The Annals of Statistics
1980, Vol. 8, No. 1, 132–146

ON THE ASYMPTOTIC PROPERTIES OF LEAST-SQUARES ESTIMATORS IN AUTOREGRESSION

BY MARTIN J. CROWDER
Surrey University

Consistency and asymptotic normality of least-squares estimators are discussed for the linear autoregressive model with explanatory variables. Few assumptions are made about the error sequence. The case of stochastic explanatory variables is also considered.

1. Introduction. The asymptotic theory of least-squares estimators in linear autoregression at present (e.g., Anderson, 1971) relies on rather restrictive assumptions. It is the purpose of this paper to show how very much weaker assumptions can suffice, thereby extending the applicability of the model. Asymptotic theory is useful when a large amount of data is available, but in such cases there will often be considerable heterogeneity within the data giving rise to serious departures from the usual assumptions. We write the basic model in the form (1.1) below involving just the mean value of Y_t, given the previous history of the process. This representation is probably the most readily interpretable and the mean is usually a feature of chief interest. The analysis, proofs of consistency and asymptotic normality of the estimators, follows from (1.1) alone, apart from various technical provisos. This is because (1.1) implies that the random errors in (1.2) form a martingale difference sequence, which is central for the various limit theorems. No particular distributional assumptions are made and most of the analysis is carried out in terms of second-moment properties only.

We consider the pth order, univariate, autoregressive scheme with q explanatory variables:

$$E[Y_t|\mathcal{F}_{t-1}] = \boldsymbol{\beta}^T\mathbf{Y}_{t-1} + \boldsymbol{\alpha}^T\mathbf{z}_t, \qquad t \geqslant 1, \tag{1.1}$$

where $\boldsymbol{\beta}^T = (\beta_1, \cdots, \beta_p)$, $\boldsymbol{\alpha}^T = (\alpha_1, \cdots, \alpha_q)$, $\mathbf{Y}_{t-1}^T = (Y_{t-1}, \cdots, Y_{t-p})$, $\mathbf{z}_t^T = (z_{1t}, \cdots, z_{qt})$; $\boldsymbol{\alpha}$ *and* $\boldsymbol{\beta}$ are the parameters to be estimated, the $\mathbf{z}_t$'s are explanatory variables, and $\mathcal{F}_{t-1}$ is the σ-field of events generated by $(\mathbf{Y}_0, Y_1, \cdots, Y_{t-1})$. Let

$$U_t = Y_t - E[Y_t|\mathcal{F}_{t-1}]. \tag{1.2}$$

It is common to assume that the U_t's are independently and identically distributed, often normally (Anderson, 1971, Chapter 5). For such reasons as those outlined above the statistician will frequently have cause to regard such strong assumptions as unrealistic and we proceed without them.

The motivation for this work lies in the statistical analysis of certain epidemiological data from an ongoing, large-scale survey among general practitioners in the

Received December 1976; revised May 1978.
AMS 1970 *subject classifications.* Primary 62M10; secondary 62J05.
Key words and phrases. Least-squares, autoregression, asymptotic estimation theory.

U.K.; this is described in Crowder and Grob (1975). We wish to apply the results of the present paper with Y_t representing the observed prevalence of a given human disease in a community at week number t, and $\mathbf{z}_t$ representing various possible explanatory variables such as weather indices, other related diseases, and environmental factors. In the previous paper a limited analysis was carried out in which Y_t was reduced to binary form (scoring 1 if Y_t exceeds a specified count and 0 if not), and a certain two-state Markov chain model was used incorporating just three z's, all weather variables.

2. Consistency and asymptotic normality of the least-squares estimators. We suppose that observations $(Y_{1-p}, \cdots, Y_0, \cdots, Y_n)$ are available and estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ by minimising the sum (over $t = 1, \cdots, n$) of squared residuals. Let $\boldsymbol{\mu}_t = E[\mathbf{Y}_t]$ and $\sigma_t^2 = V[U_t]$ for $t \geqslant 1$; unweighted least-squares is employed because the nuisance parameters σ_t^2 are taken to be unknown. The equations for the least-squares estimators $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}})$ are

$$(2.1) \qquad \Sigma \mathbf{J}_t \begin{pmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{\alpha}} \end{pmatrix} = \begin{pmatrix} \Sigma Y_t \mathbf{Y}_{t-1} \\ \Sigma Y_t \mathbf{z}_t \end{pmatrix}, \quad where \quad \mathbf{J}_t = \begin{pmatrix} \mathbf{Y}_{t-1}\mathbf{Y}_{t-1}^T & \mathbf{Y}_{t-1}\mathbf{z}_t^T \\ \mathbf{z}_t \mathbf{Y}_{t-1}^T & \mathbf{z}_t \mathbf{z}_t^T \end{pmatrix};$$

all summations are over $t = 1, \cdots, n$ unless otherwise indicated. Let $\mathbf{V}_n = E[\Sigma \mathbf{Y}_{t-1}\mathbf{Y}_{t-1}^T]$, $\mathbf{L}_n = \Sigma \boldsymbol{\mu}_{t-1}\mathbf{z}_t^T$, $\mathbf{M}_n = \Sigma \mathbf{z}_t \mathbf{z}_t^T$, $\mathbf{G}_n = \mathbf{M}_n^{-1}\mathbf{L}_n^T$ (assuming $\mathbf{M}_n^{-1}$ exists), and $\mathbf{D}_n = E[\Sigma \mathbf{J}_t]$; also let m_n and e_n be respectively the smallest eigenvalues of $\mathbf{M}_n$ and $\mathbf{E}_n = \mathbf{V}_n - \mathbf{L}_n \mathbf{M}_n^{-1} \mathbf{L}_n^T$. We will denote by z_t the vector norm $|\mathbf{z}_t| = (\Sigma_i \{\mathbf{z}_t\}_i^2)^{\frac{1}{2}}$, and by g_n the Euclidean matrix norm $|\mathbf{G}_n| = (\Sigma_{ij}\{\mathbf{G}_n\}_{ij}^2)^{\frac{1}{2}}$. We use $\sigma_0^2 = |V[\mathbf{Y}_0]|$ and $\nu_t = \Sigma_{r=0}^t \lambda^{2r}\sigma_{t-r}^2$ where λ exceeds the largest of the moduli of the roots of the polynomial $x^p - \beta_1 x^{p-1} - \cdots - \beta_p$.

Our main results on the asymptotic properties of the least-squares estimators are now summarised, all proofs being relegated to later sections.

THEOREM 1. *The following conditions are sufficient for consistency of the least-squares estimators* $(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}})$.

(a) $c_n^{-2}\Sigma \lambda^{t-1}(\Sigma_{r=0}^{n-t}\nu_r \mathbf{z}_{r+1}^t \mathbf{z}_{r+t}) \to 0$ *as* $n \to \infty$ *when* $c_n = e_n, e_n/g_n^2$ *and* m_n;

(b) $c_n^{-2}\Sigma z_t^2 \sigma_t^2 \to 0$ *as* $n \to \infty$ *when* $c_n = e_n, e_n/g_n$ *and* m_n;

(c) $c_n^{-1}\{\nu_n + (\Sigma_{r=0}^n \lambda^r z_{n-r})^2\} \to 0$ *as* $n \to \infty$ *when* $c_n = e_n$ *and* e_n/g_n;

(d) *either* $c_n^{-2}\Sigma \sigma_{1t}^2\{\nu_{t-1} + (\Sigma_{r=0}^{t-1}\lambda^r z_{t-1-r})^2\} \to 0$ *or* $c_n^{-2}\Sigma \gamma_t^2\{\Sigma_{r=0}^{t-1}\lambda^r(\gamma_{t-1-r} + z_{t-1-r})\}^2 \to 0$ *for* $c_n = e_n$ *and* e_n/g_n, *where* $\gamma_t^4 = E[U_t^4]$ *and* $\sigma_{1t}^2 \geqslant E[U_t^2|\mathcal{F}_{t-1}]$ *is a constant bound for the conditional error variance*;

(e) $c_n^{-1}\Sigma(U_t^2 - \sigma_t^2) \underset{ms}{\to} 0$ *as* $n \to \infty$ *when* $c_n = e_n$ *and* e_n/g_n;

(f) $\lambda < 1$.

THEOREM 2. *Let* $X_t = U_t(\mathbf{Y}_{t-1}^T, \mathbf{z}_t^T)\mathbf{x}$ *for a given unit vector* $\mathbf{x}$, *and write* $s_n^2 = E[\Sigma X_t^2]$, $\mathbf{C}_n = E[\Sigma U_t^2 \mathbf{J}_t]$. *Suppose that the conditions of Theorem* 1 *hold, and that for all* $\mathbf{x}$

$$s_n^{-2}\Sigma X_t^2 \to_p 1, \; s_n^{-2}\Sigma X_t^2 I\{|X_t| \geqslant \varepsilon s_n\} \to_p 0 \;\forall \varepsilon > 0,$$

where $I\{A\}$ *denotes the indicator function of the set* A. *Then the asymptotic distribution of* $\begin{pmatrix}\tilde{\boldsymbol\beta}\\ \tilde{\boldsymbol\alpha}\end{pmatrix}$ *is* $N_{p+q}\left(\begin{pmatrix}\boldsymbol\beta\\ \boldsymbol\alpha\end{pmatrix}, \mathbf{D}_n^{-1}\mathbf{C}_n\mathbf{D}_n^{-1}\right)$.

In Theorems 1 and 2 it is implicit that the $\mathbf{z}_t$'s are fixed, or that the inference is conditional on their observed values. The next result aims to drop this restriction, i.e., to allow $\{\mathbf{z}_t\}$ to be a stochastic process and obtain unconditional asymptotic results for $\tilde{\boldsymbol\alpha}$ and $\tilde{\boldsymbol\beta}$.

THEOREM 3. *Let* $\{\mathbf{z}_t\}$ *be a stochastic process independent of* $\{U_t\}$ *and suppose that the conditions of Theorems* 1 *and* 2 *each hold "in probability". Then the conclusions of those theorems are preserved.*

The assumption that $\{\mathbf{z}_t\}$ is independent of $\{U_t\}$ means that the z's are so-called exogeneous variables; they influence $\{Y_t\}$ but not vice-versa.

Three lemmas are given here because they prove useful in checking the conditions of Theorem 1. The first two lemmas provide bounds for e_n and g_n, mainly useful as orders of magnitude. The third gives a criterion for the convergence required in condition (e) in terms of the "long-distance" correlations of $\{U_t^2\}$.

LEMMA 1. $e_n \geqslant a\Sigma_{t=2}^{n-p+1}\sigma_t^2 + \inf\{\mathbf{f}^T\mathbf{G}_n^T\mathbf{M}_n'\mathbf{G}_n\mathbf{f}\}$ (inf *over* $|\mathbf{f}| = 1$) *for some* $a > 0$, *where* $\mathbf{M}_n' = \mathbf{M}_n - n\bar{\mathbf{z}}_n\bar{\mathbf{z}}_n^T$, $\bar{\mathbf{z}}_n = n^{-1}\Sigma\mathbf{z}_t$.

LEMMA 2. $g_n \leqslant am_n^{-1}\Sigma z_t(\Sigma_{r=0}^{t-1}\lambda^r z_{t-r-1})$ *for some* $a < \infty$.

LEMMA 3. *Let* $\gamma_{st} = C[U_s^2, U_t^2]$ *be the covariance between* U_s^2 *and* U_t^2, *and let* $\rho_{st} = \gamma_{st}/(\gamma_{ss}\gamma_{tt})^{\frac{1}{2}}$. *If* $|\rho_{st}| \leqslant \phi_{|s-t|}$, *and* $c_n^{-2}(\Sigma\gamma_{tt})(1 + \Sigma\phi_t) \to 0$ *as* $n \to \infty$ *then* $c_n^{-1}\Sigma(U_t^2 - \sigma_t^2) \to_{ms} 0$.

Finally in this section we give a theorem on consistency of $(\tilde{\boldsymbol\alpha}, \tilde{\boldsymbol\beta})$ whose premises are implied by those of Theorem 1. Thus Theorem 4 is more general than Theorem 1, but its conditions, involving matrices, are usually more difficult to check. Let $\mathbf{W}_{1n} = \Sigma(\mathbf{Y}_{t-1} - \mathbf{G}_n^T\mathbf{z}_t)\mathbf{z}_t^T$, $\mathbf{W}_{2n} = \Sigma(\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}^T - E[\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}^T])$, $\mathbf{W}_{3n} = \Sigma\mathbf{z}_tU_t$, and $\mathbf{W}_{4n} = \Sigma\mathbf{Y}_{t-1}U_t$.

THEOREM 4. *The least-squares estimators* $(\tilde{\boldsymbol\alpha}, \tilde{\boldsymbol\beta})$ *are consistent if the limits L*1 *to L*4 *below hold. If L*5 *also holds then their asymptotic distribution is normal as in Theorem* 2.

L1: $\mathbf{E}_n^{-1}\mathbf{W}_{1n} \to_p \mathbf{0}$, $\mathbf{G}_n\mathbf{E}_n^{-1}\mathbf{W}_{1n} \to_p \mathbf{0}$, $\mathbf{E}_n^{-1}\mathbf{G}_n^T\mathbf{W}_{1n} \to_p \mathbf{0}$, $\mathbf{G}_n\mathbf{E}_n^{-1}\mathbf{G}_n^T\mathbf{W}_{1n}^T \to_p \mathbf{0}$, $\mathbf{M}_n^{-1}\mathbf{W}_{1n}^T \to_p \mathbf{0}$;

L2: $\mathbf{E}_n^{-1}\mathbf{W}_{2n} \to_p \mathbf{0}$, $\mathbf{G}_n\mathbf{E}_n^{-1}\mathbf{W}_{2n} \to_p \mathbf{0}$;

L3: $\mathbf{E}_n^{-1}\mathbf{G}_n^T\mathbf{W}_{3n} \to_p \mathbf{0}$, $\mathbf{G}_n\mathbf{E}_n^{-1}\mathbf{G}_n^T\mathbf{W}_{3n} \to_p \mathbf{0}$, $\mathbf{M}_n^{-1}\mathbf{W}_{3n} \to_p \mathbf{0}$;

L4: $\mathbf{E}_n^{-1}\mathbf{W}_{4n} \to_p \mathbf{0}$, $\mathbf{G}_n\mathbf{E}_n^{-1}\mathbf{W}_{4n} \to_p \mathbf{0}$;

L5: $\mathbf{C}_n^{-\frac{1}{2}}\Sigma U_t\begin{pmatrix}\mathbf{Y}_{t-1}\\ \mathbf{z}_t\end{pmatrix} \sim_d N_{p+q}(\mathbf{0}, \mathbf{I}_{p+q})$.

(By $\sim_d$ we mean "is asymptotically distributed as".)

3. Discussion and examples. The very interesting paper of Anderson (1959) treats the case $q = 0, p = 1$ and is mainly concerned with unstable processes where $\lambda > 1$. The examination is extended to vector Y_t and where $\boldsymbol{\beta}$ has all eigenvectors greater than 1. The case $\lambda < 1$ is also discussed but Anderson indicates that his methods will not work for $q > 0$.

The asymptotic theory as given by Anderson (1971, pages 188–211) is derived under fairly narrow conditions: (i) the sequence $\{\sigma_t^2\}$ is assumed constant ($\sigma_t^2 = \sigma^2$ for all t) and the U_t's are assumed independent and either identically distributed or with uniformly bounded moments of order $2 + \varepsilon$ ($E[|U_t|^{2+\varepsilon}] < k$ for all t); (ii) the sequence $\{\mathbf{z}_t\}$ is assumed uniformly bounded ($|\mathbf{z}_t| < z$ for all t), although it is mentioned that this may be relaxed to $\Sigma_1^\infty \lambda^t z_t < \infty$; (iii) certain convergence assumptions are made, amounting to $n^{-1}\mathbf{M}_n \to \mathbf{M}$, $n^{-1}\mathbf{L}_n \to \mathbf{L}$, and $n^{-1}\Sigma \boldsymbol{\mu}_t \boldsymbol{\mu}_t^T \to \mathbf{H}$. The condition (i) seems likely to be violated in practice because circumstances, in particular factors governing the variation of recorded observations, change over long periods of time. Likewise, (ii) sometimes cannot be guaranteed, for instance, it fails for the familiar linear regression form with $\mathbf{z}_t = (1, t)^T$. Condition (iii) also fails for $\mathbf{z}_t = (1, t)^T$ since the elements of $\mathbf{M}_n$, for instance, are not all $0(n)$; even when $\mathbf{M}_n$ is $0(n)$ the convergence requires some stability over a long period, and this may not obtain.

In the derivations given in this paper we have worked in terms of mean-square convergence for one main reason, being the simplification made possible by the identity $E[U_s U_t] = 0, s \neq t$. This orthogonality follows immediately from the nature of $\{U_t\}$ as a martingale-differences sequence, i.e. $E[U_t | \mathcal{F}_{t-1}] = 0$ a.s., which in turn is a result solely of the definition (1.2). The much stronger property of independence has not been called upon. Likewise, homoscedasticity has played no part, either in a strong sense ($E[U_t^2 | \mathcal{F}_{t-1}] = \sigma^2$) or in a weak sense ($E[U_t^2] = \sigma^2$); only bounds on the moments of $\{U_t\}$ are needed. The conditions we have employed, such as $E[U_t^2 | \mathcal{F}_{t-1}] \leqslant \sigma_{1t}^2$ a.s., have been chosen as much with regard to their applicability as to their generality. It can be verified that at various points the arguments go through under a variety of assumptions.

We give now some examples which fall outside one or more of the conditions (i) to (iii) above, but which are amenable to the present treatment. For brevity they are simplified versions of situations which may arise in applications, and we assume $\lambda < 1$ throughout.

EXAMPLE 1. We first outline a situation which is more or less the "standard" case. That is, $z_t \leqslant z$, $\sigma_t^2 = 0(1)$, $\gamma_t^4 \leqslant \gamma^4$, $\gamma_{tt} = 0(1)$ for all t, and $\rho_{st} \to 0$ as $|s - t| \to \infty$; also assume $m_n = 0(n)$, i.e., the $\mathbf{z}_t$'s are not in the main confined to a subspace of dimension less than q. Then by Lemma 1 $e_n \geqslant a\Sigma\sigma_t^2 = 0(n)$, and by Lemma 2 $g_n \leqslant 0(1)$. Also $\nu_t = 0(1)$ for all t. We can now check the conditions of Theorem 1 as follows:

(a) $c_n^{-2}\Sigma\lambda^{t-1}(\Sigma_{r=0}^{n-t}\nu_r \mathbf{z}_{r+1}^T \mathbf{z}_{r+t}) = c_n^{-2}\Sigma\lambda^{t-1}0(n-t) = c_n^{-2}0(n) \to 0$ for $c_n = 0(n)$;

(b) $c_n^{-2}\Sigma z_t^2\sigma_t^2 = c_n^{-2}0(n) \to 0$ for $c_n = 0(n)$;
(c) $c_n^{-1}\{\nu_n + (\Sigma_{r=0}^n \lambda^r z_{n-r})^2\} = c_n^{-1}0(1) \to 0$ for $c_n = 0(n)$;
(d) $c_n^{-2}\Sigma\gamma_t^2\{\Sigma_{r=0}^{t-1}\lambda^r(\gamma_{t-1-r} + z_{t-1-r})\}^2 = c_n^{-2}0(n) \to 0$ for $c_n = 0(n)$;
(e) $c_n^{-1}\Sigma(U_t^2 - \sigma_t^2) \to_{ms} 0$ for $c_n = 0(n)$ by Lemma 3.

EXAMPLE 2. The first nonstandard situation which we consider is when $\sigma_t^2 \to \infty$, i.e., the mechanism deteriorates and runs out of control eventually. To be specific take $E[U_t^2|\mathcal{F}_{t-1}] = 0(t^\eta)$ for $0 < \eta < 1$; then $\sigma_t^2 = 0(t^\eta)$, $\nu_t = 0(t^\eta)$, and $e_n \geqslant 0(n^{\eta+1})$ by Lemma 1. Suppose that $z_t = 0(1)$, $m_n = 0(n)$, so $g_n \leqslant 0(1)$ by Lemma 2. Then conditions (a), (b), (c) and (d) of Theorem 1 can be verified as in Example 1; condition (e) holds if $|\rho_{st}| \leqslant \phi_{|s-t|}$ and $n^{-2(\eta+1)}(\Sigma\gamma_{tt})(1 + \Sigma\phi_t) \to 0$, which is achieved, for instance, with $\Sigma_1^\infty\phi_t < \infty$ and $\gamma_{tt} = 0(t^{2\eta+1})$.

EXAMPLE 3. Suppose that the effect of $\mathbf{z}_t$ fades with time, then the question arises as to whether the information about $\boldsymbol{\alpha}$ provided by observing $\{Y_t\}$ dries up too quickly for consistent estimation to be possible. For illustration take $q = 1$, $\mathbf{z}_t = (t^{-\frac{1}{2}})$, $\sigma_t^2 = 0(1)$ and $\gamma_t = 0(1)$. Then $m_n = \Sigma t^{-1} \sim \ln n$, $e_n \geqslant 0(n)$ by Lemma 1, $g_n \leqslant 0(n^{\frac{1}{2}}/\ln n)$ by Lemma 2, $\nu_t = 0(1)$ and conditions (a) to (d) of Theorem 1 may be verified as before. The additional assumption $n^{-1}(\ln n)^{-2}(\Sigma\gamma_{tt})(1 + \Sigma\phi_t) \to 0$ suffices for (e), using Lemma 3. On the other hand, if $\mathbf{z}_t$ were (t^{-1}) then $m_n = \Sigma t^{-2} \not\to \infty$ and the conditions of Theorem 1 cannot be met. Since $\mathbf{M}_n^{-1} = (\Sigma t^{-2})^{-1} \not\to \mathbf{0}$, L1 and L3 of Theorem 4 fail too, and it seems likely that consistent estimation of $\boldsymbol{\alpha}$ is impossible in this case.

EXAMPLE 4. We now give an illustration of how Theorem 4 can be brought to the rescue when the conditions of Theorem 1 fail. We take as our model $E[Y_t|\mathcal{F}_{t-1}] = \beta Y_{t-1} + \alpha_1 + \alpha_2 t$, simple linear regression on time with $\mathbf{z}_t = (1, t)^T$; also $\sigma_t^2 = 0(1)$ and $\gamma_t = 0(1)$. It turns out that $m_n^{-1}|\mathbf{W}_{1n}^T| \to \infty$ whereas $\mathbf{M}_n^{-1}\mathbf{W}_{1n}^T \to_{ms} \mathbf{0}$, the discrepancy arising from the reduction of matrix limits to scalar ones. Proceeding as before, we find that $e_n \geqslant a\Sigma\sigma_t^2 = 0(n)$, $\nu_t = 0(1)$, $m_n = 0(n)$ and $g_n \leqslant 0(n^2)$. The expressions involved in Theorem 1 turn out as follows: (a) is $c_n^{-2}0(n^3)$, (b) is $c_n^{-2}0(n^3)$, (c) is $c_n^{-1}0(n^2)$ and (d) is $c_n^{-2}0(n^3)$. Thus the conditions fail. We can retrieve those involving e_n and g_n by more careful calculations, but since this work depends on formulae to be developed later we just quote the results here: we find $g_n = 0(1)$, and use Lemma 1 to show $e_n \geqslant \mathbf{G}_n^T\mathbf{M}_n'\mathbf{G}_n = 0(n^3)$. Thus only the conditions involving m_n are outstanding and for those we have to attack the corresponding ones involving $\mathbf{M}_n$ in Theorem 4, i.e., $\mathbf{M}_n^{-1}\mathbf{W}_{1n}^T \to_p \mathbf{0}$ and $\mathbf{M}_n^{-1}\mathbf{W}_{3n} \to_{ms} \mathbf{0}$. The verifications of these limits are also given later in the Appendix. We note that such a discrepancy, as between $m_n^{-1}\mathbf{W}_{1n}^T$ and $\mathbf{M}_n^{-1}\mathbf{W}_{1n}^T$, cannot occur when the eigenvalues of $\mathbf{M}_n$ are of comparable orders, for then $m_n^{-1}\mathbf{M}_n$ is positive definite with eigenvalues lying between 1 and $m < \infty$, so $\mathbf{M}_n^{-1}\mathbf{W}_{1n}^T = (m_n^{-1}\mathbf{M}_n)^{-1}(m_n^{-1}\mathbf{W}_{1n}^T) \to \mathbf{0}$ iff $m_n^{-1}\mathbf{W}_{1n}^T \to \mathbf{0}$.

4. Theorem 4: Proof and discussion. A necessary and sufficient condition for consistency of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ would be, from (2.1),

$$(\Sigma \mathbf{J}_t)^{-1}\begin{pmatrix} \Sigma Y_t \mathbf{Y}_{t-1} \\ \Sigma Y_t \mathbf{z}_t \end{pmatrix} \to_p \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix}.$$

(If $\Sigma \mathbf{J}_t$ were singular (2.1) would have no unique solution anyway.) However, such a condition would be of little practical value, so we have chosen to multiply (2.1) by $\mathbf{D}_n^{-1}$, and confine attention to cases where $\mathbf{D}_n^{-1}\Sigma \mathbf{J}_t \to_p \mathbf{I}_{p+q}$. We have

$$\mathbf{D}_n^{-1} = \begin{bmatrix} \mathbf{E}_n^{-1} & -\mathbf{E}_n^{-1}\mathbf{G}_n^T \\ -\mathbf{G}_n\mathbf{E}_n^{-1} & \mathbf{M}_n^{-1} + \mathbf{G}_n\mathbf{E}_n^{-1}\mathbf{G}_n^T \end{bmatrix},$$

$$\mathbf{D}_n^{-1}\Sigma \mathbf{J}_t = \begin{bmatrix} \mathbf{E}_n^{-1}(\Sigma \mathbf{Y}_{t-1}\mathbf{Y}_{t-1}^T - \mathbf{G}_n^T\Sigma \mathbf{z}_t\mathbf{Y}_{t-1}^T) & \mathbf{E}_n^{-1}(\Sigma \mathbf{Y}_{t-1}\mathbf{z}_t^T - \mathbf{G}_n^T\Sigma \mathbf{z}_t\mathbf{z}_t^T) \\ \mathbf{M}_n^{-1}\Sigma \mathbf{z}_t\mathbf{Y}_{t-1}^T - \mathbf{G}_n\mathbf{E}_n^{-1}(\Sigma \mathbf{Y}_{t-1}\mathbf{Y}_{t-1}^T - \mathbf{G}_n^T\Sigma \mathbf{z}_t\mathbf{Y}_{t-1}^T) & \mathbf{I}_q - \mathbf{G}_n\mathbf{E}_n^{-1}(\Sigma \mathbf{Y}_{t-1}\mathbf{z}_t^T - \mathbf{G}_n^T\Sigma \mathbf{z}_t\mathbf{z}_t^T) \end{bmatrix},$$

$$\mathbf{D}_n^{-1}\begin{bmatrix} \Sigma Y_t\mathbf{Y}_{t-1} \\ \Sigma Y_t\mathbf{z}_t \end{bmatrix} = \begin{bmatrix} \mathbf{E}_n^{-1}(\Sigma Y_t\mathbf{Y}_{t-1} - \mathbf{G}_n^T\Sigma Y_t\mathbf{z}_t) \\ \mathbf{M}_n^{-1}\Sigma Y_t\mathbf{z}_t - \mathbf{G}_n\mathbf{E}_n^{-1}(\Sigma Y_t\mathbf{Y}_{t-1} - \mathbf{G}_n^T\Sigma Y_t\mathbf{z}_t) \end{bmatrix}.$$

But

$$\begin{aligned} \Sigma \mathbf{Y}_{t-1}\mathbf{Y}_{t-1}^T - \mathbf{G}_n^T\Sigma \mathbf{z}_t\mathbf{Y}_{t-1}^T &= (\mathbf{W}_{2n} + \mathbf{V}_n) - \mathbf{G}_n^T\Sigma \mathbf{z}_t\mathbf{Y}_{t-1}^T \\ &= \mathbf{W}_{2n} + (\mathbf{E}_n + \mathbf{G}_n^T\mathbf{L}_n^T) - \mathbf{G}_n^T\Sigma \mathbf{z}_t\mathbf{Y}_{t-1}^T \\ &= \mathbf{W}_{2n} + \mathbf{E}_n - \mathbf{G}_n^T(\Sigma \mathbf{z}_t\mathbf{Y}_{t-1}^T - \Sigma \mathbf{z}_t\mathbf{z}_t^T\mathbf{G}_n) \\ &= \mathbf{W}_{2n} + \mathbf{E}_n - \mathbf{G}_n^T\mathbf{W}_{1n}^T, \end{aligned}$$

and

$$\mathbf{M}_n^{-1}\Sigma \mathbf{z}_t\mathbf{Y}_{t-1}^T - \mathbf{G}_n = \mathbf{M}_n^{-1}\mathbf{W}_{1n}^T,$$

so

$$\mathbf{D}_n^{-1}\Sigma \mathbf{J}_t = \begin{bmatrix} \mathbf{I}_p + \mathbf{E}_n^{-1}(\mathbf{W}_{2n} - \mathbf{G}_n^T\mathbf{W}_{1n}^T) & \mathbf{E}_n^{-1}\mathbf{W}_{1n} \\ \mathbf{M}_n^{-1}\mathbf{W}_{1n}^T + \mathbf{G}_n\mathbf{E}_n^{-1}(\mathbf{W}_{2n} - \mathbf{G}_n^T\mathbf{W}_{1n}^T) & \mathbf{I}_q - \mathbf{G}_n\mathbf{E}_n^{-1}\mathbf{W}_{1n} \end{bmatrix}. \tag{4.1}$$

Also

$$\begin{aligned} \mathbf{D}_n^{-1}\begin{pmatrix} \Sigma Y_t\mathbf{Y}_{t-1} \\ \Sigma Y_t\mathbf{z}_t \end{pmatrix} &= \mathbf{D}_n^{-1}\begin{pmatrix} \Sigma \mathbf{Y}_{t-1}(\mathbf{Y}_{t-1}^T\boldsymbol{\beta} + \mathbf{z}_t^T\boldsymbol{\alpha} + U_t) \\ \Sigma \mathbf{z}_t(\mathbf{Y}_{t-1}^T\boldsymbol{\beta} + \mathbf{z}_t^T\boldsymbol{\alpha} + U_t) \end{pmatrix} \\ &= \mathbf{D}_n^{-1}\Sigma \mathbf{J}_t\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix} + \mathbf{D}_n^{-1}\begin{pmatrix} \Sigma \mathbf{Y}_{t-1}U_t \\ \Sigma \mathbf{z}_t U_t \end{pmatrix} = \mathbf{D}_n^{-1}\Sigma \mathbf{J}_t\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix} \\ &\quad + \begin{pmatrix} \mathbf{E}_n^{-1}\mathbf{W}_{4n} - \mathbf{E}_n^{-1}\mathbf{G}_n^T\mathbf{W}_{3n} \\ -\mathbf{G}_n\mathbf{E}_n^{-1}\mathbf{W}_{4n} + \mathbf{M}_n^{-1}\mathbf{W}_{3n} + \mathbf{G}_n\mathbf{E}_n^{-1}\mathbf{G}_n^T\mathbf{W}_{3n} \end{pmatrix}. \end{aligned} \tag{4.2}$$

From (4.1) and (4.2) it will be seen that the limits L1 to L4 of Theorem 4 imply that $\mathbf{D}_n^{-1}\Sigma\mathbf{J}_t \to_p \mathbf{I}_{p+q}$ and $\mathbf{D}_n^{-1}(\Sigma Y_t\mathbf{Y}_{t-1}^T, \Sigma Y_t\mathbf{z}_t^T)^T \to_p (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$, and hence that $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\alpha}}$ are consistent.

For L5 of Theorem 4 we first note that a matrix version of a theorem of Cramér (1946, Section 20.6) may be used to imply that

$$\begin{pmatrix}\tilde{\boldsymbol{\beta}}\\ \tilde{\boldsymbol{\alpha}}\end{pmatrix} \sim_d \mathbf{D}_n^{-1}\begin{pmatrix}\Sigma Y_t\mathbf{Y}_{t-1}\\ \Sigma Y_t\mathbf{z}_t\end{pmatrix} \sim_d \begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\alpha}\end{pmatrix} + \mathbf{D}_n^{-1}\begin{pmatrix}\Sigma\mathbf{Y}_{t-1}U_t\\ \Sigma\mathbf{z}_tU_t\end{pmatrix},$$

using L1 to L4 and (4.2). The import of L5 is now made clear by noting that

$$E\left[\begin{pmatrix}\Sigma\mathbf{Y}_{t-1}U_t\\ \Sigma\mathbf{z}_tU_t\end{pmatrix}\right] = \mathbf{0},\ V\left[\begin{pmatrix}\Sigma\mathbf{Y}_{t-1}U_t\\ \Sigma\mathbf{z}_tU_t\end{pmatrix}\right] = E\left[\Sigma U_t^2\mathbf{J}_t\right] = \mathbf{C}_n.$$

An estimate of the parameter covariance matrix is usually required in order to calculate confidence regions or perform hypothesis tests. These may be based on asymptotic normality if this estimator is consistent. The obvious estimator for $\mathbf{D}_n^{-1}\mathbf{C}_n\mathbf{D}_n^{-1}$ is the sample analogue, $\mathbf{V}_{1n} = (\Sigma\mathbf{J}_t)^{-1}(\Sigma U_t^2\mathbf{J}_t)(\Sigma\mathbf{J}_t)^{-1}$, and, since $\mathbf{D}_n^{-1}\Sigma\mathbf{J}_t \to_p \mathbf{I}_{p+q}$ under L1 to L4, $\mathbf{V}_{1n} \sim_p \mathbf{D}_n^{-1}\mathbf{C}_n\mathbf{D}_n^{-1}$ whenever

$$\mathbf{D}_n^{-1}\Sigma\left(U_t^2\mathbf{J}_t - E\left[U_t^2\mathbf{J}_t\right]\right) \to_p \mathbf{0}.$$

This holds, for instance, in the case of strong homoscedasticity, when $E[U_t^2|\mathcal{F}_{t-1}] = \sigma^2$, for then $E[U_t^2\mathbf{J}_t] = \sigma^2E[\mathbf{J}_t]$, and the assertion reduces to $\mathbf{D}_n^{-1}(\Sigma\mathbf{J}_t - \mathbf{D}_n) \to_p \mathbf{0}$. However, in such special circumstances alternative estimators may be more convenient. Let $\mathbf{V}_{2n} = (\Sigma\mathbf{J}_t)^{-1}n^{-1}\Sigma U_t^2$, then, under L1 to L4 and (e) of Theorem 1, $\mathbf{V}_{2n} \sim_p n^{-1}\mathbf{D}_n^{-1}\Sigma\sigma_t^2$, so $\mathbf{V}_{2n}$ provides consistent estimation for less computation whenever

$$E\left[\mathbf{D}_n^{-1}\Sigma U_t^2\mathbf{J}_t\right] \div E\left[n^{-1}\Sigma U_t^2\right] \to 1;$$

strong homoscedasticity is one case in which this obtains.

5. Theorems 1 and 2. First we need to develop some basic formulae and this is facilitated by writing the model in the following matrix form (see Anderson, 1971, Section 5.3): let

$$\mathbf{B} = \begin{pmatrix}\beta_1 & \beta_2 & \cdots & \beta_p\\ \mathbf{I}_{p-1} & & & \mathbf{0}\end{pmatrix},\ \mathbf{A} = \begin{pmatrix}\boldsymbol{\alpha}^T\\ \mathbf{0}\end{pmatrix},\ \mathbf{U}_t = (U_t, 0, \cdots, 0)^T,$$

$\mathbf{B}$, $\mathbf{A}$, $\mathbf{U}_t$ being respectively $p \times p, p \times q, p \times 1$. Then (1.1) and (1.2) become

(5.1) $$\mathbf{Y}_t = \mathbf{B}\mathbf{Y}_{t-1} + \mathbf{A}\mathbf{z}_t + \mathbf{U}_t,\ t \geqslant 1,$$

(5.2) $$E\left[\mathbf{U}_t|\mathcal{F}_{t-1}\right] = \mathbf{0}.$$

It follows from (5.1) that

(5.3) $$\mathbf{Y}_t = \Sigma_{r=0}^{t-s}\mathbf{B}^r(\mathbf{U}_{t-r} + \mathbf{A}\mathbf{z}_{t-r}) + \mathbf{B}^{t-s+1}\mathbf{Y}_{s-1},\ 1 \leqslant s \leqslant t,$$

(5.4) $$\boldsymbol{\mu}_t = \Sigma_{r=0}^{t-s}\mathbf{B}^r\mathbf{A}\mathbf{z}_{t-r} + \mathbf{B}^{t-s+1}\boldsymbol{\mu}_{s-1},\ 1 \leqslant s \leqslant t,$$

and from (5.2) that

(5.5) $$E[\mathbf{U}_t\mathbf{U}_s^T] = \mathbf{0},\ E[\mathbf{U}_t\mathbf{Y}_s^T] = \mathbf{0}, \qquad s < t.$$

Subtract (5.4) from (5.3), postmultiply by $\mathbf{Y}_{s-1}^T$, and take expectations:

(5.6)

$$C[\mathbf{Y}_t, \mathbf{Y}_{s-1}] = E[(\mathbf{Y}_t - \boldsymbol{\mu}_t)(\mathbf{Y}_{s-1} - \boldsymbol{\mu}_{s-1})^T] = \mathbf{B}^{t-s+1}V[\mathbf{Y}_{s-1}], \qquad 1 \leqslant s \leqslant t.$$

Similarly, using (5.3) and (5.4) with $s = 1$,

(5.7) $$V[\mathbf{Y}_t] = \Sigma_{r=0}^{t-1}\mathbf{B}^r\begin{pmatrix} \sigma_{t-r}^2 & 0 \cdots 0 \\ \mathbf{0} & \end{pmatrix}(\mathbf{B}^T)^r + \mathbf{B}^tV[\mathbf{Y}_0](\mathbf{B}^T)^t$$

$$= \Sigma_{r=0}^{t-1}\sigma_{t-r}^2\mathbf{b}_r\mathbf{b}_r^T + \mathbf{B}^tV[\mathbf{Y}_0](\mathbf{B}^T)^t, \qquad t \geqslant 1,$$

where $\mathbf{b}_r$ is the first column of $\mathbf{B}^r$. The eigenvalues of $\mathbf{B}$ are the roots of the polynomial $x^p - \beta_1x^{p-1} - \cdots - \beta_p$ (see Anderson, 1971, Section 5.3), and all have moduli less than or equal to λ. The elements of $\mathbf{B}^r$ are then bounded above in modulus by $b\lambda^r$ for some $b < \infty$; for convenience take $b \geqslant 1$. It follows that $|\mathbf{B}^r| < pb\lambda^r$, and

(5.8) $$|\boldsymbol{\mu}_t| \leqslant pb\alpha\Sigma_{r=0}^{t-1}\lambda^r z_{t-r} + pb\lambda^t|\boldsymbol{\mu}_0| = pb\alpha\Sigma_{r=0}^{t}\lambda^r z_{t-r}, \qquad t \geqslant 0,$$

where $\alpha = |\mathbf{A}| = |\boldsymbol{\alpha}|$, $z_0 = |\boldsymbol{\mu}_0|/\alpha$; also

(5.9) $$|\{V[\mathbf{Y}_t]\}_{ij}| \leqslant |V[\mathbf{Y}_t]| \leqslant p^2b^2\Sigma_{r=0}^{t-1}\lambda^{2r}\sigma_{t-r}^2 + p^2b^2\lambda^{2t}|V[\mathbf{Y}_0]| = p^2b^2\nu_t,$$

and

(5.10) $$E[|\mathbf{Y}_t|^2] = \Sigma_{i=0}^{p-1}(V[Y_{t-i}] + \mu_{t-i}^2) = \text{trace } V[\mathbf{Y}_t] + |\boldsymbol{\mu}_t|^2$$

$$\leqslant p^3b^2\nu_t + p^2b^2\alpha^2(\Sigma_{r=0}^{t}\lambda^r z_{t-r})^2.$$

The $q \times q$ matrix $\mathbf{M}_n$ is symmetric, positive semidefinite and has smallest eigenvalue m_n. Thus

$$m_n = \inf \mathbf{f}^T\mathbf{M}_n\mathbf{f} = \inf \Sigma|\mathbf{f}^T\mathbf{z}_t|^2 \quad \text{(inf over unit vectors } \mathbf{f}\text{)},$$

and $m_{n+1} \geqslant m_n$. If $m_n = 0$ there exists $\mathbf{f} \neq \mathbf{0}$ such that $\mathbf{f}^T\mathbf{z}_t = 0$ for $t = 1, \cdots, n$ and $\mathbf{z}_1, \cdots, \mathbf{z}_n$ are linearly dependent. As a result any multiple of $\mathbf{f}$ may be added to $\boldsymbol{\alpha}$ in (1.1) without any detectable change in the model, i.e., $\boldsymbol{\alpha}$ is unidentifiable. We assume throughout that $m_n > 0$ for some n, and hence for all subsequent n, justifying the definitions in Section 2 of $\mathbf{G}_n$ and $\mathbf{E}_n$. In fact we need $m_n \to \infty$, which means that $\{\mathbf{z}_t\}$ must continue indefinitely to probe all q dimensions.

The proof of Theorem 1 is now given; we show that conditions (a) to (f) imply L1 to L4 of Theorem 4.

PROOF THAT (a) $\Rightarrow$ L1. We have

$$\mathbf{W}_{1n} = \Sigma\mathbf{Y}_{t-1}\mathbf{z}_t^T - \mathbf{G}_n^T\Sigma\mathbf{z}_t\mathbf{z}_t^T = \Sigma\mathbf{Y}_{t-1}\mathbf{z}_t^T - \mathbf{L}_n = \Sigma(\mathbf{Y}_{t-1} - \boldsymbol{\mu}_{t-1})\mathbf{z}_t^T,$$

so

$$E\left[|\mathbf{W}_{1n}|^2\right] = E\left[\operatorname{trace}(\mathbf{W}_{1n}\mathbf{W}_{1n}^T)\right] = \operatorname{trace} \Sigma_{s,t}\mathbf{z}_s^T\mathbf{z}_t C\left[\mathbf{Y}_{s-1}, \mathbf{Y}_{t-1}\right]$$
$$= \Sigma_t z_t^2 \operatorname{trace} V\left[\mathbf{Y}_{t-1}\right] + 2\Sigma_{s<t}\mathbf{z}_s^T\mathbf{z}_t \operatorname{trace}\left(\mathbf{B}^{t-s}V\left[\mathbf{Y}_{s-1}\right]\right).$$

Hence, using (5.9), and recalling that $|\{\mathbf{B}^r\}_{ij}| \leqslant b\lambda^r$ for $1 \leqslant b < \infty$,

$$E\left[|\mathbf{W}_{1n}|^2\right] \leqslant p^3b^2\Sigma_t z_t^2 v_{t-1} + 2p^4b^3\Sigma_{s<t}\mathbf{z}_s^T\mathbf{z}_t\lambda^{t-s}v_{s-1}$$
$$\leqslant 2p^4b^3\Sigma_{s\leqslant t}\mathbf{z}_s^T\mathbf{z}_t\lambda^{t-s}v_{s-1} = 2p^4b^3\Sigma\lambda^{t-1}\left(\Sigma_{r=0}^{n-t}v_r\mathbf{z}_{r+1}^T\mathbf{z}_{r+t}\right).$$

Now the elements of $\mathbf{E}_n^{-1}$ are bounded above in modulus by e_n^{-1}, so

$$E\left[|\mathbf{E}_n^{-1}\mathbf{W}_{1n}|^2\right] \leqslant |\mathbf{E}_n^{-1}|^2E\left[|\mathbf{W}_{1n}|^2\right] \leqslant p^2e_n^{-2}E\left[|\mathbf{W}_{1n}|^2\right].$$

Thus the condition

$$e_n^{-2}\Sigma\lambda^{t-1}\left(\Sigma_{r=0}^{n-t}v_r\mathbf{z}_{r+1}^T\mathbf{z}_{r+t}\right) \to 0$$

in (a) of Theorem 1 is sufficient for $\mathbf{E}_n^{-1}\mathbf{W}_{1n} \to_{ms} \mathbf{0}$ which implies $\mathbf{E}_n^{-1}\mathbf{W}_{1n} \to_p \mathbf{0}$ in L1 of Theorem 4. Likewise, the other conditions in (a) imply the limits in L1.

PROOF THAT (a) TO (f) $\Rightarrow$ L2. The approach is based on that of Anderson (1971, Section 5.5.6). We first seek conditions under which $e_n^{-1}\mathbf{W}_{2n} \to_p \mathbf{0}$ since this implies $\mathbf{E}_n^{-1}\mathbf{W}_{2n} \to_p \mathbf{0}$. From (5.1)

$$\mathbf{Y}_t\mathbf{Y}_t^T = \mathbf{B}\mathbf{Y}_{t-1}\mathbf{Y}_t^T + \mathbf{A}\mathbf{z}_t\mathbf{Y}_t^T + \mathbf{U}_t\mathbf{Y}_t^T,$$

and

$$\mathbf{B}\mathbf{Y}_{t-1}\mathbf{Y}_t^T = \mathbf{B}\mathbf{Y}_{t-1}\left(\mathbf{Y}_{t-1}^T\mathbf{B}^T + \mathbf{z}_t^T\mathbf{A}^T + \mathbf{U}_t^T\right),$$

so

$$\mathbf{Y}_t\mathbf{Y}_t^T - \mathbf{B}\mathbf{Y}_{t-1}\mathbf{Y}_{t-1}^T\mathbf{B}^T = \mathbf{A}\mathbf{z}_t\mathbf{Y}_t^T + \mathbf{B}\mathbf{Y}_{t-1}\mathbf{z}_t^T\mathbf{A}^T + \mathbf{U}_t\mathbf{Y}_t^T + \mathbf{B}\mathbf{Y}_{t-1}\mathbf{U}_t^T$$
$$= \left(\mathbf{A}\mathbf{z}_t\mathbf{Y}_{t-1}^T\mathbf{B}^T\right)_{+T} + \mathbf{A}\mathbf{z}_t\mathbf{z}_t^T\mathbf{A}^T + \left(\mathbf{A}\mathbf{Z}_t\mathbf{U}_t^T\right)_{+T} + \left(\mathbf{U}_t\mathbf{Y}_{t-1}^T\mathbf{B}^T\right)_{+T} + \mathbf{U}_t\mathbf{U}_t^T,$$

using $(\mathbf{C})_{+T}$ to denote $\mathbf{C} + \mathbf{C}^T$, the sum of the matrix and its transpose, for convenience. Sum over $1 \leqslant t \leqslant n$ and subtract expectations:

$$(5.11)\quad \begin{aligned}\left(\mathbf{W}_{2n} - \mathbf{B}\mathbf{W}_{2n}\mathbf{B}^T\right) &= \left(\mathbf{Y}_0\mathbf{Y}_0^T - E\left[\mathbf{Y}_0\mathbf{Y}_0^T\right]\right) - \left(\mathbf{Y}_n\mathbf{Y}_n^T - E\left[\mathbf{Y}_n\mathbf{Y}_n^T\right]\right)\\ &+ \left(\mathbf{A}\mathbf{W}_{1n}^T\mathbf{B}^T\right)_{+T} + \left(\mathbf{A}\Sigma\mathbf{z}_t\mathbf{U}_t^T\right)_{+T} + \left(\mathbf{B}\Sigma\mathbf{Y}_{t-1}\mathbf{U}_t^T\right)_{+T} + \Sigma\left(\mathbf{U}_t\mathbf{U}_t^T - E\left[\mathbf{U}_t\mathbf{U}_t^T\right]\right).\end{aligned}$$

If $e_n \to \infty$, as required in (b) for example, $e_n^{-1}(\mathbf{Y}_0\mathbf{Y}_0^T - E[\mathbf{Y}_0\mathbf{Y}_0^T]) \to_p \mathbf{0}$ (assuming that $\mathbf{Y}_0\mathbf{Y}_0^T$ has finite expectation), and under (a) $e_n^{-1}(\mathbf{A}\mathbf{W}_{1n}^T\mathbf{B}^T)_{+T} \to_{ms} \mathbf{0}$. Also, by (e)

$$e_n^{-1}\left|\Sigma\left(\mathbf{U}_t\mathbf{U}_t^T - E\left[\mathbf{U}_t\mathbf{U}_t^T\right]\right)\right| = e_n^{-1}\left|\Sigma\left(U_t^2 - \sigma_t^2\right)\right| \to_{ms} 0.$$

We must further consider

(b′) $$e_n^{-1}\Sigma\mathbf{z}_t\mathbf{U}_t^T \to_{ms} \mathbf{0},$$

(c′) $$e_n^{-1}\left(\mathbf{Y}_n\mathbf{Y}_n^T - E\left[\mathbf{Y}_n\mathbf{Y}_n^T\right]\right) \to \mathbf{0} \quad \text{in first mean,}$$

(d′) $$e_n^{-1}\Sigma\mathbf{Y}_{t-1}\mathbf{U}_t^T \to_{ms} \mathbf{0},$$

where the labels correspond to those in Theorem 1. Taking (b′) first we have

$$E\left[|\Sigma \mathbf{z}_t \mathbf{U}_t^T|^2\right] = \text{trace } \Sigma_{s,t} \mathbf{z}_s^T \mathbf{z}_t E\left[\mathbf{U}_s \mathbf{U}_t^T\right] = \Sigma z_t^2 \sigma_t^2,$$

so (b′) is implied by (b). For (c′) note that from (5.10)

$$E\left[|\mathbf{Y}_n \mathbf{Y}_n^T|\right] \leq E\left[|\mathbf{Y}_n|^2\right] \leq p^3 b^2 \nu_n + p^2 b^2 \alpha^2 (\Sigma_{r=0}^{n} \lambda^r z_{n-r})^2,$$

and so (c) ensures (c′). For (d′) we calculate

$$E\left[|\Sigma \mathbf{Y}_{t-1} \mathbf{U}_t^T|^2\right] = \text{trace } \Sigma_{s,t} E\left[\mathbf{Y}_{s-1} \mathbf{U}_s^T \mathbf{U}_t \mathbf{Y}_{t-1}^T\right]$$

$$= \text{trace } \Sigma E\left[U_t^2 \mathbf{Y}_{t-1} \mathbf{Y}_{t-1}^T\right] = \Sigma E\left[U_t^2 |\mathbf{Y}_{t-1}|^2\right].$$

The two alternative possibilities considered in (d) work out as follows. First, if $E[U_t^2|\mathcal{F}_{t-1}] \leq \sigma_{1t}^2$ a.s. then, using (5.10),

$$E\left[U_t^2|\mathbf{Y}_{t-1}|^2\right] \leq \sigma_{1t}^2 E\left[|\mathbf{Y}_{t-1}|^2\right] \leq p^3 b^2 \sigma_{1t}^2 \left\{\nu_{t-1} + \alpha^2 (\Sigma_{r=0}^{t-1} \lambda^r z_{t-1-r})^2\right\}$$

and so (d′) obtains if

$$e_n^{-2} \Sigma \sigma_{1t}^2 \left\{\nu_{t-1} + (\Sigma_{r=0}^{t-1} \lambda^r z_{t-1-r})^2\right\} \to 0.$$

Second, if the U_t's have finite fourth moments γ_t^4 then

$$E\left[U_t^2|\mathbf{Y}_{t-1}|^2\right] = E\left[U_t^2 |\Sigma_{r=0}^{t-2} \mathbf{B}^r (\mathbf{U}_{t-1-r} + \mathbf{A}\mathbf{z}_{t-1-r}) + \mathbf{B}^{t-1} \mathbf{Y}_0|^2\right]$$

$$= \text{trace } E\Big[\Sigma_{r,s} \mathbf{B}^r (\mathbf{U}_{t-1-r} + \mathbf{A}\mathbf{z}_{t-1-r})(\mathbf{U}_{t-1-s} + \mathbf{A}\mathbf{z}_{t-1-s})^T (\mathbf{B}^T)^s$$

$$+ 2\Sigma_r \mathbf{B}^r (\mathbf{U}_{t-1-r} + \mathbf{A}\mathbf{z}_{t-1-r}) \mathbf{Y}_0^T (\mathbf{B}^T)^{t-1} + \mathbf{B}^{t-1} \mathbf{Y}_0 \mathbf{Y}_0^T (\mathbf{B}^T)^{t-1}\Big]$$

$$\leq \Sigma_{r,s} p^2 b^2 \lambda^{r+s} \left\{\gamma_t^2 \gamma_{t-1-r} \gamma_{t-1-s} + 2q\alpha \gamma_t^2 \sigma_{t-1-r} z_{t-1-s} + q^2 \alpha^2 \sigma_t^2 z_{t-1-r} z_{t-1-s}\right\}$$

$$+ 2\Sigma_r p^2 b^2 \lambda^{r+t+1} \left\{\gamma_t^2 \gamma_{t-1-r} E\left[|\mathbf{Y}_0|^4\right]^{\frac{1}{4}} + q\alpha \gamma_t^2 z_{t-1-s} E\left[|\mathbf{Y}_0|^2\right]^{\frac{1}{2}}\right\}$$

$$+ p^2 b^2 \lambda^{2t-2} \gamma_t^2 E\left[|\mathbf{Y}_0 \mathbf{Y}_0^T|^2\right]^{\frac{1}{2}}$$

$$\leq p^2 b^2 \gamma_t^2 \left\{\Sigma_{r=0}^{t-2} \lambda^r (\gamma_{t-1-r} + q\alpha z_{t-1-r}) + \lambda^{t-1} E\left[|\mathbf{Y}_0|^4\right]^{\frac{1}{4}}\right\}^2.$$

It follows that (d′) holds if

$$e_n^{-2} \Sigma \gamma_t^2 \left\{\Sigma_{r=0}^{t-1} \lambda^r (\gamma_{t-1-r} + z_{t-1-r})\right\}^2 \to 0$$

where $\gamma_0 = E[|\mathbf{Y}_0|^4]^{\frac{1}{4}}$.

Returning now to (5.11) we have shown that (a) to (e) imply that

$$e_n^{-1}\left(\mathbf{W}_{2n} - \mathbf{B}\mathbf{W}_{2n}\mathbf{B}^T\right) \to \mathbf{0} \quad \text{in first mean.}$$

Premultiply this equation by $\mathbf{B}^i$, postmultiply by $(\mathbf{B}^T)^i$, and add the results over $i = 1, \cdots, r-1$. Then for each r

$$e_n^{-1}\left(\mathbf{W}_{2n} - \mathbf{B}^r \mathbf{W}_{2n} (\mathbf{B}^T)^r\right) \to \mathbf{0} \quad \text{in first mean.}$$

It follows that for each $\varepsilon > 0$ there exists $N(\varepsilon)$ such that

$$\varepsilon \geqslant E\left[e_n^{-1}|\mathbf{W}_{2n} - \mathbf{B}^r\mathbf{W}_{2n}(\mathbf{B}^T)^r|\right] \geqslant E\left[e_n^{-1}|\mathbf{W}_{2n}|\right](1 - |\mathbf{B}^r|^2)$$

for $n \geqslant N(\varepsilon)$. Under (f) $\lambda < 1$ and we can choose r sufficiently large so that $|\mathbf{B}^r|^2 < p^2b^2\lambda^{2r} < 1$ independently of n, whence $E[e_n^{-1}|\mathbf{W}_{2n}|] < \varepsilon$ for $n \geqslant N(\varepsilon)$ and $e_n^{-1}\mathbf{W}_{2n} \to \mathbf{0}$ in first mean.

For the other half of L2 it is sufficient that $g_n e_n^{-1}\mathbf{W}_{2n} \to 0$ in first mean, so each of the conditions above are also required to hold when multiplied by g_n.

PROOF THAT (b) $\Rightarrow$ L3 AND (d) $\Rightarrow$ L4. These have essentially been covered in the treatment of (b′) and (d′) above.

PROOF OF THEOREM 2. We verify that the extra conditions in Theorem 2, additional to those in Theorem 1, are sufficient for L5 of Theorem 4. Let X_t be defined as in Theorem 2 and $S_n = \Sigma X_t$. Since $E[X_t|\mathcal{F}_{t-1}] = 0$ a.s. S_n is a martingale. Write

$$s_n^2 = E\left[S_n^2\right] = \Sigma E\left[X_t^2\right] = \Sigma \mathbf{x}^T E\left[U_t^2 \mathbf{J}_t\right]\mathbf{x} = \mathbf{x}^T\mathbf{C}_n\mathbf{x}.$$

Various sets of conditions are given in Brown (1971) and Scott (1973) under which $S_n/s_n \to_d N(0, 1)$. Scott's set (A), comprising two limit conditions on the process $\{X_t^2\}$, is given in Theorem 2. The first limit is a stability requirement and the second is a Lindeberg-type condition. It follows by application of the Cramér-Wold device (1936) that in large samples the distribution of $\Sigma(\mathbf{Y}_{t-1}^T, \mathbf{z}_t^T)^T U_t$ is approximately normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{C}_n$.

6. Theorem 3: Stochastic explanatory variables. The results so far have been proved for a fixed sequence $\{\mathbf{z}_t\}$. If $\{\mathbf{z}_t\}$ is a stochastic process, and the conditions of Theorems 1 and 2 hold for all realizations which are likely to occur, it seems reasonable to expect the results to retain their validity in the wider context. The basic formulae of Section 5 have to be reinterpreted. All expectations are now taken conditionally on the z-process, and the working goes through as before provided we may use such properties as

$$E\left[U_t|\{\mathbf{z}_t\}\right] = 0,\ E\left[U_s U_t|\{\mathbf{z}_t\}\right] = \delta_{st}\sigma_t^2(\{\mathbf{z}_t\}) \text{ for } s \leqslant t.$$

For instance, (5.4) becomes

$$\boldsymbol{\mu}_t(\{\mathbf{z}_t\}) = E\left[\mathbf{Y}_t|\{\mathbf{z}_t\}\right] = \Sigma_{r=0}^{t-s}\mathbf{B}^r\mathbf{A}\mathbf{z}_{t-r} + \mathbf{B}^{t-s+1}\boldsymbol{\mu}_{s-1}(\{\mathbf{z}_t\}).$$

It is sufficient to make the rather natural assumption that the process $\{\mathbf{z}_t\}$ is independent of $\{U_t\}$. The limit results L1 and L4 persevere, now conditionally on $\{\mathbf{z}_t\}$, and we wish to make them unconditional. To take a specific example, we have shown in Section 5 that

$$e_n^{-2}\Sigma\lambda^{t-1}\left(\Sigma_{r=0}^{n-t}\nu_r\mathbf{z}_{r+1}^T\mathbf{z}_{r+t}\right) \to 0 \Rightarrow p_n(\varepsilon) \to 0\ \forall\varepsilon > 0$$

where $p_n(\varepsilon) = P[|\mathbf{E}_n^{-1}\mathbf{W}_{1n}| \geqslant \varepsilon|\{\mathbf{z}_t\}]$. Thus if

$$e_n^{-2}\Sigma\lambda^{t-1}\left(\Sigma_{r=0}^{n-t}\nu_r\mathbf{z}_{r+1}^T\mathbf{z}_{r+t}\right) \to_p 0$$

then $p_n(\varepsilon) \to_p 0$. But this implies that the unconditional probability

$$P\left[|\mathbf{E}_n^{-1}\mathbf{W}_{1n}| \geqslant \varepsilon\right] = E\left[p_n(\varepsilon)\right] \to 0,$$

by the dominated convergence theorem. Thus if $\{\mathbf{z}_t\}$ satisfies the condition "in probability" the limit result holds unconditionally. A similar argument applies to each of the results L1 to L4, and so, if the various limits hold in probability, the least-squares estimators are weakly consistent.

The argument leading to asymptotic normality requires no further modification. The martingale property, $E[X_t|\mathcal{F}_{t-1}] = 0$ a.s., is preserved because

$$E\left[U_t\mathbf{z}_t|\mathcal{F}_{t-1}\right] = 0 \text{ a.s.}$$

7. Lemmas 1, 2, and 3.

PROOF OF LEMMA 1. First note that $E[\Sigma(\mathbf{Y}_{t-1} - \mathbf{G}_n^T\mathbf{z}_t)\mathbf{z}_t^T] = E[\Sigma\mathbf{Y}_{t-1}\mathbf{z}_t^T - \mathbf{L}_n] = \mathbf{0}$. Therefore

$$\begin{aligned}\mathbf{E}_n = \mathbf{V}_n - \mathbf{L}_n\mathbf{M}_n^{-1}\mathbf{L}_n^T &= E\left[\Sigma\mathbf{Y}_{t-1}Y_{t-1}^T\right] - \mathbf{G}_n^T\Sigma\mathbf{z}_t\boldsymbol{\mu}_{t-1}^T\\ &= E\left[\Sigma(\mathbf{Y}_{t-1} - \mathbf{G}_n^T\mathbf{z}_t)(\mathbf{Y}_{t-1} - \mathbf{G}_n^T\mathbf{z}_t)^T\right],\end{aligned}$$

so $\mathbf{E}_n$ is the residual inner product matrix of $\{\mathbf{Y}_{t-1}\}$ after orthogonal projection onto $\{\mathbf{z}_t\}$. Thus

$$e_n = \inf \mathbf{f}^T\mathbf{E}_n\mathbf{f} = \inf \Sigma E\left[|\mathbf{f}^T(\mathbf{Y}_{t-1} - \mathbf{G}_n^T\mathbf{z}_t)|^2\right] \quad (\text{inf over } |\mathbf{f}| = 1).$$

Let

$$\mathbf{Y}_{0,t-1} = E\left[\mathbf{Y}_{t-1}|\mathcal{F}_0\right] = \Sigma_{r=0}^{t-2}\mathbf{B}^r\mathbf{A}\mathbf{z}_{t-1-r} + \mathbf{B}^{t-1}\mathbf{Y}_0, \qquad t \geqslant 2,$$

$$\mathbf{U}_{0,t-1} = \mathbf{Y}_{t-1} - \mathbf{Y}_{0,t-1} = \Sigma_{r=0}^{t-2}\mathbf{B}^r\mathbf{U}_{t-1-r} = \Sigma_{r=0}^{t-2}\mathbf{b}_rU_{t-1-r}, \qquad t \geqslant 2,$$

then

$$E\left[\mathbf{U}_{0,t-1}|\mathcal{F}_0\right] = \mathbf{0}\ a.s.,\ E\left[\mathbf{U}_{0,t-1}\mathbf{Y}_{0,t-1}^T\right] = \mathbf{0},$$

and

$$E\left[|\mathbf{f}^T(\mathbf{Y}_{t-1} - \mathbf{G}_n^T\mathbf{z}_t)|^2\right] = E\left[|\mathbf{f}^T\mathbf{U}_{0,t-1}|^2\right] + E\left[|\mathbf{f}^T(\mathbf{Y}_{0,t-1} - \mathbf{G}_n^T\mathbf{z}_t)|^2\right].$$

Now for $t \geqslant 2$,

$$E\left[|\mathbf{f}^T\mathbf{U}_{0,t-1}|^2\right] = E\left[\left(\Sigma_{r=0}^{t-2}\mathbf{f}^T\mathbf{b}_rU_{t-1-r}\right)^2\right] = \Sigma_{r=0}^{t-2}a_r^2\sigma_{t-r-1}^2$$

where $a_r = \mathbf{f}^T\mathbf{b}_r$, so

$$\Sigma_{t=2}^nE\left[|\mathbf{f}^T\mathbf{U}_{0,t-1}|^2\right] = \Sigma_{t=2}^n\Sigma_{r=0}^{t-2}\sigma_{t-1-r}^2a_r^2 = \Sigma_{t=2}^n\sigma_{t-1}^2(\Sigma_{r=0}^{n-t}a_r^2).$$

Also

$$E\left[|\mathbf{f}^T(\mathbf{Y}_{0,t-1} - \mathbf{G}_n^T\mathbf{z}_t)|^2\right] = \Sigma|\mathbf{f}^T\mathbf{G}_n^T(\mathbf{z}_t - \bar{\mathbf{z}}_n)|^2 + \Sigma E\left[|\mathbf{f}^T(\mathbf{Y}_{0,t-1} - E\left[\mathbf{Y}_{0,t-1}\right])|^2\right],$$

and

$$\Sigma|\mathbf{f}^T\mathbf{G}_n^T(\mathbf{z}_t - \bar{\mathbf{z}}_n)|^2 = \mathbf{f}^T\mathbf{G}_n^T\mathbf{M}_n'\mathbf{G}_n\mathbf{f}.$$

Thus we have

$$e_n = \inf\{E[|\mathbf{f}^T(\mathbf{Y}_0 - \mathbf{G}_n^T\mathbf{z}_1)|^2] + \Sigma_{t=2}^n \sigma_{t-1}^2(\Sigma_{r=0}^{n-t} a_r^2) + \mathbf{f}^T\mathbf{G}_n^T\mathbf{M}_n'\mathbf{G}_n\mathbf{f}$$
$$+ \Sigma E[|\mathbf{f}^T(\mathbf{Y}_{0,t-1} - E[\mathbf{Y}_{0,t-1}])|^2]\}.$$

Now

$$\Sigma_{t=2}^n \sigma_{t-1}^2(\Sigma_{r=0}^{n-t} a_r^2) \geqslant \Sigma_{t=2}^{n-p+1} \sigma_{t-1}^2(\Sigma_{r=0}^{p-1} a_r^2)$$

and, setting $a(\mathbf{f}) = \Sigma_{r=0}^{p-1} a_r^2$, the proof will be completed by showing that $a = \inf a(\mathbf{f}) > 0$ (inf over $|\mathbf{f}| = 1$). Let $\mathbf{f}^T = (f_1, \cdots, f_p)$. Then $a_0 = f_1$, $a_1 = f_1\beta_1 + f_2$, and generally we find that

$$a_r = f_{r+1} + \Sigma_{s=1}^r b_s f_s, \quad 1 \leqslant r \leqslant p-1,$$

where b_s is a sum of products of $\beta_1, \cdots, \beta_s$ of maximum degree s. Suppose that $a(\mathbf{f}) \leqslant \varepsilon^2$. The $|a_r| \leqslant \varepsilon$ for $0 \leqslant r \leqslant p-1$, and so $|f_1| \leqslant \varepsilon$ and $|f_2| \leqslant \varepsilon(1 + |\beta_1|) = c_2\varepsilon$, say. By induction, if $|f_s| \leqslant c_s\varepsilon$ for $s \leqslant r$,

$$|f_{r+1}| \leqslant |a_r| + \Sigma_{s=1}^r |b_s f_s| \leqslant \varepsilon + \Sigma_{s=1}^r |b_s| c_s \varepsilon = c_{r+1}\varepsilon,$$

say. Hence

$$|\mathbf{f}|^2 = \Sigma_{s=1}^p f_s^2 \leqslant \varepsilon^2(1 + \Sigma_{s=2}^p c_s^2).$$

But $|\mathbf{f}|^2 = 1$, so ε^2 cannot be smaller than $(1 + \Sigma_{s=2}^p c_s^2)^{-1}$, thus

$$a = \inf a(\mathbf{f}) \geqslant (1 + \Sigma_{s=2}^p c_s^2)^{-1} > 0.$$

Proof of Lemma 2. This follows immediately from (5.8):

$$g_n|\mathbf{M}_n^{-1}\mathbf{L}_n^T| \leqslant qm_n^{-1}\Sigma z_t|\boldsymbol{\mu}_{t-1}| \leqslant pqb\alpha m_n^{-1}\Sigma z_t(\Sigma_{r=0}^{t-1}\lambda^r z_{t-r-1}).$$

Lastly we discuss Lemma 3 which involves a stability property of the sequence $\{U_t^2\}$. Such laws of large numbers tend to hold when the sequence possesses a mixing property, i.e., dependence between terms diminishes with increasing distance. There are many convergence theorems for such cases, see e.g., Révész (1967, Section 8.2), Billingsley (1968, Section 20), Iosifescu and Theodorescu (1969, Section 1.1.3.2). However, the following result does not seem to have been placed on record previously. It should be more easily appreciated in applied work since the criterion involves correlations and deals with mean-square convergence, unlike most of the available theorems.

Lemma. *Consider the sequence* $\{\xi_t\}$ *of random variables and let* $\gamma_{st} = C[\xi_s, \xi_t]$, $\rho_{st} = \gamma_{st}/(\gamma_{ss}\gamma_{tt})^{\frac{1}{2}}$. *Then*

$$|V[\Sigma\xi_t] - \Sigma V[\xi_t]| \leqslant (a_n + b_n)\Sigma\gamma_{tt},$$

where $a_n = \max_{1 \leqslant i < n}\Sigma_{j=i+1}^n|\rho_{ij}|$, $b_n = \max_{1 < i \leqslant n}\Sigma_{j=1}^{i-1}|\rho_{ij}|$.

Note. This formula, involving correlations, is an analogue of Theorem 1.1.10 of Iosifescu and Theodorescu (1969) which employs the more restrictive "dependence coefficient."

PROOF. We have

$$|V[\Sigma\xi_t] - \Sigma V[\xi_t]| = 2|\Sigma_{s<t}\gamma_{st}| \leqslant \Sigma_{s<t}|\rho_{st}|(\gamma_{ss} + \gamma_{tt})$$
$$= \Sigma_{s=1}^{n-1}\gamma_{ss}(\Sigma_{t=s+1}^{n}|\rho_{st}|) + \Sigma_{t=2}^{n}\gamma_{tt}(\Sigma_{s=1}^{t-1}|\rho_{st}|) \leqslant a_n\Sigma_1^{n-1}\gamma_{ss} + b_n\Sigma_2^n\gamma_{tt} \leqslant (a_n + b_n)\Sigma\gamma_{tt}.$$

COROLLARY. *If* $|\rho_{st}| \leqslant \phi_{|s-t|}$ *and* $c_n^{-2}(1 + \Sigma\phi_t)\Sigma\gamma_{tt} \to 0$ *then*

$$c_n^{-1}\Sigma(\xi_t - E[\xi_t]) \to_{ms} 0.$$

PROOF. First

$$a_n \leqslant \max_{1\leqslant i<n}\Sigma_{j=i+1}^{n}\phi_{j-i} = \max_{1\leqslant i<n}\Sigma_{j=1}^{n-i}\phi_j = \Sigma_{j=1}^{n-1}\phi_j \qquad \text{and} \qquad b_n \leqslant \Sigma_{j=1}^{n-1}\phi_j.$$

Therefore

$$|V[\Sigma\xi_t] - \Sigma V[\xi_t]| \leqslant 2(\Sigma\phi_j)(\Sigma\gamma_{tt}),$$

so

$$V[\Sigma\xi_t] \leqslant (1 + 2\Sigma\phi_j)\Sigma\gamma_{tt},$$

and the result follows immediately.

For Lemma 3 the corollary is to be applied with $\xi_t = U_t^2$ and many practical situations will be covered by the following:

(a) $\sigma_t^2 = 0(1)$, so that $e_n \geqslant 0(n)$ by Lemma 1;

(b) $\gamma_{tt} = 0(1)$, so $\Sigma\gamma_{tt} = 0(n)$;

(c) $|\rho_{st}| \leqslant \phi_{|s-t|}$, and $\phi_j \to 0$ as $j \to \infty$, so $n^{-1}\Sigma\phi_j \to 0$.

Such cases satisfy the condition of Lemma 3.

APPENDIX

We give here an outline of the calculations for Example 4 of Section 2. We have to justify $e_n = 0(n^3)$, $g_n = 0(1)$, $\mathbf{M}_n^{-1}\mathbf{W}_{1n}^T \to_p \mathbf{0}$ and $\mathbf{M}_n^{-1}\mathbf{W}_{3n} \to_p \mathbf{0}$. From (5.4), with $p = 1$ and $\mathbf{z}_t = (1, t)^T$,

$$\mu_t = \Sigma_{r=0}^{t-1}\beta^r[\alpha_1 + \alpha_2(t - r)] + \beta^t\mu_0 = \mu_0\beta^t + \frac{\alpha_1}{1-\beta}(1 - \beta^t)$$
$$+ \frac{\alpha_2}{(1-\beta)^2}[t(1-\beta) - \beta(1-\beta^t)].$$

We can now evaluate $\Sigma_1^n\mu_t$, $\Sigma_1^n t\mu_t$ and hence $\mathbf{L}_n = (\mu_0 + \Sigma_1^{n-1}\mu_t, \mu_0 + \Sigma_1^{n-1}(t + 1)\mu_t)$. Then, after some further algebra, $\mathbf{G}_n = \mathbf{M}_n^{-1}\mathbf{L}_n^T$ is found to have both elements of $0(1)$, so $g_n = 0(1)$. Also

$$\mathbf{M}_n' = \mathbf{M}_n - n\bar{\mathbf{z}}_n\bar{\mathbf{z}}_n^T = \begin{pmatrix} 0 & 0 \\ 0 & n(n^2-1)/12 \end{pmatrix}$$

and so, by Lemma 1,

$$e_n \geqslant \mathbf{G}_n^T\mathbf{M}_n'\mathbf{G}_n = \tfrac{1}{12}n(n^2 - 1)\{\mathbf{G}_n\}_2^2 = 0(n^3).$$

Now

$$\mathbf{M}_n^{-1}\mathbf{W}_{3n} = \begin{pmatrix} 0(n^{-1})\Sigma U_t + 0(n^{-2})\Sigma tU_t \\ 0(n^{-2})\Sigma U_t + 0(n^{-3})\Sigma tU_t \end{pmatrix}$$

and

$$E\left[(\Sigma U_t)^2\right] = \Sigma\sigma_t^2 = 0(n),\ E\left[(\Sigma tU_t)^2\right] = \Sigma t^2\sigma_t^2 = 0(n^3)$$

so $\mathbf{M}_n^{-1}\mathbf{W}_{3n} \to_{ms} \mathbf{0}$. Finally we consider $\mathbf{M}_n^{-1}\mathbf{W}_{1n}^T \to_p \mathbf{0}$. We have

$$E\left[|\mathbf{M}_n^{-1}\mathbf{W}_{1n}^T|^2\right] = \operatorname{trace}\left(\mathbf{M}_n^{-1}E\left[\mathbf{W}_{1n}^T\mathbf{W}_{1n}\right]M_n^{-1}\right),$$

and, using similar calculations to those in Section 5 based on (5.6) and (5.7),

$$E\left[\mathbf{W}_{1n}^T\mathbf{W}_{1n}\right] = \Sigma_{s,t}\mathbf{z}_s\mathbf{z}_t^TC\left[Y_{s-1}, Y_{t-1}\right] = \begin{pmatrix} 0(n) & 0(n^2) \\ 0(n^2) & 0(n^3) \end{pmatrix}.$$

It can now be verified that $\mathbf{M}_n^{-1}E[\mathbf{W}_{1n}^T\mathbf{W}_{1n}]\mathbf{M}_n^{-1} \to \mathbf{0}$, whereas $m_n^{-2}E[|\mathbf{W}_{1n}|^2] \to \infty$.

Acknowledgments. I would like to thank the referee for constructive comments on the first draft of this paper.

NOTE. Since this paper was accepted I have been shown one by Anderson and Taylor (1979), shortly to appear, which has some degree of overlap with the material here. The main differences are that they (i) treat the multivariate case (Y_t a vector in (1.1)), (ii) consider *strong* consistency but not asymptotic normality of the least-squares estimate, and (iii) make assumptions of (a) strong homoscedasticity on the error sequence, and (b) comparability of the orders of magnitude of the eigenvalues of $\mathbf{M}_n$.

REFERENCES

[1] ANDERSON, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Ann. Math. Statist.* **30** 676–687.

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

[3] ANDERSON, T. W. and TAYLOR, J. B. (1979). Strong consistency of least squares estimates in dynamic models. *Ann. Statist.* **7** 484–489.

[4] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[5] BROWN, B. M. (1971). Martingale central limit theorems. *Ann. Math. Statist.* **42** 59–66.

[6] CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.

[7] CRAMÉR, H. and WOLD, H. (1936). Some theorems on distribution functions. *J. London Math. Soc.* **11** 290–295.

[8] CROWDER, M. J. and GROB, P. R. (1975). A logit model for infectious diseases. *Appl. Statist.* **24** 85–94.

[9] IOSIFESCU, M. and THEODORESCU, R. (1969). *Random Processes and Learning.* Springer-Verlag, Berlin.

[10] RÉVÉSZ, P. (1968). *The Laws of Large Numbers.* Academic Press, New York.

[11] SCOTT, D. J. (1973). Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. *Advances Appl. Probability* **5** 119–137.

DEPARTMENT OF MATHEMATICS
SURREY UNIVERSITY
GUILDFORD, SURREY GU2 5XH
ENGLAND