The Annals of Statistics
1995, Vol. 23, No. 3, 1048–1072

LOG-PERIODOGRAM REGRESSION OF TIME SERIES WITH LONG RANGE DEPENDENCE[1]

BY P. M. ROBINSON

London School of Economics

This paper discusses the estimation of multiple time series models which allow elements of the spectral density matrix to tend to infinity or zero at zero frequency and be unrestricted elsewhere. A form of log-periodogram regression estimate of differencing and scale parameters is proposed, which can provide modest efficiency improvements over a previously proposed method (for which no satisfactory theoretical justification seems previously available) and further improvements in a multivariate context when differencing parameters are a priori equal. Assuming Gaussianity and additional conditions which seem mild, asymptotic normality of the parameter estimates is established.

1. Introduction. In the analysis of stationary time series the behaviour of the spectral density around zero frequency is often of interest. For a scalar covariance stationary process X_t, $t = 0, \pm 1, \dots,$ assume absolute continuity of the spectral distribution function, so there is a spectral density $f(\lambda)$, $-\pi < \lambda \le \pi$, such that the autocovariance

$$\gamma_j = E\big[(X_1 - E(X_1))(X_{1+j} - E(X_1))\big] = \int_{-\pi}^{\pi} \cos(j\lambda) f(\lambda)\, d\lambda.$$

Many time series models imply $0 < f(0) < \infty$, but empirical observation is sometimes consistent with the possibilities that either $f(0) = \infty$ or $f(0) = 0$. These can be examined by means of a model of the form

$$f(\lambda) \sim C\lambda^{-2d} \quad \text{as } \lambda \to 0+, \tag{1.1}$$

where "$\sim$" means that the ratio of left- and right-hand sides tends to 1, C is positive and finite and $-\frac{1}{2} < d < \frac{1}{2}$ because for $d \ge \frac{1}{2}$ a function behaving like λ^{-2d} as $\lambda \to 0+$ will not be integrable so that covariance stationarity cannot obtain, while $d > -\frac{1}{2}$ corresponds to an invertibility condition in parametric models with property (1.1). When $d > 0$, X_t is often said to have "long range dependence." Property (1.1) is also useful in the modeling of nonstationary processes. If there is a somewhat greater degree of nonstationarity in the process than implied by a unit root, (1.1) with $d > 0$ could be a useful model for first differences, while $d < 0$ can model the first differences of a process which is nonstationary, but less so than a unit root process. Two

Received June 1992; revised March 1994.
[1] Research supported by ESRC Grant R000233609.
AMS 1991 *subject classifications*. Primary 62M10; secondary 60G18, 62G05.
Key words and phrases. Long range dependence, log-periodogram regression, least squares, generalized least squares, limiting distribution theory.

parametric models for $f(\lambda)$ over all frequencies which imply (1.1) are the "fractional noise" spectral density

$$f(\lambda) = \frac{4\sigma^2\Gamma(2d)}{(2\pi)^{3+2d}}\cos(\pi d)\sin^2\left(\frac{1}{2}\lambda\right)\sum_{j=-\infty}^{\infty}\left|j + \left(\frac{\lambda}{2\pi}\right)\right|^{-2-2d}, \quad -\pi < \lambda \le \pi, \tag{1.2}$$

and the fractional autoregressive integrated moving average (ARIMA) spectral density

$$f(\lambda) = \frac{\sigma^2}{2\pi}|1 - e^{i\lambda}|^{-2d}\left|\frac{a(e^{i\lambda})}{b(e^{i\lambda})}\right|^2, \qquad \pi < \lambda \le \pi, \tag{1.3}$$

where $a(\cdot)$ and $b(\cdot)$ are polynomials, all of whose roots lie outside the unit circle. A recent review of the long range dependence literature is Robinson (1994c).

In applications, C and d are generally unknown. Introduce the discrete Fourier transform and periodogram

$$w(\lambda) = (2\pi n)^{-1/2}\sum_{t=1}^{n} X_t e^{it\lambda}, \qquad I(\lambda) = |w(\lambda)|^2, \tag{1.4}$$

and also

$$\lambda_j = 2\pi j/n,\ a_j = -\log\big(4\sin^2(\lambda_j/2)\big),$$

$$\bar{a} = (m-l)^{-1}\Sigma_{j=l+1}^{m} a_j,\ S_l = \Sigma_{j=l+1}^{m}(a_j - \bar{a})^2$$

and

$$\bar{d}(l) = \sum_{j=l+1}^{m}(a_j - \bar{a})\log I(\lambda_j)/S_l, \qquad 0 \le l < m < n,$$

which is the least squares estimate of d in the "regression model" given by the identity

$$\log I(\lambda_j) = c + da_j + U_j, \qquad j = l+1,\dots,m,$$

where $c = \log C - \eta$, η is Euler's constant $\eta = 0.5772\dots$ and $U_j = \log\{(4\sin^2(\lambda_j/2))^d I(\lambda_j)/C\} + \eta$. Geweke and Porter-Hudak (1983) made the important contribution of introducing $\bar{d}(0)$. They assumed, as in (1.1), that $f(\lambda)$ can be approximated by $C(4\sin^2(\lambda/2))^{-d}$ in only a neighbourhood of zero frequency, so an asymptotic theory will require that m tend to infinity more slowly than n. When a finite-parameter model for the spectrum across all frequencies can be accurately specified, $\bar{d}(0)$ is bound to be asymptotically less efficient than a Gaussian maximum likelihood estimate of d; indeed, its relative efficiency will be zero. Now that Gaussian parametric estimates for long range dependent time series models have been rigorously justified by Fox and Taqqu (1986), Giraitis and Surgailis (1990) and others, they thus provide an attractive alternative. However, if the parametric model has been misspecified, these estimates will be inconsistent, whereas if both autoregressive and moving average orders in the fractional ARIMA are overstated, there is a loss of identifiability. So far as computation is concerned, the Gaussian parametric methods require software that is not very widely available, whereas least squares is extremely familiar. At least when n is large, the weak model assumptions apparently underlying $\bar{d}(0)$ make it rather appealing, and it has been frequently used in empirical work [for applications of this and related estimates, see, e.g., Cheung (1993), Cheung and Lai (1993), Diebold and Rudebusch (1989), Geweke and Porter-Hudak (1983), Porter-Hudak (1990), Shea (1991) and Tschernig (1993)]. A rival semiparametric estimate was recently proposed by Robinson (1994a) and shown to be consistent under very mild conditions, but its limiting distributional theory is rather complicated [Lobato and Robinson (1993)].

Geweke and Porter-Hudak (1983) attempted a proof of asymptotic properties of $\bar{d}(0)$ in the case $-\frac{1}{2} < d < 0$, only, and the following theorem has implications for their method of proof. In the theorem, (1.1) is augmented by a related [see Yong (1974)] condition on γ_j. The theorem merely extends an earlier result of Künsch (1986), and we omit the straightforward proof.

THEOREM 1. *Let* (1.1) *hold and* $\gamma_j \sim C b_d j^{2d-1}$ *as* $j \to \infty$, *where* $b_d = 2\Gamma(1-2d)\sin \pi d$. *Then for* $0 < |d| < \frac{1}{2}$ *and any fixed positive or nonnegative integers* j *and* k,

(1.5)
$$\lim_{n\to\infty} E\left\{\frac{I(\lambda_j)}{f(\lambda_j)}\right\} = \frac{b_d |j|^{2d}}{(2\pi)^{1-2d}} \times \left\{4\int_0^1 u^{2d-1}(u-1)\sin^2(\pi j u)\,du + \frac{1}{d(2d+1)}\right\},$$

and, defining

$$P_d(j,k) = \frac{-2b_d|jk|^d}{(2\pi)^{1-2d}(j+k)}\int_0^1 u^{2d-1}\{\sin(2\pi j u) + \sin(2\pi k u)\}\,du, \qquad j+k \neq 0,$$

we have

$$\lim_{n\to\infty} E\left\{\frac{w(\lambda_j)w(\lambda_k)}{f(\lambda_j)^{1/2} f(\lambda_k)^{1/2}}\right\} = P_d(j,k), \tag{1.6}$$

and, if X_t *is Gaussian*,

$$\lim_{n\to\infty} \mathrm{Var}\{\frac{I(\lambda_j)}{f(\lambda_j)}\} = 2P_d^2(j,j), \tag{1.7}$$

$$\lim_{n\to\infty} \mathrm{Cov}\{\frac{I(\lambda_j)}{f(\lambda_j)}, \frac{I(\lambda_k)}{f(\lambda_k)}\} = P_d^2(j,k) + P_d^2(j,-k), \qquad j \pm k \neq 0. \tag{1.8}$$

In the case $0 < d < \frac{1}{2}$, Künsch (1986) previously obtained a formula which (after division by 2π) is equivalent to (1.5), leading him to propose $\bar{d}(l)$ for $l > 0$, though he did not establish its asymptotic distribution. Künsch's important observations were noted by Smith (1989, 1993), but have been overlooked in much of the literature. [Smith (1993), Delgado and Robinson (1994) and Lee and Robinson (1993) carried out empirical applications of $\bar{d}(l)$ and related estimates for $l > 0$].

Theorem 1 does not imply that choosing $l > 0$ is essential in order to achieve good asymptotic properties of $\bar{d}(l)$, though it suggests that this may be a desirable practical policy. The intuition behind $\bar{d}(0)$ is based on a belief that the U_j are approximately uncorrelated and homoscedastic. However, because (1.5) and (1.7) vary with j, while (1.8) is nonzero, the theorem implies that for both $-\frac{1}{2} < d < 0$ and $0 < d < \frac{1}{2}$, the normalized periodograms $I(\lambda_j)/f(\lambda_j)$, and thus the U_j, are asymptotically neither independent nor identically distributed when $n \to \infty$ but j stays fixed. To justify $\bar{d}(0)$, Geweke and Porter-Hudak (1983) argued [as a preliminary to applying the Lindeberg–Feller central limit theorem (CLT)] that for fixed m, the U_j, $j = 1, \dots, m$, are approximately iid to any specified degree of accuracy for n large enough when $-\frac{1}{2} < d < 0$. Having met this accuracy criterion, m is replaced by $m + 1$ and n is increased suitably, and so on, leading to a sequence such that as $n \to \infty$, the effect of replacing U_j, $j = 1, \dots, m$, by an iid triangular array (call it $\tilde{U}_j$, $j = 1, \dots, m$) is asymptotically negligible. However, Theorem 1 implies that we cannot choose a finite m to initiate this sequence. Part of Theorem 1 rests on the Gaussianity assumption, but the asymptotic non-iid property is likely to hold much more generally; indeed (1. 7) and (1.8) are true when $0 < d < \frac{1}{2}$ and X_t is instead fourth-order stationary with $\Sigma\Sigma\Sigma_{j,k,l=-\infty}^{\infty}|\mathrm{cum}(X_0, X_j, X_k, X_l)| < \infty$. Independently, and at around the same time as our work, Hurvich and Beltrao (1993, 1994) obtained results corresponding to Theorem 1, though theirs are presented in a different form.

Apparently independently of Künsch's (1986) earlier work, $\bar{d}(0)$ was recently criticised by Agiakloglou, Newbold and Wohar (1993) in the case $0 < d < \frac{1}{2}$. For $d < 0$, Hassler (1993b) [correcting arguments in Hassler (1993a)] claimed to have shown that if X_t is Gaussian, $S_l^{1/2}(\bar{d}(l) - d) \to_d N(0, \pi^2/6)$ for suitable $l > 0$. Hassler noted that for suitable $\lambda > 0$, $E\{I(\lambda)/f(\lambda) - J(\lambda)\}^2 = O(n^{-\beta})$ for some $\beta > 0$, where the $J(\lambda_j)$, $0 < j < n/2$, the periodograms of standard normal white noise innovations of X_t under model (1.3), are exactly iid for all n. Hassler claimed that these properties imply that the $I(\lambda_j)/f(\lambda_j)$, $l < j < n/2$, converge to iid variates, for l increasing suitably with n, and thence that the corresponding U_j can be replaced by the iid $\tilde{U}_j$ in $S_l^{1/2}(\bar{d}(l) - d)$, to give $S_l^{-1/2}\Sigma_{j=l+1}^{m}(a_j - \bar{a})\tilde{U}_j$, which tends to a $N(0, \pi^2/6)$ variate by a straightforward application of the Lindeberg–Feller CLT. However, the difference between $S_l^{1/2}(\bar{d}(l) - d)$ and $S_l^{-1/2}\Sigma_{j=l+1}^{m}(a_j - \bar{a})\tilde{U}_j$ is $S_l^{-1/2}\Sigma_{j=l+1}^{m}(a_j - \bar{a})(U_j - \tilde{U}_j)$, and Hassler's arguments do not appear to amount to a proof that the latter quantity is $o_p(1)$, especially bearing in mind that $S_l^{-1/2}$ decreases only at rate $m^{-1/2}$. We thus disagree with Hassler's claim to have verified the limiting distribution of $\bar{d}(l)$ [and for similar reasons, his claim to have completed Kashyap and Eom's (1988) proof of asymptotic normality of a parametric log-periodogram regression estimate].

Yajima (1989) gave useful results for the limiting joint distribution of $w(\lambda)$ at finitely many fixed λ, but not at an increasing number of λ_j on an interval which degenerates to zero as $n \to \infty$, so these results do not form a basis for theory for $\bar{d}(l)$ with $m \to \infty$. Despite the theoretical and practical interest that has been shown in $\bar{d}(l)$, no satisfactory asymptotic distribution theory seems to be available for it, for any value of d. The present paper attempts to provide one. We deal simultaneously with $0 < d < \frac{1}{2}$ and $-\frac{1}{2} < d < 0$ (and $d = 0$), unusually in the asymptotic theory of series satisfying (1.1), and correcting an impression which might have been conveyed, that a proof for $d < 0$ is relatively easy, while for $d > 0$ it is much harder. Our theory also covers estimates of the scale parameter C in (1.1). We in fact propose a more general class of estimates of C and d, including ones which have modestly superior asymptotic efficiency to $\bar{d}(0)$. We treat directly a multivariate model, and also show that further improvements are possible when differencing parameters are known to be common to more than one series. On the other hand, our work provides justification for tests that different time series share a common differencing parameter. A large-sample theory similar to ours may be useful in completing proofs for related estimates of Janacek (1982) and Kashyap and Eom (1988), and efficiency improvements similar to ours can be made to these estimates.

The regularity conditions we impose are in one sense extremely strong and in others, rather weak. We assume Gaussianity of the time series, finding this very helpful because of the complicated and nonlinear way in which parameter estimates depend on the data, but conjecture that a limit distribution theory can be obtained under more general distributional assumptions. A desirable side effect of the Gaussianity assumption is that further assumptions on the time series can be expressed purely in terms of the spectral density (or spectral density matrix) f, and our assumptions on f seem weak. On a neighbourhood of zero frequency, we make stronger assumptions on f than (1.1), but we impose no restrictions on f elsewhere, apart from the integrability necessary for covariance stationarity, and in this respect our conditions are milder than those conjectured by Geweke and Porter-Hudak (1983).

The following section introduces the multivariate model, modified and improved estimates and related test statistics. Regularity conditions are introduced in Section 3, where asymptotic results are presented. Proofs of these conditions appear in Sections 4 and 5.

2. Multivariate semiparametric model, estimates and test statistics. Let X_t now represent a G-dimensional vector with gth element X_{gt}, $g = 1, \dots, G$. We assume that X_t has a spectral density matrix given by $E\{(X_1 - E(X_1))(X_{1+j} - E(X_1))'\} = \int_{-\pi}^{\pi} e^{ij\lambda} f(\lambda)\, d\lambda$. The (g, h)th element of $f(\lambda)$, the cross spectral density of X_{gt} and X_{ht} when $g \neq h$, is denoted $f_{gh}(\lambda)$. The gth diagonal element of $f(\lambda)$, $f_{gg}(\lambda)$, is the power spectral density of

X_{gt}. For C_g, d_g satisfying $0 < C_g < \infty$, $-\frac{1}{2} < d_g < \frac{1}{2}$, we assume that

$$f_{gg}(\lambda) \sim C_g \lambda^{-2d_g} \quad \text{as } \lambda \to 0+, \tag{2.1}$$

for $g = 1, \dots, G$.

We can motivate this model as follows. Let ξ_t be a G-dimensional zero-mean unobservable stationary process having spectral density matrix which is continuous and nonsingular at the origin, and

$$X_t = E(X_t) + \sum_{j=-\infty}^{\infty} A_j \xi_{t-j}, \tag{2.2}$$

where the A_j are $G \times G$ matrices. Denote by $\alpha_g(\lambda)'$ the gth row of $A(\lambda) = \sum_{j=-\infty}^{\infty} A_j e^{ij\lambda}$ and suppose that for some $d_g \in (-\frac{1}{2}, \frac{1}{2})$, $\alpha_g(\lambda)\lambda^{d_g}$ tends to a finite, nonnull vector as $\lambda \to 0$, for $g = 1, \dots, G$. Then (2.1) holds. As an example, take ξ_t to be a stationary and invertible vector ARMA process and let each X_{gt} be formed by separate fractional differencing of the corresponding ξ_t element, so that

$$A(\lambda) = \operatorname{diag}\left\{(1 - e^{i\lambda})^{-d_1}, \dots, (1 - e^{i\lambda})^{-d_G}\right\}. \tag{2.3}$$

The periodogram of X_{gt}, $t = 1, \dots, n$, is denoted by

$$I_g(\lambda) = (2\pi n)^{-1}\left|\sum_{t=1}^{n} X_{gt} e^{it\lambda}\right|^2, \qquad g = 1, \dots, G.$$

Let J be a given integer greater than or equal to 1. Define

$$Y_{gk}^{(J)} = \log\left\{\sum_{j=1}^{J} I_g(\lambda_{k+j-J})\right\}, \qquad g = 1, \dots, G,\ k = l + J, l + 2J, \dots, m.$$

In the asymptotic theory l and m both tend to infinity with n, but more slowly, while $l/m \to 0$ also. We have for simplicity assumed here that $m - l$ is a multiple of J, but because J is held fixed as $n \to \infty$, the end effects in case this is untrue when $J > 1$ are easily shown to exert a negligible influence on asymptotic properties. Whatever the value of J, the $Y_{gk}^{(J)}$ for the given values of k use all the $I_g(\lambda_j)$, $j = l + 1, \dots, m$, once, but when $J > 1$, there is a pooling of contributions from adjacent frequencies.

Define the unobservable random variables $U_{gk}^{(J)}$ by

$$\begin{aligned} Y_{gk}^{(J)} &= c_g^{(J)} - d_g(2\log\lambda_k) + U_{gk}^{(J)}, \\ h &= 1, \dots, G,\ k = l + J, l + 2J, \dots, m, \end{aligned} \tag{2.4}$$

where $c_g^{(J)} = \log C_j + \psi(J)$ and ψ is the digamma function, $\psi(z) = (d/dz)\log\Gamma(z)$, where Γ is the gamma function. Now if the C_g, d_g, $g = 1, \dots, G$, are all functionally unrelated parameters and if one pretends that the vectors $U_k^{(J)} = (U_{1k}^{(J)}, \dots, U_{Gk}^{(J)})'$ are uncorrelated and homoscedastic with zero mean, it is well known that least squares estimation of the G equations (2.4) provides the best linear unbiased estimates of the $c_g^{(J)}, d_g$, even in the presence of correlation between elements of $U_k^{(J)}$. Of course the $U_k^{(J)}$ do not have the former properties, but we will show that the least squares estimates have the same limiting distributional behaviour as if such properties held.

The least squares estimates of $c^{(J)} = (c_1^{(J)}, \ldots, c_G^{(J)})'$ and $d = (d_1, \ldots, d_G)'$ are $\tilde{c}^{(J)}$ and $\tilde{d}^{(J)}$, given by

$$\begin{bmatrix} \tilde{c}^{(J)} \\ \tilde{d}^{(J)} \end{bmatrix} = \operatorname{vec}\left\{Y^{(J)\prime}Z^{(J)}\left(Z^{(J)\prime}Z^{(J)}\right)^{-1}\right\}, \tag{2.5}$$

where $Z^{(J)} = (Z_{l+J}, Z_{l+2J}, \ldots, Z_m)'$, $Y^{(J)} = (Y_1^{(J)}, \ldots, Y_G^{(J)})$, $Z_K = (1, -2\log\lambda_k)'$ and $Y_g^{(J)} = (Y_{g,l+J}^{(J)}, Y_{g,l+2J}^{(J)}, \ldots, Y_{g,m}^{(J)})'$. In the case $J = 1$, $\tilde{d}^{(J)}$ is $\bar{d}(l)$ apart from the multivariate generalization and our use of $-2\log\lambda_k$ as a regressor in place of $-\log(4\sin^2\lambda_k/2)$. We have stressed $-2\log\lambda_k$ owing to its slightly greater notational and computational simplicity, and it is easily shown that the same results go through if $-\log(4\sin^2\frac{1}{2}\lambda_k)$ is used.

Denoting the gth element of $\tilde{c}^{(J)}$ by $\tilde{c}_g^{(J)}$, we can estimate C_g by $\tilde{C}_g^{(J)} = \exp(\tilde{c}_g^{(J)} - \psi(J))$. We can also compute diagnostic statistics associated with least squares estimates. The residual vectors are $\tilde{U}_k^{(J)} = Y_k^{(J)} - \tilde{c}^{(J)} + \tilde{d}^{(J)}(2\log\lambda_k)$, $k = l+J, l+2J, \ldots, m$, and the matrix of sample variances and covariances is $\tilde{\Omega}^{(J)} = \{J/(m-l)\}\sum_k' \tilde{U}_k^{(J)}\tilde{U}_k^{(J)\prime}$, where the primed sum is over $k = l+J, l+2J, \ldots, m$. A standard error for $\tilde{d}_g^{(J)}$, the gth element of $\tilde{d}^{(J)}$, is given by the square root of the $(g+G)$th diagonal element of $(Z^{(J)\prime}Z^{(J)})^{-1} \otimes \tilde{\Omega}^{(J)}$. Consider the homogeneous restriction

$$H_0\colon Pd = 0, \tag{2.6}$$

where P is an $H \times G$ matrix of rank $H < G$. For example, we might assert that one or more d_g are zero, to indicate the absence of long range dependence, or that two or more of the d_g are equal. A statistic for testing (2.6) is

$$\tilde{d}^{(J)\prime}P'\left[(0, P)\left\{\left(Z^{(J)\prime}Z^{(J)}\right)^{-1} \otimes \tilde{\Omega}^{(J)}\right\}(0, P)'\right]^{-1} P\tilde{d}^{(J)}. \tag{2.7}$$

In the case $H = 1$, a corresponding one-sided statistic is constructed in the obvious way.

We can incorporate the assumption that some or all the G series share a common d-parameter. An analogous assumption is taken for granted in much of the unit root literature. There is a large literature concerning multivariate processes with $d = 0$ and one might respond to nonrejection of a hypothesis of equality of d-parameters (following a test such as that described in the previous paragraph) by exploiting this information in the parameter estimation. By suitably pooling information from different series with common d-parameter, we could expect to obtain estimates which are asymptotically more efficient than $\tilde{c}^{(J)}$ and $\tilde{d}^{(J)}$. Impose the restrictions

$$d = Q\theta, \tag{2.8}$$

where Q is a given $G \times K$ matrix of rank $K < G$ and θ is a K-dimensional column vector of functionally unrelated parameters. Any set of $G - K$ homogeneous linear restrictions on d can be so represented. Consider

$$\begin{bmatrix} \hat{c}^{(J)} \\ \hat{\theta}^{(J)} \end{bmatrix} = \left\{ Q_1'(Z^{(J)'}Z^{(J)} \otimes \tilde{\Omega}^{(J)-1})Q_1 \right\}^{-1} Q_1' \operatorname{vec}(\tilde{\Omega}^{(J)-1} Y^{(J)'} Z^{(J)}), \tag{2.9}$$
$$Q_1 = \begin{bmatrix} I_G & 0 \\ 0 & Q \end{bmatrix},$$

where I_G is the G-rowed identity matrix and $\hat{d}^{(J)} = Q\hat{\theta}^{(J)}$. Denoting the gth element of $\hat{c}^{(J)}$ by $\hat{c}_g^{(J)}$, we can estimate C_g by $\hat{C}_g^{(J)} = \exp(\hat{c}_g^{(J)} - \psi(J))$. When there are no restrictions, so $Q = I_G$, (2.9) reduces to (2.5). A standard error for the kth element of $\hat{\theta}^{(J)}$ is given by the square root of the $(G+k)$th diagonal element of the inverse matrix in (2.9). A statistic for testing the further restrictions

$$H_0': S\theta = 0, \tag{2.10}$$

where S is a $L \times K$ matrix of rank $L < K$, is given by

$$\hat{\theta}^{(J)'} S' \left[(0, S) \left\{ Q_1'(Z^{(J)'}Z^{(J)} \otimes \tilde{\Omega}^{(J)-1}) Q_1 \right\} (0, S)' \right]^{-1} S \hat{\theta}^{(J)}. \tag{2.11}$$

3. Limiting distribution theory. In order to establish the limiting distribution of statistics introduced in the previous section, a number of assumptions are introduced. We strengthen (2.1) as follows:

ASSUMPTION 1. There exist $C_g \in (0, \infty)$, $d_g \in (-\frac{1}{2}, \frac{1}{2})$ and $\alpha \in (0, 2]$ such that

$$f_{gg}(\lambda) = C_g \lambda^{-2d_g} + O(\lambda^{\alpha - 2d_g}) \quad \text{as } \lambda \to 0+, \ g = 1, \dots, G.$$

ASSUMPTION 2. In a neighbourhood $(0, \varepsilon)$ of the origin, $f_{gh}(\lambda)$ is differentiable and

$$\left| \frac{d}{d\lambda} f_{gh}(\lambda) \right| = O(\lambda^{-1-d_g-d_h}) \quad \text{as } \lambda \to 0+, \ g, h = 1, \dots, G.$$

Define the coherency between X_{gt} and X_{ht}: $R_{gh}(\lambda) = f_{gh}(\lambda)/f_{gg}^{1/2}(\lambda) f_{hh}^{1/2}(\lambda)$ [see, e.g., Brillinger (1975), page 297]. Introduce the following assumption:

ASSUMPTION 3. For some $\beta \in [0, 2]$,

$$\left| R_{gh}(\lambda) - R_{gh}(0) \right| = O(\lambda^{\beta}) \quad \text{as } \lambda \to 0+, \ g < h = 2, \dots, G.$$

No assumptions whatsoever are imposed on f outside a neighbourhood of the origin, apart from integrability implied by covariance stationarity. Assumption 1 strengthens (2.1) by imposing a rate of convergence of $f_{gg}(\lambda)/C_g \lambda^{-2d_g}$ to 1; a similar assumption was made by Robinson (1994b) in studying the averaged periodogram. Assumption 2 is similar to assumptions of Fox and Taqqu (1986) and Giraitis and Surgailis (1990) in studying the asymptotic properties of Gaussian parameter estimates in scalar series and assumptions of Robinson (1994b) in a semiparametric context. Assumption 3 is vacuous when $G = 1$. When $G > 1$, it holds with $0 < \beta < 1$ when all $R_{gh}(\lambda)$ are in Lip(β); it holds for $\beta > 1$ when their derivatives are in Lip($\beta - 1$), and are zero at $\lambda = 0$ [the real part of $R_{gh}(\lambda)$ automatically has zero derivative at $\lambda = 0$ when $\beta > 1$]. Note that for any vector process admitting a spectral density matrix, including one with power spectral densities which are zero or infinite at some frequencies, $|R_{gh}(\lambda)| \le 1$ for all λ [see Brillinger (1975), page 297]. Assumptions 1–3 are simultaneously satisfied with $\alpha = \beta = 1$ in the case X_t is given by (2.2) and (2.3) when the spectral density matrix of ξ_t is nonsingular and boundedly differentiable in a neighbourhood of the origin. If this matrix is twice boundedly differentiable with zero first derivative at $\lambda = 0$, we can take $\alpha = \beta = 2$, as is the case with certain fractional ARIMA processes. [Note that $(\sin \lambda/\lambda)^{-2d} = 1 + O(\lambda^2)$ as $\lambda \to 0$ for all d.] Assumptions 1 and 2 hold with $\alpha = 2$ in models (1.2) and (1.3).

Before listing further assumptions, we present some results that depend only on Assumptions 1–3 and that are of considerable importance to our limiting distributional theory: they concern the limiting covariance properties of discrete Fourier transforms at frequency λ_j such that j is allowed to increase with n while $j/n \to 0$ as $n \to \infty$. The results provide a contrast to those of Theorem 1, where j was fixed, and they are of some wider relevance to the theory of frequency domain analysis of time series with spectra having zeros or singularities, and seem original even in the case $G = 1$. The proof, which involves careful use of truncation in order that the mild Assumptions 1–3 can suffice, is in Section 4. Denote by $w_g(\lambda)$ the gth element of $w(\lambda)$ given by (1.4) with vector X_t and introduce the scaled discrete Fourier transform $v_g(\lambda) = w_g(\lambda)/(C_g^{1/2}\lambda^{-d_g})$.

THEOREM 2. *Let Assumptions* 1–3 *hold. Then for* $g, h = 1, \dots, G$ *and any sequences of positive integers* $j = j(n)$ *and* $k = k(n)$ *such that* $j > k$ *and* $j/n \to 0$ *as* $n \to \infty$,

$$\text{(a)} \qquad E\{v_g(\lambda_j)\bar{v}_h(\lambda_j)\} = R_{gh}(0) + O\left[\frac{\log j}{j} + \left(\frac{j}{n}\right)^{\min(\alpha,\beta)}\right],$$

$$\text{(b)} \qquad E\{v_g(\lambda_j)v_h(\lambda_j)\} = O\left(\frac{\log j}{j}\right),$$

$$\text{(c)} \qquad E\{v_g(\lambda_j)\bar{v}_h(\lambda_k)\} = O\left(\frac{\log j}{k}\right),$$

$$\text{(d)} \qquad E\{v_g(\lambda_j)v_h(\lambda_k)\} = O\left(\frac{\log j}{k}\right).$$

For the limiting distributional properties of the estimates of Section 2 we need three further assumptions and some additional notation.

Introduce the matrix of coherencies at the origin, $R(0) = (R_{gh}(0))$. The following assumption will not hold if any of the series is perfectly predictable from the others.

ASSUMPTION 4. $R(0)$ is nonsingular.

As discussed in Section 1, we introduce the next assumption:

ASSUMPTION 5. $\{X_t, t = 1, 2 \dots\}$ is a Gaussian process.

Finally a condition on the bandwidth and trimming numbers is introduced:

ASSUMPTION 6. As $n \to \infty$,

$$\frac{m^{1/2} \log m}{l} + \frac{l(\log n)^2}{m} + \frac{m^{1+1/2\min(\alpha,\beta)}}{n} \to 0.$$

Assumption 6 implies that the weakest possible upper bound on m relative to n is $m^5/n^4 \to 0$, whereas acceptable l and m sequences exist for any $\alpha, \beta > 0$. Some limited knowledge is now available concerning the choice of m. Robinson (1994b) derived formulae for m which are optimal in the sense of asymptotically minimizing mean squared error of the (unlogged) periodograms averaged over m frequencies near the origin, in the case of (1.1) with $0 < d < \frac{1}{2}$. Feasible "plug-in" versions of the optimal m were given by Delgado and Robinson (1993). As in other problems involving trimming numbers, there seems to be no optimality theory to guide the choice of l in a given practical problem.

To motivate Assumptions 4 and 5, introduce real-valued $v_g^R(\lambda)$ and $v_g^I(\lambda)$ such that $v_g(\lambda) = v_g^R(\lambda) + iv_g^I(\lambda)$, so

$$U_{gk}^{(J)} = \log\left[\sum_{j=1}^{J}\left\{v_g^R(\lambda_{k+j-J})^2 + v_g^I(\lambda_{k+j-J})^2\right\}e^{-\psi(J)}\right]. \tag{3.1}$$

Now introduce the vector $v(\lambda) = (v_1^R(\lambda), \dots, v_G^R(\lambda), v_1^I(\lambda), \dots, v_G^I(\lambda))'$, where variances and covariances of elements of $v(\lambda_j)$ and $v(\lambda_k)$ can be deduced from those of the $v_g(\lambda_j)$ and $v_g(\lambda_k)$ and their complex conjugates. Approximations to the latter, with error bounds, were given in Theorem 2. These indicate a sense in which the $v(\lambda_j)$, for j increasing slowly with n, can be regarded as approximately uncorrelated with mean zero [because $w(\lambda_j) = (2\pi n)^{-1/2}\sum_{t=1}^{n}(X_t - EX_1)\exp(it\lambda_j)$ for $1 \le j < n$] and covariance matrix

$$R = \frac{1}{2}\begin{bmatrix} R_R & -R_I \\ R_I & R_R \end{bmatrix},$$

where R_R and R_I are the real and imaginary parts of $R(0)$. (The matrix R is Hermitian, so R_R is symmetric and $R_I = -R_I'$.) Assumption 5 implies the $v(\lambda_j)$ are Gaussian and thus that the approximate uncorrelatedness can be reinterpreted as approximate independence. Thus, introduce the $2G$-dimensional vector variates

$$V_j \sim NID(0, R), \qquad j = l+1, \dots, m. \tag{3.2}$$

Assumption 4 is equivalent to positive definiteness of R [see Hannan (1970), page 224], so the V_j have a nonsingular distribution. Denoting by V_{gj} the gth element of V_j, introduce the variates

$$W_{gk}^{(J)} = \log\left[\sum_{j=1}^{J}\left(V_{g,k+j-J}^2 + V_{g+G,k+j-J}^2\right)e^{-\psi(J)}\right], \tag{3.3}$$

for $k = l+J, l+2J, \dots, m$. Because $R_{gg}(0) = 1$ for each g, it follows that the diagonal elements of R_R and R_I are all, respectively, unity and zero, so that $\sum_{j=1}^{J}(V_{g,k+j-J}^2 + V_{g+G,k+j-J}^2) \sim \frac{1}{2}\chi_{2J}^2$ for each g, k. Thus [see, e.g., Johnson and Kotz (1970), pages 167 and 181] $E(W_{gk}^{(J)}) = 0$ and $W_{gk}^{(J)}$ has finite moments of all orders. Denote the covariance matrix of $W_k^{(J)} = (W_{1k}^{(J)}, \dots, W_{Gk}^{(J)})'$ by $\Omega^{(J)}$. From Johnson and Kotz [(1970), page 181], the diagonal elements of $\Omega^{(J)}$ are all $\psi'(J)$, where $\psi'(z) = (d/dz)\psi(z)$. Further, independence of the V_j implies independence of $W_{l+J}^{(J)}, W_{l+2J}^{(J)}, \dots, W_m^{(J)}$. Thus if the $U_{gk}^{(J)}$ in (2.4) can indeed be replaced by the $W_{gk}^{(J)}$ without affecting the limit distribution of our centred and suitably scaled estimates, we can apply the Lindeberg–Feller CLT to complete the proof, as discussed in Section 1. The proof that the $U_{gk}^{(J)}$ can be replaced by the $W_{gk}^{(J)}$ rests heavily on the error bounds established in Theorem 2, but also involves considerable additional manipulation.

The proofs of Theorems 3 and 4 below, which deal, respectively, with the least squares and generalized least squares estimates of Section 2, are in Section 5.

THEOREM 3. *Let Assumptions* 1–6 *hold. Then as* $n \to \infty$,

$$\begin{bmatrix} \dfrac{m^{1/2}}{\log n}\left(\tilde{c}^{(J)} - c^{(J)}\right) \\ 2m^{1/2}\left(\tilde{d}^{(J)} - d\right) \end{bmatrix} \to_d N\left[0, J\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \otimes \Omega^{(J)}\right] \tag{3.4}$$

and the covariance matrix in the limiting distribution is consistently estimated by

$$J\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \otimes \tilde{\Omega}^{(J)}. \tag{3.5}$$

REMARK 1. Theorem 3 provides simple formulae for standard errors, interval estimates and test statistics, but a version which corresponds to the methodology based on standard output from a regression package, presented in Section 2, is readily deduced from Theorem 3,

$$\left\{\sum_k{}'(2\log\lambda_k)^2-\frac{J}{m-l}\left(\sum_k{}'2\log\lambda_k\right)^2\right\}^{1/2} \tag{3.6}$$
$$\times\,\tilde{\Omega}^{(J)-1/2}(\tilde{d}^{(J)}-d)\to_d N(0,I_G),$$

because the quantity in braces is $4m/J+o(m)$ as $n\to\infty$. We likewise deduce the usual large-sample interpretation of regression t-ratios and that (2.7) has a limiting χ_H^2 distribution under (2.6).

REMARK 2. The normal distribution in Theorem 3 is singular, the covariance matrix having rank G owing to perfect negative correlation in the limiting joint distribution of $\tilde{c}_g^{(J)}$ and $\tilde{d}_g^{(J)}$, for each g. Individually, the vectors $\tilde{c}^{(J)}$ and $\tilde{d}^{(J)}$ have nonsingular limiting distributions.

REMARK 3. For each g, $2m^{1/2}(\tilde{d}_g^{(J)}-d_g)\to_d N(0,J\psi'(J))$. In the case $J=1$, as in $\bar{d}(l)$, $J\psi'(J)=\pi^2/6=1.645$. The recurrence relation $\psi'(J+1)=\psi'(J)-J^{-2}$ indicates that the function $J\psi'(J)$ decreases in J, taking values 1.289 at $J=2$ and 1.185 at $J=3$, and tending to 1 as $J\to\infty$. One can modify $\bar{d}(l)$ in the opposite direction, logging the squared cosine and sine transforms separately, to give $2(m-l)$ rather than $(m-l)/J$ "observations." It may be shown that the formulae of Theorem 3 apply to this estimate with $J=\frac{1}{2}$, but because $\frac{1}{2}\psi'(\frac{1}{2})=2.467$, there is now substantial efficiency loss.

REMARK 4. In inference on a single d_g one can use $J\psi'(J)$ in place of the gth diagonal element of $\tilde{\Omega}^{(J)}$, but in inferences involving more than one d_g, one would use the latter, including relevant off-diagonal elements, in order to ensure nonnegative definiteness.

REMARK 5. $\tilde{c}^{(J)}$ converges more slowly than $\tilde{d}^{(J)}$.

REMARK 6. A simple application of the "delta" method provides a CLT for the estimate $\tilde{C}^{(J)}=(\tilde{C}_1^{(J)},\dots,\tilde{C}_G^{(J)})'$ of the vector $C=(C_1,\dots,C_G)'$ of scale parameters in (2.1):

$$(m/\log n)^{1/2}(\tilde{C}^{(J)}-C)\to_d N\big(0,J\operatorname{diag}\{C\}\,\Omega^{(J)}\operatorname{diag}\{C\}\big).$$

REMARK 7. In practical applications, X_t may be a vector of unobservable errors, which can be expressed as a parametric or nonparametric filter of observable time series. Then if X_t can be proxied by residuals $\hat{X}_t$ by means of estimates of the filter which ignore the long range dependence of X_t but are consistency-robust to this, with a suitably fast rate of convergence, our results should still go through for estimates of the C_g, d_g based on the $\hat{X}_t$.

Now we turn to the generalized least squares estimates.

THEOREM 4. *Let Assumptions* 1–6 *hold. Then as* $n \to \infty$,

$$\begin{bmatrix} \dfrac{m^{1/2}}{\log n}(\hat{c}^{(J)} - c^{(J)}) \\ 2m^{1/2}(\hat{d}^{(J)} - d) \end{bmatrix} \tag{3.7}$$

$$\to_d N\left(0, J\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \otimes Q\left(Q'\Omega^{(J)-1}Q\right)^{-1}Q'\right)$$

and the covariance matrix in the limiting distribution is consistently estimated by

$$J\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \otimes Q\left(Q'\tilde{\Omega}^{(J)-1}Q\right)^{-1}Q'. \tag{3.8}$$

Because $\Omega^{(J)} - Q(Q'\Omega^{(J)-1}Q)^{-1}Q'$ is nonnegative definite, the usual sort of efficiency improvement achieved by generalized least squares is confirmed. We have equivalently $2m^{1/2}(\hat{\theta}^{(J)} - \theta) \to_d N(0, J(Q'\Omega^{(J)-1}Q)^{-1})$. The remarks following Theorem 3 continue to apply with obvious modifications. In particular, we have an analogous result to (3.6)—the standard errors mentioned in Section 2 can be used instead to normalize individual $\hat{d}_g^{(J)}$—and (2.11) has an asymptotic χ_L^2 distribution under (2.10).

For scalar X_t, Hassler (1993a) proposed an estimate of d, which we denote $\underline{d}(l)^*$, and which involves not our pooling, but replacing each $I(\lambda_j)$ in $\bar{d}(l)$ by a smoothed nonparametric estimate $\hat{f}(\lambda_j)$ of $f(\lambda_j)$. He did not give an asymptotic distribution theory but an order of magnitude for the variance of $d(l)^*$ when $d < -1$; his proof seems to entail assertions that the $\hat{f}(\lambda_j)$ are asymptotically uncorrelated across adjacent λ_j [compare, however, Hannan (1970), pages 280–281] and that the contribution from the $O(m)$ variance terms in $\mathrm{Var}\{d(l)^*)\}$ [arising from the summation form of $d(l)^*$] is not dominated asymptotically by that from the $O(m^2)$ covariance terms.

4. Proof of Theorem 2. Fix g, h and for notational simplicity write $\varepsilon_{jk} = k^{-1}\log j\lambda_j^{-d_g}\lambda_k^{-d_h}$, $\rho(\lambda) = f_{gh}(\lambda)$, $\rho'(\lambda) = (d/d\lambda)f_{gh}(\lambda)$ and $\delta = d_g + d_h$. The proof of (a) is in two parts. First we show that

$$E\{w_g(\lambda_j)\bar{w}_h(\lambda_j)\} - \rho(\lambda_j) = O(\varepsilon_{jj}) \tag{4.1}$$

and then that

$$\rho(\lambda_j) - C_g^{1/2}C_h^{1/2}\lambda_j^{-\delta}R_{gh}(0) = O\left[\left(\frac{j}{n}\right)^{\min(\alpha,\beta)}\lambda_j^{-\delta}\right]. \tag{4.2}$$

The left-hand side of (4.1) is

$$\int_{-\pi}^{\pi}\{\rho(\lambda) - \rho(\lambda_j)\}K(\lambda - \lambda_j)\,d\lambda, \tag{4.3}$$

where $K(\lambda)$ is proportional to Fejér's kernel,

$$K(\lambda) = (2\pi n)^{-1}|\Sigma\Sigma_{t,s=1}^{n} e^{i(t-s)\lambda}|^2.$$

Assumption 1 implies that we can pick ε so small that for some $C_\varepsilon < \infty$, $|\rho(\lambda)| \le f_{gg}^{1/2}(\lambda) f_{hh}^{1/2}(\lambda) \le C_\varepsilon |\lambda|^{-\delta}$ and $|\rho'(\lambda)| \le C_\varepsilon |\lambda|^{-1-\delta}$, for $\lambda \in (-\varepsilon, 0) \cup (0, \varepsilon)$, $g, h = 1, \dots, G$. For such ε, the component of (4.3) due to integration over $(-\pi, -\varepsilon) \cup (\varepsilon, \pi)$ has bound

$$\left| \int_{-\pi}^{-\varepsilon} + \int_{\varepsilon}^{\pi} \right| \le \max_{|\lambda| \ge \varepsilon} K(\lambda - \lambda_j) \int_{-\pi}^{\pi} \left| \rho(\lambda) - \rho(\lambda_j) \right| d\lambda = O\left(n^{-1}\left(1 + \lambda_j^{-\delta}\right)\right)$$

by elementary inequalities and because $2\lambda_j < \varepsilon$ for n sufficiently large, because

$$K(\lambda) = O\left((n\lambda^2)^{-1}\right), \qquad 0 < |\lambda| < \pi \tag{4.4}$$

[see, e.g., Zygmund (1977), page 89], because $|\rho(\lambda)| \le f_{gg}^{1/2}(\lambda) f_{hh}^{1/2}(\lambda)$ and because $\int_{-\pi}^{\pi} f_{gg}(\lambda)\, d\lambda = \operatorname{Var}\{X_{gt}\} < \infty$. However $n^{-1} = O((j/n)^{1+\delta} j^{-1} \lambda_j^{-\delta}) = o(\varepsilon_{jj})$ because

$$1 + \delta > 0. \tag{4.5}$$

For n large enough there is also a contribution to (4.3) bounded by

$$\left| \int_{-\varepsilon}^{-\lambda_j/2} \right| \le \left\{ \max_{\lambda_j/2 \le \lambda \le \varepsilon} \frac{|\rho(\lambda)|}{\lambda^{(1-\delta)/2}} \right\} \int_{\lambda_j/2}^{\pi} \lambda^{(1-\delta)/2} K(\lambda + \lambda_j)\, d\lambda$$

$$+ \left| \rho(\lambda_j) \right| \int_{\lambda_j/2}^{\pi} K(\lambda + \lambda_j)\, d\lambda.$$

This is $O(j^{-1}\lambda_j^{-\delta})$ because the factor in braces is $O(\max_{\lambda_j/2 \le \lambda \le \varepsilon} \lambda^{-(1+\delta)/2}) = O(\lambda_j^{-(1+\delta)/2})\lambda$ due to (4.5), because

$$\int_{\lambda_j/2}^{\pi} \lambda^{(1-\delta)/2} K(\lambda + \lambda_j)\, d\lambda = O\left(n^{-1} \int_{\lambda_j/2}^{\infty} \lambda^{-(3+\delta)/2}\, d\lambda \right) = O\left(n^{-1} \lambda_j^{-(1+\delta)/2}\right)$$

from (4.4) and (4.5) and because $\int_{\lambda_j/2}^{\pi} K(\lambda + \lambda_j)\, d\lambda = O(n^{-1} \int_{\lambda_j}^{\infty} \lambda^{-2}\, d\lambda) = O(j^{-1})$. By an identical argument, $|\int_{2\lambda_j}^{\varepsilon}| = O(j^{-1}\lambda_j^{-\delta})$. The mean value theorem gives

$$\left| \int_{\lambda_j/2}^{2\lambda_j} \right| \le \left\{ \max_{\lambda_j/2 \le \lambda \le 2\lambda_j} |\rho'(\lambda)| \right\} \int_{\lambda_j/2}^{2\lambda_j} |\lambda - \lambda_j| K(\lambda - \lambda_j)\, d\lambda. \tag{4.6}$$

The factor in braces is $O(\lambda_j^{-1-\delta})$ by Assumptions 1 and 2. Writing $K(\lambda) = (2\pi n)^{-1} |D(\lambda)|^2$, where Dirichlet's kernel $D(\lambda) = \Sigma_{t=1}^{n} e^{it\lambda}$ satisfies

$$|D(\lambda)| \le 2|\lambda|^{-1}, \qquad 0 < |\lambda| < \pi, \tag{4.7}$$

[see, e.g., Zygmund (1977), pages 49–51] and, from Lemma 5 of Robinson (1994b),

$$\int_{-C\lambda_j}^{C\lambda_j} |D(\lambda)|\, d\lambda = O(\log j) \tag{4.8}$$

for $C < \infty$, it follows that (4.6) $= O(\varepsilon_{jj})$. To complete the proof of (4.1),

$$\left|\int_{-\lambda_j/2}^{\lambda_j/2}\right| \le \max_{|\lambda|\le \lambda_j/2} K(\lambda - \lambda_j)\int_{-\lambda_j/2}^{\lambda_j/2}\{|\rho(\lambda)| + |\rho(\lambda_j)|\}\,d\lambda$$

$$= O\left(n^{-1}\lambda_j^{-2}\left\{\int_0^{\lambda_j}\lambda^{-\delta}\,d\lambda + \lambda_j^{1-\delta}\right\}\right) = O(\varepsilon_{jj}).$$

The left-hand side of (4.2) is dominated by

$$f_{gg}^{1/2}(\lambda_j)f_{hh}^{1/2}(\lambda_j)\left|1 - \frac{C_g^{1/2}C_h^{1/2}\lambda_j^{-\delta}}{f_{gg}^{1/2}(\lambda_j)f_{hh}^{1/2}(\lambda_j)}\right| + |R_{gh}(\lambda_j) - R_{gh}(0)|C_g^{1/2}C_h^{1/2}\lambda_j^{-\delta}.$$

As $n \to \infty$, the second term is $O(\lambda_j^{\beta-\delta})$ from Assumption 3, whereas the first term is $O(\lambda_j^{\alpha-\delta})$ from the inequality $1 - x^2 \le 2(1 - x)$, for $0 < x < 1$, and Assumption 1. Thus (a) is proved.

To prove (b), for $1 < j < n/2$,

$$E\{w_g(\lambda_j)w_h(\lambda_j)\} = (2\pi n)^{-1}\int_{-\pi}^{\pi}\{\rho(\lambda) - \rho(\lambda_j)\}E_{j,-j}(\lambda)\,d\lambda, \tag{4.9}$$

where $E_{jk}(\lambda) = (2\pi n)^{-1}D(\lambda_j - \lambda)D(\lambda - \lambda_k)$, because

$$\int_{-\pi}^{\pi} E_{j,-k}(\lambda)\,d\lambda = 0 \quad \text{for } 1 \le j + k < n. \tag{4.10}$$

Decompose (4.9) as

$$\int_{-\pi}^{-\varepsilon} + \int_{-\varepsilon}^{-2\lambda_j} + \int_{-2\lambda_j}^{-\lambda_j/2} + \int_{-\lambda_j/2}^{\lambda_j/2} + \int_{\lambda_j/2}^{2\lambda_j} + \int_{2\lambda_j}^{\varepsilon} + \int_{\varepsilon}^{\pi}.$$

The calculations are much as before and so we present them in a more abbreviated form. We have

$$\left|\int_{-\pi}^{-\varepsilon} + \int_{\varepsilon}^{\pi}\right| = O\left(\frac{1}{n\varepsilon^2}\int_{-\pi}^{\pi}\{|\rho(\lambda)| + |\rho(\lambda_j)|\}\,d\lambda\right) = O(j^{-1}\lambda_j^{-\delta}),$$

$$\left|\int_{-\varepsilon}^{-2\lambda_j} + \int_{2\lambda_j}^{\varepsilon}\right| = O\left(\left\{\max_{2\lambda_j\le\lambda\le\varepsilon}\frac{|\rho(\lambda)|}{\lambda^{(1-\delta)/2}}\right\}\right.$$

$$\left.\times\frac{1}{n}\int_{\lambda_j}^{\infty}\lambda^{-(3+\delta)/2}\,d\lambda + \frac{|\rho(\lambda_j)|}{n}\int_{\lambda_j}^{\infty}\lambda^{-2}\,d\lambda\right)$$

$$= O(\varepsilon_{jj}),$$

$$\left|\int_{-2\lambda_j}^{-\lambda_j/2} + \int_{\lambda_j/2}^{2\lambda_j}\right| = O\left(\left\{\max_{\lambda_j/2\le\lambda\le 2\lambda_j}|\rho'(\lambda)|\right\}\frac{1}{n}\int_{-3\lambda_j}^{3\lambda_j}|D(\lambda)|\,d\lambda\right) = O(\varepsilon_{jj}),$$

$$\left|\int_{-\lambda_j/2}^{\lambda_j/2}\right| = O\left(\max_{|\lambda|\le\lambda_j/2}|E_{j,-j}(\lambda)|\int_0^{\lambda_j}\{|\rho(\lambda)| + |\rho(\lambda_j)|\}\,d\lambda\right) = o(\varepsilon_{jj}).$$

Thus (b) is proved.

To prove (c), write $E\{w_g(\lambda_j)\overline{w}_h(\lambda_k)\} = \int_{-\pi}^{\pi} \rho(\lambda)E_{jk}(\lambda)\,d\lambda$, for $O < k < j < n$. This can be expanded as

$$\int_{(\lambda_j+\lambda_k)/2}^{2\lambda_j} \{\rho(\lambda) - \rho(\lambda_j)\}E_{jk}(\lambda)\,d\lambda \tag{4.11}$$

$$+\int_{\lambda_k/2}^{(\lambda_j+\lambda_k)/2} \{\rho(\lambda) - \rho(\lambda_k)\}E_{jk}(\lambda)\,d\lambda \tag{4.12}$$

$$-\{\rho(\lambda_j) - \rho(\lambda_k)\}\int_{\lambda_k/2}^{(\lambda_j+\lambda_k)/2} E_{jk}(\lambda)\,d\lambda \tag{4.13}$$

$$+\left\{\int_{2\lambda_j}^{\pi} + \int_{-\pi}^{\lambda_k/2}\right\}\{\rho(\lambda) - \rho(\lambda_j)\}E_{jk}(\lambda)\,d\lambda \tag{4.14}$$

because of (4.10). From Assumption 2, (4.7) and (4.8), it follows that (4.11) is bounded by

$$\left\{\max_{(\lambda_j+\lambda_k)/2\le\lambda\le 2\lambda_j} |\rho'(\lambda)|\right\} n^{-1}\int_{(\lambda_j+\lambda_k)/2}^{2\lambda_j} |D(\lambda-\lambda_k)|\,d\lambda = O(\varepsilon_{jj}),$$

which is $O(\varepsilon_{jk})$ for $j > k$. Next, (4.12) is bounded by

$$\begin{aligned}\max_{\lambda_k/2\le\lambda\le(\lambda_j+\lambda_k)/2} &|\rho'(\lambda)| n^{-1}\int_{\lambda_k/2}^{(\lambda_j+\lambda_k)/2} |D(\lambda_j-\lambda)|\,d\lambda \\ &= O\left(\frac{\log j}{k}\lambda_k^{-\delta}\right) = O(\varepsilon_{jk})\end{aligned}$$

when $k \ge j/2$ and by

$$\begin{aligned}\left\{\max_{\lambda_k/2\le\lambda\le\lambda_j} |\rho(\lambda)| + |\rho(\lambda_k)|\right\}&\int_{\lambda_k/2}^{(\lambda_j+\lambda_k)/2} |E_{jk}(\lambda)|\,d\lambda \\ &= O\left((\lambda_j^{-\delta} + \lambda_k^{-\delta})(j-k)^{-1}\int_0^{\lambda_j} |D(\lambda)|\,d\lambda\right) \\ &= O(\varepsilon_{jk})\end{aligned}$$

when $k < j/2$. Next, (4.13) is bounded by

$$\begin{aligned}(\lambda_j - \lambda_k)\left\{\max_{\lambda_k\le\lambda\le\lambda_j} |\rho'(\lambda)|\right\}&\int_{\lambda_k/2}^{(\lambda_j+\lambda_k)/2} |E_{jk}(\lambda)|\,d\lambda \\ &= O\left(\lambda_k^{-1-\delta}n^{-1}\int_0^{\lambda_j} |D(\lambda)|\,d\lambda\right) = O(\varepsilon_{jk})\end{aligned}$$

when $k \ge j/2$, and by $O(\{|\rho(\lambda_j)| + |\rho(\lambda_k)|\}j^{-1}\int_0^{\lambda_j}|D(\lambda)|\,d\lambda) = O(\varepsilon_{jk})$ when $k < j/2$, as in the evaluation of (4.12). To estimate (4.14), first consider the integral on $[2\lambda_j, \varepsilon]$, where ε and n are chosen as in the proof of (a). This integral is bounded by

$$\left\{\max_{\lambda_j \le \lambda \le \varepsilon} \frac{|\rho(\lambda)|}{\lambda^{(1-\delta)/2}}\right\}\int_{2\lambda_j}^{\varepsilon} \lambda^{(1-\delta)/2}|E_{jk}(\lambda)|\,d\lambda$$

$$= O\left(\left\{\max_{\lambda_j \le \lambda \le \varepsilon} \lambda^{-(1+\delta)/2}\right\} n^{-1}\int_{\lambda_j}^{\infty} \lambda^{-(3+\delta)/2}\,d\lambda\right) = O\left(j^{-1}\lambda_j^{-\delta}\right) = O(\varepsilon_{jk}).$$

The components of (4.14) due to the integrals on $[-\lambda_k/2, \lambda_k/2]$ and $[-\lambda_j, -\lambda_k/2]$ are, respectively,

$$O\left(\frac{1}{n\lambda_j\lambda_k}\int_{-\lambda_k/2}^{\lambda_k/2}\{|\rho(\lambda)| + |\rho(\lambda_j)|\}\,d\lambda\right) = O(\varepsilon_{jk})$$

and

$$O\left(\max_{\lambda_k/2 \le \lambda \le \lambda_j} |\rho(\lambda)| j^{-1}\int_0^{\lambda_j}|D(\lambda - \lambda_k)|\,d\lambda\right)$$

$$= O\left(\frac{\log j}{j}\left(\lambda_j^{-\delta} + \lambda_k^{-\delta}\right)\right) = O(\varepsilon_{jk}).$$

The component of (4.14) due to the integral on $[-\varepsilon, -\lambda_j]$ is handled like that on $[2\lambda_j, \varepsilon]$, and finally

$$\left|\int_{-\pi}^{-\varepsilon} + \int_{\varepsilon}^{\pi}\right| = O\left(n^{-1}\varepsilon^{-2}\int_{-\pi}^{\pi}\{|\rho(\lambda)| + |\rho(\lambda_j)|\}\,d\lambda\right)$$

$$\doteq O\left(n^{-1}\left(1 + \lambda_j^{-\delta}\right)\right) = O(\varepsilon_{jk}).$$

Thus (c) is proved.

Finally, to prove (d) write

$$E\{w_g(\lambda_j)w_h(\lambda_k)\} = \int_{-\pi}^{\pi}\{\rho(\lambda) - \rho(\lambda_j)\}E_{j,-k}(\lambda)\,d\lambda.$$

The remaining details are much like those in the proof of (c), but easier because there is no need to distinguish between close and distant j, k. We have

$$\left|\int_{-\pi}^{-\varepsilon} + \int_{\varepsilon}^{\pi}\right| = O\left(\frac{1}{n\varepsilon^2}\int_{-\pi}^{\pi}\{|\rho(\lambda)| + |\rho(\lambda_j)|\}\,d\lambda\right) = o(\varepsilon_{jk}),$$

$$\left|\int_{-\varepsilon}^{-2\lambda_j} + \int_{2\lambda_j}^{\varepsilon}\right| = O\left(\left\{\max_{\lambda_j \le \lambda \le \varepsilon} \frac{|\rho(\lambda)|}{\lambda^{(1-\delta)/2}}\right\} n^{-1}\int_{\lambda_j}^{\infty}\lambda^{-(3+\delta)/2}\,d\lambda\right.$$

$$\left. + \frac{|\rho(\lambda_j)|}{n}\int_{\lambda_j}^{\infty}\lambda^{-2}\,d\lambda\right) = O(\varepsilon_{jk}),$$

$$\left|\int_{-2\lambda_j}^{\lambda_k/2}\right| = O\left(\left[\left\{\max_{\lambda_k/2\le\lambda\le2\lambda_j}|\rho(\lambda)|\right\}\right]j^{-1}\int_0^{3\lambda_j}|D(\lambda)|\,d\lambda\right) = O(\varepsilon_{jk}),$$

$$\left|\int_{-\lambda_k/2}^{\lambda_k/2}\right| = O\left(\frac{1}{n\lambda_j\lambda_k}\int_{-\lambda_k/2}^{\lambda_k/2}\{|\rho(\lambda)|+|\rho(\lambda_j)|\}\,d\lambda\right) = O(\varepsilon_{jk}),$$

$$\left|\int_{\lambda_k/2}^{\lambda_j/2}\right| \le \left\{\max_{\lambda_k/2\le\lambda\le\lambda_j}|\rho(\lambda)|\right\}j^{-1}\int_0^{\lambda_j}|D(\lambda)|\,d\lambda = O(\varepsilon_{jk}),$$

$$\left|\int_{\lambda_j/2}^{2\lambda_j}\right| = O\left(\left\{\max_{\lambda_j/2\le\lambda\le2\lambda_j}|\rho'(\lambda)|\right\}n^{-1}\int_0^{3\lambda_j}|D(\lambda)|\,d\lambda\right) = O(\varepsilon_{jk}). \qquad \square$$

5. Proofs of Theorems 3 and 4. We shall prove these theorems more or less simultaneously. The proof is via the method of moments. We show that each moment of any linear combination of the variates on the left-hand side of either (3.4) or (3.7) converges to the corresponding moment of the normal distribution implied by the respective right-hand sides, and then appeal to the Frechet–Shohat "moment convergence theorem" [see, e.g., Loeve (1977), page 187] and unique determination of the normal distribution by its moments. To accomplish this, we first use Theorem 2(c) and (d) to show that the moments differ negligibly from those which would arise if the $U_k^{(J)}$ (like the $W_k^{(J)}$) were actually independent. Then we use Theorem 2 to show that these latter moments in turn differ negligibly from those which would arise if the $U_k^{(j)}$ (like the $W_k^{(J)}$) were actually independent. Then we use Theorem 2 to show that these latter moments in turn differ negligibly from those which would arise if the $v(\lambda_j)$ had identical joint distributions to those of the V_j [see (3.2)]. We are left with a variate which is straightforwardly found to be asymptotically normal by the Lindeberg–Feller CLT. Because each moment of this variate is bounded uniformly in n, it follows from Theorem A of Loeve [(1977), page 185] that all of its moments converge to those of the same normal distribution.

For ease of presentation we omit the (J) superscript throughout. Define $z_k = -2\log\lambda_k$ and $S(A) = Q_1'(Z'Z\otimes A)Q_1$, $s(B) = (\Sigma_k U_k'B', \Sigma_k z_k' U_k' B'Q)'$ for $G\times G$ matrices A and B, and put $T(A) = S(A)^{-1}s(A)$. Then $T(I_G) = (\tilde{c}' - c', \tilde{d}' - d')'$ when $Q = I_G$ and $T(\hat{\Omega}^{-1}) = (\hat{c}' - c', \hat{\theta}' - \theta')'$. By inversion of partitioned matrices, for nonsingular A,

$$S(A)^{-1} = \frac{(m-l)/J}{|Z'Z|}\begin{bmatrix}\bar{z}Q\\ -I_G\end{bmatrix}(Q'AQ)^{-1}\begin{bmatrix}\bar{z}Q\\ -I_G\end{bmatrix}' + \frac{J}{m-l}\begin{bmatrix}A^{-1} & 0\\ 0 & 0\end{bmatrix}, \tag{5.1}$$

where $\bar{z} = (J/(m-l))\Sigma_k' z_k$. Also

$$[\bar{z}Q', -I_G]s(B) = -Q'B\sum_k(z_k-\bar{z})U_k, \qquad \begin{bmatrix}A^{-1} & 0\\ 0 & 0\end{bmatrix}s(B) = \begin{pmatrix}A^{-1}B\\ 0\end{pmatrix}\sum_k{}' U_k.$$

By approximation of sums by integrals we have, as $n \to \infty$,

$$\sum_k{}' z_k = (2m/J)\{\log n + 1\} + O(l \log n), \tag{5.2}$$

$$\sum_k{}' z_k^2 = (4m/J)\{(\log n)^2 + 2(\log n) + 2\} + O\big(l(\log n)^2\big), \tag{5.3}$$

and thus

$$|Z'Z| = 4m^2/J^2 + O\big(lm(\log n)^2\big). \tag{5.4}$$

From (5.2), (5.4) and Assumption 6, as $n \to \infty$,

$$(m^{1/2}/\log n)\bar{z} = 2m^{1/2}\big(1 + O(\log n)^{-1}\big),$$

$$(J/(m-l))|Z'Z| = 4m/J + O\big(l(\log n)^2\big).$$

Putting

$$\Delta = \begin{bmatrix} (m^{1/2}/\log n) I_G & 0 \\ 0 & 2m^{1/2} I_G \end{bmatrix},$$

it follows that

$$\begin{aligned} \Delta S(A)^{-1} s(B) = -J^{1/2}&\left(\begin{bmatrix} Q \\ -I_G \end{bmatrix}(Q'AQ)^{-1}Q'B, \begin{bmatrix} -A^{-1}QB \\ 0 \end{bmatrix}\right) \\ &\times \left(\frac{J}{m + o(m)}\right)^{1/2} \sum_k{}' \begin{bmatrix} \frac{1}{2}(z_k - \bar{z})U_k \\ U_k/\log n \end{bmatrix}. \end{aligned} \tag{5.5}$$

Assumptions 4 and 5 imply that there exists no set of constants $\omega_1, \dots, \omega_G$, not all zero, such that $\sum_{g=1}^{G} \omega_g \log W_{gk} = 0$ a.s.; thus Ω is nonsingular. Then (3.4) and (3.7) will follow, if, as $n \to \infty$,

$$\begin{aligned} (J/m)^{1/2} \sum_k{}' (z_k - \bar{z})U_k &\to_d N(0, 4\Omega), \\ (m^{1/2} \log n)^{-1} \sum_k{}' U_k &\to_p 0. \end{aligned} \tag{5.6}$$

Consistency of (3.5) and (3.8) evidently follows if $\tilde{\Omega} \to_p \Omega$. However,

$$\begin{aligned} \tilde{\Omega} &= \frac{J}{m-l} \sum_k{}' U_k U_k' - (\tilde{c} - c, \tilde{d} - d)\frac{J}{m-l} Z'Z(\tilde{c} - c, \tilde{d} - d)' \\ &= \frac{J}{m-l} \sum_k{}' U_k U_k' + O_p\left(\frac{(\log n)^2}{m}\right) \quad \text{as } n \to \infty, \end{aligned} \tag{5.7}$$

from (3.4), (5.2) and (5.3). Thus consistency of (3.5) and (3.8) follows if

$$\frac{1}{m} \sum_k{}' U_k U_k' \to_p \Omega \quad \text{as } n \to \infty. \tag{5.8}$$

Also, from (5.2) and (5.3),

$$\Delta^{-1}S(A)\Delta^{-1} = (\log n)^2(1 + o(1))Q_1'\left\{\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \otimes A\right\}Q_1 \tag{5.9}$$

as $n \to \infty$ and from (5.1), (5.2) and (5.4),

$$(\Delta^{-1}S(A)\Delta^{-1})^{-1} \to J(Q', -I_G)'(Q'AQ)^{-1}(Q', -I_G) \quad \text{as } n \to \infty, \tag{5.10}$$

$$\Delta T(\tilde{\Omega}^{-1}) = \left(I_{G+K} + \left(\Delta^{-1}S(\Omega^{-1})\Delta^{-1}\right)^{-1}\Delta^{-1}S(\tilde{\Omega}^{-1} - \Omega^{-1})\Delta^{-1}\right)^{-1} \tag{5.11}$$

$$\times \left(\Delta T(\Omega^{-1}) + \Delta S(\Omega^{-1})^{-1}s(\tilde{\Omega}^{-1} - \Omega^{-1})\right). \tag{5.12}$$

It follows from (5.9), (5.10) and Slutzky's theorem that the right-hand side of (5.11) is $I_{G+K} + o_p(1)$ if $\tilde{\Omega} = \Omega + o_p((\log n)^{-2})$, which, from (5.7) and Assumption 6, is equivalent to

$$\frac{J}{m}\sum_k{}' U_k U_k' = \Omega + o_p\left((\log n)^{-2}\right) \quad \text{as } n \to \infty. \tag{5.13}$$

Also $\tilde{\Omega} \to_p \Omega$, (5.5) and (5.6) together imply (5.12) is $\Delta T(\Omega^{-1}) + o_p(1)$ so that (3.7) is a consequence of (5.6) and (5.13). It thus remains to establish (5.6) and (5.13) [which implies (5.8)]. Now (5.6) is in turn a consequence of

$$(J/m)^{1/2}\sum_k{}' a_k \nu' U_k \to_d N(0, \nu'\Omega\nu) \quad \text{as } n \to \infty, \tag{5.14}$$

for any $G \times 1$ vector ν and triangular array $a_{kn} = a_k$ satisfying, as $n \to \infty$,

$$\max_k |a_k| = o(m), \qquad \sum_k{}' a_k^2 \sim \frac{m}{J}, \qquad \sum_k{}' |a_k|^p = O(m) \quad \text{for all } p \geq 1. \tag{5.15}$$

This claim is clearly true in the case of the second part of (5.6). In the case of the first part, note that the first two parts of (5.15) follow from $z_k - \bar{z} = O(\log m)$ and (5.4), while the third part is verified by

$$\sum_k{}' |z_k - \bar{z}|^p = \sum_k{}' \left|\log k - \frac{J}{m-l}\sum_k{}' \log k\right|^p \leq 4^{p-1}\sum_k{}' \left|\log\left(\frac{k}{m}\right)\right|^p + O(m)$$
$$= O\left(m\left\{\int_0^1 |\log x|^p\,dx + 1\right\}\right) = O(m),$$

because $\int_0^1 |\log x|^p\,dx < \infty$ for all p. Write $\chi_k = (J/m)^{1/2} a_k \nu' U_k$. Fix an integer N; $E(\sum_k{}' \chi_k)^N$ is a sum of finitely many terms of the form

$$\sum_{k_1}{}' \cdots \sum_{k_M}{}' E\left(\prod_{i=1}^{M} \chi_{k_i}^{N_{k_i}}\right), \tag{5.16}$$

where $N_{k_1}, \ldots, N_{k_M}$ are all positive and sum to N and $1 \leq M \leq N$. Fix such M and $N_{k_1}, \ldots, N_{k_M}$. Introduce the $2GJ \times 1$ vector $v_k^* = (v(\lambda_{k+1-J})', v(\lambda_{k+2-J})', \ldots, v(\lambda_k)')'$ and the $2GJM \times 1$ vector $v^* = (v_{k_1}^{*\prime}, \ldots, v_{k_M}^{*\prime})$. Following the discussion of Section 3, v^* is normally distributed with zero mean. Theorem 2 implies that

$$E(v_j^* v_k^{*\prime}) = I_J \otimes R + O\left(\frac{\log j}{j} + \left(\frac{j}{n}\right)^{\min(\alpha,\beta)}\right), \qquad j = k,$$

$$= O\left(\frac{\log j}{k}\right) \qquad (j > k),$$

as $n \to \infty$. It follows from Assumption 6 that $\Sigma^* = E(v^* v^{*\prime})$ satisfies

$$(5.17)\quad \Sigma^* = I_{JM} \otimes R + O\left(\frac{\log m}{l} + \left(\frac{m}{n}\right)^{\min(\alpha,\beta)}\right) = I_{JM} \otimes R + o(m^{-1/2}),$$

as $n \to \infty$. For n sufficiently large, $\Psi = \Sigma^{*-1}$ exists by (5.17) and Assumption 4. Denote by Ψ_{ij} the (i, j)th $2GJ \times 2GJ$ submatrix of Ψ and write

$$\tilde{\Psi} = \begin{bmatrix} \Psi_{11} & & 0 \\ & \ddots & \\ 0 & & \Psi_{MM} \end{bmatrix}, \qquad \overline{\Psi} = \Psi - \tilde{\Psi}.$$

It follows that

$$(5.18)\quad \tilde{\Psi} = I_{JM} \otimes R^{-1} + o(m^{-1/2}), \qquad \overline{\Psi} = o(m^{-1/2}) \quad \text{as } n \to \infty.$$

Now denote by φ_p the density function of a p-dimensional standard normal variate. Then (5.16) is

$$(5.19)\qquad \sum_{k_1}{}' \cdots \sum_{k_M}{}' |\Psi|^{1/2} \int \left(\prod_{i=1}^{M} \chi_{k_i}^{N_{k_i}}\right) \varphi_{2GJM}(\Psi^{1/2} v^*)\, dv^*,$$

for n sufficiently large. Consider

$$(5.20)\qquad \sum_{k_1}{}' \cdots \sum_{k_M}{}' |\Psi|^{1/2} \prod_{i=1}^{M} \left\{\int \chi_{k_i}^{N_{k_i}} \varphi_{2GJ}\left(\Psi_{ii}^{1/2} v_{k_i}^*\right) dv_{k_i}^*\right\}.$$

The difference between (5.19) and (5.20) is

$$(5.21)\qquad \sum_{k_1}{}' \cdots \sum_{k_M}{}' |\Psi|^{1/2} \int \left(\prod_{i=1}^{M} \chi_{k_i}^{N_{k_i}}\right) \varphi_{2GJM}(\tilde{\Psi}^{1/2} v^*) \times \left\{\exp\left(-\tfrac{1}{2} v^{*\prime} \overline{\Psi} v^*\right) - 1\right\} dv^*.$$

For any positive integer r the mean value theorem indicates that $|e^u - \sum_{t=0}^{r-1} u^t/t!| \le |u|^r e^{|u|}/r!$, for all u. For all $\varepsilon > 0$ there exists $C_\varepsilon < \infty$ such that $|u|^r \le C_\varepsilon e^{\varepsilon |u|}$ for all u. Following (5.18), choose n so large that $\|\overline{\Psi}\| < \varepsilon$, where $\|\cdot\|$ is the Euclidean norm, that is, $\exp(\frac{1}{2}|v^{*\prime}\overline{\Psi}v^*|) \le \exp(\frac{1}{2}\varepsilon\|v^*\|^2)$. Again using (5.18), $|v^{*\prime}\overline{\Psi}v^*| = o(m^{-1/2}\|v^*\|^2)$ uniformly in v^*. Thus

$$\left|\exp\left(-\frac{1}{2}v^{*\prime}\overline{\Psi}v^*\right) - \sum_{t=0}^{r-1}\frac{\left(-\frac{1}{2}v^{*\prime}\overline{\Psi}v^*\right)^t}{t!}\right| = o\left(m^{-r/2}\exp(\varepsilon\|v^*\|^2)\right)$$

as $n \to \infty$, uniformly in v^*. Thus the difference between (5.21) and

$$(5.22)\quad \sum_{k_1}{}' \cdots \sum_{k_M}{}' |\Psi|^{1/2} \int \left(\prod_{i=1}^{M} \chi_{k_i}^{N_{k_i}}\right) \varphi_{2GJM}(\tilde{\Psi}^{-1/2}v^*) \sum_{t=1}^{r-1} \frac{\left(-(1/2)v^{*\prime}\tilde{\Psi}v^*\right)^t}{t!}\, dv^*$$

is

$$(5.23)\quad o\left(m^{-r/2}\sum_{k_1}{}' \cdots \sum_{k_M}{}' |\Psi|^{1/2} \int \prod_{i=1}^{M} |\chi_{k_i}^{N_{k_i}}| \times \exp\left\{-\tfrac{1}{2}v^{*\prime}\left(\tilde{\Psi} - \varepsilon I_{2GJM}\right)v^*\right\} dv^*\right).$$

In view of (5.17), $|\Psi| = O(1)$, while $\frac{1}{2}v^*(\tilde{\Psi} - \varepsilon I_{2GJM})v^* > \frac{1}{2}\eta\|v^*\|^2$ for some $\eta > 0$. Because $\|\chi_k\| \le (J/m)^{1/2}|a_k|\,\|\nu\|\,\|U_k\|$, we deduce from finiteness of moments of all order of the log of a chi-squared variate that (5.23) is $o(m^{M-N/2-r/2}) \to 0$ on choosing $r = \max(2M - N, 0)$. Now (5.22) makes a contribution only when such $r \ge 2$, which occurs only when $2M - N \ge 2$. Let D be the number of N_{k_i} which equal 1. Clearly $D \ge 2M - N$, that is, $D > t$ for $t = 1, \ldots, r - 1 = 2M - N - 1$ in (5.22). Note that $v^{*\prime}\overline{\Psi}v^*$ is bilinear in the $v_{k_i}^*$ and for each $t = 1, \ldots, r - 1$, consider the following two possible circumstances. In the first, $(v^{*\prime}\overline{\Psi}v^*)^t$ is an odd function of the elements of $v_{k_i}^*$ for at least one k_i. Then because U_k and $\varphi_{2GJM}(\tilde{\Psi}^{1/2}v^*)$ are even in the $v_{k_i}^*$ and the integral is well defined, it follows that there is a zero contribution to the tth summand of (5.22). The other possibility is that $(v^{*\prime}\overline{\Psi}v^*)^t$ is an even function of all elements of v^*. Thus it cannot involve more than t of the $v_{k_i}^*$. The corresponding t or fewer k_i can overlap with the D k_i for which $N_{k_i} = 1$, but because $D > r - 1$, the $(k_1, \ldots, k_M, t)$th summand in (5.22) can be written (taking $N_{k_1} = \cdots = N_{k_{D-t}} = 1$ without loss of generality)

$$(5.24)\quad |\Psi|^{1/2} \prod_{i=1}^{D-t} \left\{\int \chi_{k_i}\varphi_{2GJ}\left(\Psi_{ii}^{1/2}v_{k_i}^*\right) dv_{k_i}^*\right\}$$

$$(5.25)\quad \times \int \frac{\left(-1/2v^{*\prime}\overline{\Psi}v^*\right)^t}{t!}\left\{\prod_{i=D-t+1}^{M} \chi_{k_i}^{N_{k_i}}\varphi_{2GJ}\left(\Psi_{ii}^{1/2}v_{k_i}^*\right) dv_{k_i}^*\right\}.$$

From (5.18),

$$(5.26)\quad \varphi_{2GJ}\left(\Psi_{ii}^{1/2}v_{k_i}^*\right) = \varphi_{2GJ}\left(\left(I_J \otimes R^{-1/2}\right)v_{k_i}^*\right)\left(1 + o\left(m^{-1/2}\|v_{k_i}^*\|^2\right)\right)$$

uniformly in $v_{k_i}^*$ and $\int \chi_k \varphi_{2GJ}((I_J \otimes R^{-1/2})v_k^*)\, dv_k^* = 0$ because the W_{gk} defined by (3.2) and (3.3) were seen to have zero means. For all positive p and q, uniformly in k,

$$\int \|m^{1/2}\chi_k\|^p \|v_k^*\|^q \varphi_{2GJ}\big((I_J \otimes R^{-1/2})v_k^*\big)\, dv_k^* = O(|a_k|^p),$$

from $\log(x) = O(x^{\varepsilon} + x^{-\varepsilon})$ for any $x > 0$, $\varepsilon > 0$ and finiteness of all moments of normal variates. Thus (5.24) is $o(m^{-(D-t)/2}\prod_{i=1}^{D-t}|a_{k_i}|)$ and (5.25) is $o(m^{-t/2}\prod_{i=D-t+1}^{M}|a_{k_i}|^{N_{k_i}})$. It follows from the third part of (5.15) that (5.22) is $o(m^{M-N/2-D/2}) \to 0$ as $n \to \infty$. Thus we have shown that (5.21) $\to 0$ as $n \to \infty$. Now from (5.18), $|\Psi| = |R|^{-JM} + o(m^{-1/2})$ and

$$|R|^{-J/2} \int \chi_{k_i}^{N_{k_i}} \varphi_{2GJ}\big(\Psi_{ii}^{1/2} v_{k_i}^*\big)\, dv_{k_i}^* = \mu_{k_i}^{(N_{k_i})}(1 + o(m^{-1/2})),$$

where $\mu_k^{(p)} = |R|^{-J/2}\int \chi_k^p \varphi_{2GJ}((I_J \otimes R^{-1/2})v_k^*)\, dv_k^*$. The difference between (5.20) and

$$\prod_{i=1}^{M} \sum_{k_i}{}' \mu_{k_i}^{(N_{k_i})} \tag{5.27}$$

is readily seen to be $o(m^{M-N/2-(1/2)\max(1,D)})$ using (5.26), and using $M - N/2 - D/2 \le 0$ when $D \ge 1$ and $M \le N/2$ when $D = 0$. However, $\mu_k^{(p)} = (J/m)^{p/2}E(a_k \nu' W_k)^p$, so that by independence of the W_k, (5.27) is

$$E\left\{ \sum_{k_1}{}' \cdots \sum_{k_M}{}' \prod_{i=1}^{M} \left(J^{1/2}\, \frac{a_{k_i} \nu' W_{k_i}}{m^{1/2}} \right)^{N_{k_i}} \right\}.$$

Because N, M and $k_1, \dots, k_M$ are arbitrary, we have just shown that the moments of $\sum_k' \chi_k$ differ negligibly from those of the variate $(J/m)^{1/2}\sum_k' a_k \nu' W_k$. However, the latter $\to_d N(0, \nu'\Omega\nu)$, on applying (5.15) and the Lindeberg–Feller CLT, in view of the fact that the W_k are iid with zero mean, covariance matrix Ω and finite moments. For reasons indicated in Section 3, this completes the proof of (5.6). To prove (5.13), note that

$$E\|(J/(m-l)) \sum_k{}' U_k U_k' - \Omega\|^2$$

$$= (J/(m-l))^2 \Bigg[\sum_k{}' \operatorname{tr}\{E(U_k U_k' - \Omega)^2\} + \sum_{k_1 \ne k_2}{}' \sum{}' \operatorname{tr}\{(U_{k_1}U_{k_1}' - \Omega)(U_{k_2}U_{k_2}' - \Omega)\} \Bigg]. \tag{5.28}$$

Applying (5.26) with $M = 1$ and arguing as before indicates that $E(U_k U_k' - \Omega)^2 = E(W_k W_k' - \Omega)^2 + o(m^{-1/2}) = O(1)$ uniformly. Employing (5.22) with $M = r = 2$ and arguing as before indicates that for $k_1 \ne k_2$, $E(U_{k_1}U_{k_1} - \Omega)(U_{k_2}U_{k_2} - \Omega) = o(m^{-1})$ uniformly. It follows that (5.28) is $O(m^{-1})$. From

Assumption 6, $(\log n)^4/m = o(m/l^2) = o((m^{1/2} \log m/l)^2) \to 0$, so that (5.13) is true. □

REFERENCES

AGIAKLOGLOU, C., NEWBOLD, P. and WOHAR, M. (1993). Bias in an estimator of the fractional difference parameter. *J. Time Ser. Anal.* **14** 235–246.

BRILLINGER, D. R. (1975). *Time Series, Data Analysis and Theory*. Holden-Day, San Francisco.

CHEUNG, Y.-W. (1993). Long memory in foreign-exchange rates. *Journal of Business and Economic Statistics* **11** 93–101.

CHEUNG, Y.-W. and LAI, K. S. (1993). A fractional cointegration analysis of purchasing power parity. *Journal of Business and Economic Statistics* **11** 103–112.

DELGADO, M. A. and ROBINSON, P. M. (1993). Optimal spectral bandwidth for long memory. Preprint.

DELGADO, M. A. and ROBINSON, P. M. (1994). New methods for the analysis of long memory time series: Application to Spanish inflation. *Journal of Forecasting* **13** 97–107.

DIEBOLD, F. X. and RUDEBUSCH, G. D. (1989). Long memory and persistence in aggregate output. *Journal of Monetary Economics* **24** 189–209.

FOX, R. and TAQQU, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14** 517–532.

GEWEKE, J. and PORTER-HUDAK, S. (1983). The estimation and application of long memory time series models. *J. Time Ser. Anal.* **4** 221–238.

GIRAITIS, L. and SURGAILIS, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate. *Probab. Theory Related Fields* **86** 87–104.

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

HASSLER, U. (1993a). Regression of spectral estimators with fractionally integrated time series. *J. Time Ser. Anal.* **14** 369–380.

HASSLER, U. (1993b). Corrigendum: The periodogram regression. *J. Time Ser. Anal.* **14** 549.

HURVICH, C. M. and BELTRAO, K. I. (1993). Asymptotics for the low-frequency ordinates of the periodogram of a long memory time series. *J. Time Ser. Anal.* **14** 455–472.

HURVICH, C. M. and BELTRAO, K. I. (1994). Acknowledgement of priority for "Asymptotics for the low-frequency ordinates of the periodogram of a long-memory time series."*J. Time Ser. Anal.* **15** 64.

JANACEK, G. J. (1982). Determining the degree of differencing for time series via the log spectrum. *J. Time Ser. Anal.* **3** 177–183.

JOHNSON, N. L. and KOTZ, S. (1970). *Continuous Univariate Distributions—I*. Wiley, New York.

KASHYAP, R. and EOM, K. (1988). Estimation in long-memory time series model. *J. Time Ser. Anal.* **9** 35–41.

KÜNSCH, H. R. (1986). Discrimination between monotonic trends and long-range dependence. *J. Appl. Probab.* **23** 1025–1030.

LEE, D. K. C. and ROBINSON, P. M. (1993). Semiparametric exploration of long memory in stock prices. *J. Statist. Plann. Inference* To appear.

LOBATO, I. and ROBINSON, P. M. (1993). Averaged periodogram estimation of long memory. *J. Econometrics*. To appear.

LOEVE, M. (1977). *Probability Theory I*. Springer, Berlin.

PORTER-HUDAK, S. (1990). An application of the seasonal fractionally differenced model to the monetary aggregates. *J. Amer. Statist. Assoc.* **85** 338–344.

ROBINSON, P. M. (1994a). Semiparametric analysis of long-memory time series. *Ann. Statist.* **22** 515–539.

ROBINSON, P. M. (1994b). Rates of convergence and optimal spectral bandwidth for long range dependence. *Probab. Theory Related Fields*. **99** 443–473.

ROBINSON, P. M. (1994c). Times series with strong dependence. In *Advances in Econometrics. Sixth World Congress* (C. A. Sims, ed.) **1** 47–96. Cambridge Univ. Press.

SHEA, G. S. (1991). Uncertainty and implied variance bounds in long-memory models of the interest rate term structures. *Empirical Economics* **16** 287–312.

SMITH, R. L. (1989). Discussion of "Space-time modeling with long-memory dependence assessing Ireland's wind power resource," by J. Haslett and A. E. Raftery. *J. Roy. Statist. Soc. Ser. C* **38** 1–50.

SMITH, R. L. (1993). Long-range dependence and global warming. In *Statistics of the Environment* (V. Barnett and K. Feridun Turkman, eds.) 141–161. Wiley, New York.

TSCHERNIG, R. (1993). Long memory in foreign exchange rates revisited. Preprint.

YAJIMA, Y. (1989). A central limit theorem of Fourier transforms of strongly dependent stationary processes. *J. Time Ser. Anal.* **10** 375–383.

YONG, C. H. (1974). *Asymptotic Behaviour of Trigonometric Series*. Chinese Univ. Hong Kong.

ZYGMUND, A. (1977). *Trigonometric Series*. Cambridge Univ. Press.

LONDON SCHOOL OF ECONOMICS
HOUGHTON STREET
LONDON WC2A 2AE
UNITED KINGDOM