sample variance matrix inverse. (submitted for publication)

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic Press, New York.

Hannan, E. J. (1976). The asymptotic distribution of serial covariances. *Ann. Statist.* **4**, 396–399.

Hannan, E. J. (1987). Approximations of linear systems. In *Time Series and Econometric Modeling* (Edited by I. B. MacNeil and G. J. Humphrey), 1–12, D. Reidel, Dordrecht.

Hosoya, Y. and Taniguchi, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes. *Ann. Statist.* **10**, 132–154.

Loève, M. (1977). *Probability Theory I*, 4th edition. Springer-Verlag, New York.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd edition. John Wiley, New York.

Rozanov, Yu A. (1967). *Stationary Random Processes.* Holden-Day, San Francisco.

Serfling, R. J. (1968). Contributions to central limit theory for dependent variables. *Ann. Math. Statist.* **39**, 1158–1175.

Titchmarsh, E. C. (1939). *The Theory of Functions*, 2nd edition. Oxford University Press, London.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57**, 307–333.

Wei, C.-Z. (1987). Adaptive prediction by least squares predictors in stochastic regression models with applications to time series. *Ann. Statist.* **15**, 1667–1682.

Bureau of the Census, Statistical Research Division, Room 3000–4, Washington, DC 20233, U.S.A.

Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan.

(Received March 1992; accepted December 1992)