ROBUST COVARIANCE MATRIX ESTIMATION: "HAC" Estimates with Long Memory/Antipersistence Correction

P.M. Robinson*
London School of Economics

July 8, 2004

Abstract

Smoothed nonparametric estimates of the spectral density matrix at zero frequency have been widely used in econometric inference, because they can consistently estimate the covariance matrix of a partial sum of a possibly dependent vector process. When elements of the vector process exhibit long memory or antipersistence such estimates are inconsistent. We propose estimates which are still consistent in such circumstances, adapting automatically to memory parameters that can vary across the vector and be unknown.

*Research supported by ESRC Grant R000239936. Corresponding address: Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: p.m.robinson@lse.ac.uk

1 INTRODUCTION

We discuss a form of "automated" inference that extends a familiar feature of modern econometric practice to incorporate a flexible form of modelling which has attracted considerable recent interest. "Heteroscedasticity-and-autocorrelation-consistent" (HAC) covariance matrix estimation is commonly employed in inference based on statistics that involve a partial sum of vector-valued random variables that are not assumed serially uncorrelated or homoscedastic; "long run" covariance matrix estimation is another name for the same kind of procedure. Such statistics do not themselves attempt to correct for supposed autocorrelation or heteroscedasticity, but rather the aim is to robustify inference. Popular econometric references include Newey and West (1987), Andrews (1991), and the methods go back to earlier statistical references, such as Jowett (1955), Hannan (1957), Brillinger (1979). The autocorrelation typically presumed is "$I(0)$", in the sense that, for homoscedastic, covariance stationary processes, there is a finite and positive definite spectral density at zero frequency. These properties fail in case of long memory or antipersistent processes, and the usual HAC estimates are then inconsistent, leading to asymptotically invalid tests and inconsistent interval estimates.

We robustify the estimates to ensure consistency in the event of long memory or antipersistence. It is not required that we know whether either of these features pertains, and consistency in the $I(0)$ case is preserved. We deal with a vector process whose components can have memory parameters that are possibly different and unknown.

The following section briefly discusses HAC estimation that presumes $I(0)$ behaviour. Section 3 develops our robustified version. The paper stresses methods, avoiding detailed regularity conditions or proofs.

2 COVARIANCE MATRIX ESTIMATION FOR $I(0)$ SERIES

Consider a $p \times 1$ vector-valued sequence x_t, $t = 0, \pm 1, ...$ For the purpose of a concise discussion we take the elements of x_t to be jointly covariance stationary, later mentioning possible departures. We assume x_t has zero mean and absolutely continuous spectral distribution matrix. Defining the autocovariance matrices

$$\gamma(j) = Ex_0 x_j', \quad j = 0, \pm 1, ...,$$

the spectral density matrix $f(\lambda)$, $\lambda \in (-\pi, \pi]$, is given by

$$\gamma(j) = \int_{-\pi}^{\pi} f(\lambda) e^{ij\lambda} d\lambda,$$

and is Hermitian non-negative definite.

For $n \geq 1$ define the arithmetic mean

$$\bar{x} = n^{-1} \sum_{t=1}^{n} x_t.$$

The covariance matrix of $\bar{x}$ is

$$E\bar{x}\bar{x}' = n^{-1} \left[\gamma(0) + \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \{\gamma(j) + \gamma(-j)\} \right]. \tag{2.1}$$

If $f(\lambda)$ is continuous at $\lambda = 0$, Fejér's theorem indicates that

$$nE(\bar{x}\bar{x}') \to 2\pi f(0), \quad \text{as } n \to \infty. \tag{2.2}$$

Under a variety of additional conditions, $n^{\frac{1}{2}}\bar{x}$ satisfies a central limit theorem, so that, if $f(0)$ is also positive definite,

$$n^{\frac{1}{2}}\bar{x} \to \mathcal{N}(0, 2\pi f(0)), \quad \text{as } n \to \infty. \tag{2.3}$$

Note that

$$2\pi f(0) = \sum_{j=-0}^{\infty} \gamma(j).$$

Large sample inference based on $\bar{x}$ thus requires consistent estimates of $f(0)$. These could result from an assumed parametric model for $\gamma(j)$, $j = 0, \pm 1, ...$, or, equally for $f(\lambda)$, $\lambda \in (-\pi, \pi]$, an obvious example being a stationary and invertible autoregressive moving average process of prescribed orders. However, if either of the orders is under-specified, or both are over-specified, $f(0)$ will be inconsistently estimated.

As the Weierstrass approximation theorem hints, this theoretical drawback can be overcome if the autoregressive or moving average orders are regarded as increasing, slowly, with sample size n. In particular Berk (1974) justified the consistency of autoregression-based spectral density estimates. The autoregressive order can here be thought of as a smoothing number. Such estimates have been employed and modified in the HAC econometric literature, but this has been more influenced by spectral density estimates developed still earlier in the statistical literature, entailing an alternative form of smoothing, and based on quadratic functions of the data, in particular weighted autocovariance spectral estimates (see e.g. Grenander and Rosenblatt, 1957, Parzen, 1957). We shall discuss instead a closely related class of quadratic estimate, stressed by Brillinger (1975), which is not much used by econometricians in the HAC context, but yields more conveniently to the necessary modifications required in the following section, than weighted autocovariance forms which have already been extensively discussed in the econometric literature.

Define the periodogram matrix

$$I(\lambda) = \frac{1}{2\pi n} \left(\sum_{t=1}^{n} x_t e^{it\lambda} \right) \left(\sum_{t=1}^{n} x_t e^{-it\lambda} \right)'. \tag{2.4}$$

For an integer $m \in [1, n/2]$, introduce a sequence of non-negative weights w_{jm}, $j = 0, ..., m$, such that $w_{-j,m} = w_{jm}$ and $\sum_{j=-m}^{m} w_{jm} = 1$. Define

$$\hat{f}(0) = \sum_{j=-m}^{m} w_{jm} I(\lambda_j), \tag{2.5}$$

for $\lambda_j = 2\pi j/n$. The simplest version of (2.5) takes equal weights, $w_{jm} = 1/(2m+1)$. Under suitable conditions on $\{x_t\}$, on the w_{jm}, and on m (such that m increases with n but more slowly), we have

$$\hat{f}(0) \to_p f(0), \quad \text{as } n \to \infty.$$

Various rules have been suggested for choosing the bandwidth m, possibly to satisfy some optimality criterion, such as cross-validation, as well as rules of thumb. Optimality theory for choice of the w_{jm} is also available. The estimates considered by Brillinger (1975) are more general than (2.5), allowing weighted summation over all Fourier frequencies λ_j. However, the weights must again concentrate around zero to an extent that increases slowly with sample size, and the form (2.5) fits in conveniently with that of narrow-band estimates of memory parameters, which have predominated in the semiparametric memory estimation literature relevant to the following section. For each choice of weights $\{w_{jm}\}$ one can effectively find a choice of lag weights, approximately related to the w_{jm} by Fourier transformation, that can be employed in a corresponding weighted autocovariance spectral estimate of $f(0)$, which typically has very similar asymptotic properties to those of its weighted periodogram twin (2.5). Note that the stated conditions on the w_{jm} guarantee that $\hat{f}(0)$ is non-negative definite. It is possible to refine (2.5) by employing different bandwidths and weights across the elements, though the non-negative definite property is less easy to enforce. Refinements such as pre-whitening and tapering are also available, to reduce bias in $\hat{f}(0)$ due to "leakage" from remote frequencies.

The description "HAC" appears to stress "heteroscedasticity" at least as much as "autocorrelation" but whereas there is explicit allowance for the latter in $\hat{f}(0)$ and rival estimates, there is none for the former, and the robustness to heteroscedasticity essentially just appeals to long-standing limit theorems for non-identically distributed variates. For example in the special case of serially uncorrelated x_t, such that $Ex_t x_t' = \Omega_t$, suppose $\bar{\Omega} = n^{-1} \sum_{t=1}^{n} \Omega_t \to \Omega$ as $n \to \infty$. Then from (2.5), $E\hat{f}(0) = (2\pi)^{-1} \sum_{j=-m}^{m} w_{jm} \bar{\Omega} \to (2\pi)^{-1}\Omega$. Even the usual covariance matrix estimate motivated by uncorrelated, homoscedastic variates, $\hat{\Omega} = n^{-1} \sum_{t=1}^{n} x_t x_t'$, satisfies $E\hat{\Omega} = \bar{\Omega} \to \Omega$ as $n \to \infty$, and so can also be called "heteroscedasticity-consistent". The econometric HAC literature has stressed mixing conditions, and extensions thereof, which are designed mainly to describe dependence but also allow a degree of heterogeneity. It would be possible to allow for such heterogeneity in the discussion of the following section, but because again no explicit correction for heteroscedasticity is involved we prefer the simplicity of presentation gained by maintaining the covariance stationarity assumption.

3 COVARIANCE MATRIX ESTIMATION FOR NON-$I(0)$ SERIES

Of crucial importance in the preceding discussion was the $I(0)$ assumption, that $f(\lambda)$ be continuous and positive definite at $\lambda = 0$. To relax this requirement, suppose that

$$f(\lambda) \sim h(\lambda) G \overline{h(\lambda)}, \quad \text{as } \lambda \to 0+, \tag{3.1}$$

where G is a finite, positive definite matrix with (a, b)-th element g_{ab}

$$h(\lambda) = diag\left\{ e^{\frac{id_1\pi}{2}} \lambda^{-d_1}, ..., e^{\frac{id_p\pi}{2}} \lambda^{-d_p} \right\}, \tag{3.2}$$

for $d_j \in (-\frac{1}{2}, \frac{1}{2})$, $j = 1, ..., p$, the overbar means complex conjugation, and $\sim$ means that the ratio of real parts, and of imaginary parts, of corresponding elements of the matrices on the left and right hand sides of (3.1) tends to 1. If $d_1 = ... = d_p = 0$, (3.1) holds with $G = f(0)$ under the $I(0)$ assumption. Slightly more generally, if $d_a + d_b = 0$, for some a, b, the (a, b)-th element of $f(\lambda)$, $f_{ab}(\lambda)$, satisfies $f_{ab}(0) = g_{ab} \cos \frac{\pi}{2}(d_a - d_b)$, which can again be consistent with $f_{ab}(\lambda)$ being continuous at $\lambda = 0$. For $d_a + d_b > 0$, on the other hand, $f_{ab}(\lambda)$ diverges as $\lambda \to 0+$, whilst for $d_a + d_b < 0$, $f_{ab}(0) = 0$. Of course when $a \neq b$ $f_{ab}(0) = 0$ also occurs, for any d_a, d_b, if $g_{ab} = 0$. The d_j are called "memory parameters".

To motivate (3.1), (3.2), Theorem III-1 of Yong (1974) gives

$$(1 - e^{i\lambda})^d \sim e^{-\frac{id\pi}{2}} \lambda^d, \quad \text{as } \lambda \to 0+. \tag{3.3}$$

The left hand side is the frequency response function of the fractional difference operator. An important special case of $f(\lambda)$ satisfying (3.1) is the spectral density matrix of a stationary, non-cointegrated and invertible fractionally integrated autoregressive moving average system, with possibly distinct memory parameters $d_1, ..., d_p$.

Whilst, in a nonparametric setting, we do not want to impose such a parametric model, nevertheless we need to supplement (3.1), when at least one d_j is non-zero, by an assumption that is easily satisfied in that parametric model. We have to approximate the right side of (2.1) for large n, and this can be achieved by approximating $\gamma(j) + \gamma(-j)$ for large j. For some a, b such that $d_a + d_b \neq 0$, denote by $\gamma_{ab}(j)$ the (a, b)-th element of $\gamma(j)$. For λ close to zero, $f_{ab}(\lambda)$ has real part

$$\mathcal{R}e\{f_{ab}(\lambda)\} = \frac{1}{2\pi}\gamma_{ab}(0) + \frac{1}{2\pi}\sum_{j=1}^{\infty}\{\gamma_{ab}(j) + \gamma_{ab}(-j)\}\cos j\lambda.$$

On the other hand, from (3.1) it follows that

$$\mathcal{R}e\{f_{ab}(\lambda)\} \sim g_{ab}\lambda^{-d_a-d_b}\cos\frac{\pi}{2}(d_a - d_b), \quad \text{as } \lambda \to 0+. \tag{3.4}$$

An important topic in the trigonometric series literature concerns the asymptotic behaviour of Fourier coefficients that provide the power law behaviour found in (3.4), a detailed reference being Yong (1974). Consider a function

$$r(\lambda) = \sum_{j=1}^{\infty} s(j) \cos j\lambda,$$

for λ close to zero. Yong (1974, Theorems III-1, III-10, III-12 and III-17) gave conditions on the $s(j)$ such that, for some $\beta \neq 0$,

$$s(j) \sim \beta j^{-\alpha}, \quad \text{as } j \to \infty, \tag{3.5}$$

is equivalent to

$$r(\lambda) \sim \frac{\beta\pi}{2\Gamma(\alpha)\cos\frac{\alpha\pi}{2}} \lambda^{\alpha-1}, \quad \text{as } \lambda \to 0+, \tag{3.6}$$

when $\alpha \in (0, 1)$. Yong (1974, Theorem III-27) showed that if

$$r(0) = 0, \tag{3.7}$$

(3.5) implies (3.6) for $\alpha \in (1, 3)$.

We apply these properties with $s(j) = \gamma_{ab}(j)+\gamma_{ab}(-j)$, $r(\lambda) = 2\pi\mathcal{R}e\{f_{ab}(\lambda)\}$ and $\alpha = 1 - d_a - d_b$. We deduce from (3.4), (3.6), reflection formula for the Gamma function, and trigonometric identities, that

$$\begin{aligned}
\beta &= \frac{2\pi g_{ab}(\sin\pi d_a + \sin\pi d_b)}{\Gamma(d_a + d_b)\sin\pi(d_a + d_b)} \\
&= \frac{2\pi g_{ab}\cos\frac{\pi}{2}(d_a - d_b)}{\Gamma(d_a + d_b)\cos\frac{\pi}{2}(d_a + d_b)} \\
&= 2\pi g_{ab}\Gamma(1 - d_a - d_b)\left\{\frac{1}{\Gamma(d_a)\Gamma(1 - d_a)} + \frac{1}{\Gamma(d_b)\Gamma(1 - d_b)}\right\},
\end{aligned}$$

to give three alternative expressions. Note that if $d_a + d_b < 0$, (3.4) implies $f_{ab}(0) = 0$, so (3.7) is indeed relevant. On the other hand, for $d_a + d_b > 0$, the conditions of Yong (1974) can be checked in case of plausible autocovariance sequences.

We can now deduce that

$$\gamma_{ab}(0) + \sum_{j=1}^{n-1}\left(1 - \frac{j}{n}\right)\{\gamma_{ab}(j) + \gamma_{ab}(-j)\} \sim 2\pi g_{ab} q(d_a, d_b) n^{d_a+d_b}, \tag{3.8}$$

as $n \to \infty$, where

$$q(u, v) = \frac{\sin\pi u + \sin\pi v}{\Gamma(u + v + 2)\sin\pi(u + v)}, \tag{3.9}$$

in which we arbitrarily employ the first and the three equivalent expressions for β above. This follows by approximating sums by integrals, though in case

$d_a + d_b < 0$ (implying $\sum_{j=-\infty}^{\infty} \gamma_{ab}(j) = 0$) one first writes the left side of (3.8) as

$$-\sum_{j=n}^{\infty} \{\gamma_{ab}(j) + \gamma_{ab}(-j)\} - \frac{1}{n}\sum_{j=1}^{n-1} j\{\gamma_{ab}(j) + \gamma_{ab}(-j)\}.$$

We deduce (cf. (2.2))

$$D_n E(\bar{x}\bar{x}')D_n \to 2\pi G \circ Q(d_1, ..., d_p), \quad \text{as } n \to \infty, \tag{3.10}$$

where

$$D_n = diag\left\{n^{\frac{1}{2}-d_1}, ..., n^{\frac{1}{2}-d_p}\right\},$$

$Q(d_1, ..., d_p)$ is the $p \times p$ matrix with (a, b)-th element $q(d_a, d_b)$, and $\circ$ denotes Hadamard product. Since the right side of (3.10) is the limit of a sequence of non-negative definite matrices, it also is non-negative definite. If it is positive definite we can deduce under suitable additional conditions (for example if x_t is a linear process in stationary, conditionally homoscedastic martingale differences)

$$D_n\bar{x} \to_d N\left(0, 2\pi G \circ Q(d_1, ..., d_p)\right) \tag{3.11}$$

(cf. (2.3)).

The above discussion strictly covers only elements such that $d_a + d_b \neq 0$, but (3.8) applies when $d_a + d_b = 0$ for some a, b, because from the second equality of (3.9) it agrees with $f_{ab}(0) = g_{ab}\cos\frac{\pi}{2}(d_a - d_b)$ (which follows from (3.4)). This includes the case $d_a = d_b = 0$, and in the full $I(0)$ case $d_1 = ... = d_p = 0$, (3.10) reduces to (2.2). Thus (3.10) generalizes (2.2), indeed $q(u, v)$ is continuous at $u + v = 0$ (and all $u, v \in (-\frac{1}{2}, \frac{1}{2})$).

Given estimates $\hat{d}_1, ..., \hat{d}_p$ and $\hat{G}$ such that

$$(\log n)\left(\hat{d}_j - d_j\right) \to_p 0, \quad j = 1, ..., p, \quad \hat{G} \to_p G, \tag{3.12}$$

we can replace (3.11) by the useful result

$$\left\{2\pi\hat{G} \circ Q\left(\hat{d}_1, ..., \hat{d}_p\right)\right\}^{-\frac{1}{2}} \hat{D}_n\bar{x} \to_d N(0, I_p),$$

where $\hat{D}_n = diag\left\{n^{\frac{1}{2}-\hat{d}_1}, ..., n^{\frac{1}{2}-\hat{d}_p}\right\}$ and I_p is the $p \times p$ identity matrix. The rate requirement in (3.12) is due to the need to approximate the norming factors $n^{\frac{1}{2}-d_j}$ by the $n^{\frac{1}{2}-\hat{d}_j}$.

An acronym has become almost obligatory. The best that emerged to describe our robustified variance estimate of $\bar{x}$,

$$\hat{D}_n^{-1}\left\{2\pi\hat{G} \circ Q(\hat{d}_1, ..., \hat{d}_p)\right\}\hat{D}_n^{-1}, \tag{3.13}$$

was MAC: Memory-Autocorrelation-Consistent. Partly for reasons given at the end of the previous section, and partly for the sake of an acronym that slips easily off the tongue, reference to "H" for "heteroscedasticity" is suppressed.

Robinson (1998) proposed covariance matrix estimates for parameter estimates in regression models in which regressors and disturbances can satisfy a condition like (3.1), and need thus not be $I(0)$. His estimates are nonparametric but under his conditions do not require a user-chosen bandwidth. However he required the product of regressors and disturbances to be $I(0)$.

Estimates of the $\hat{d}_j$ satisfying (3.12) under suitable conditions are readily available, such as log periodogram, local Whittle and averaged periodogram estimates, all of which are "semiparametric" in character, being based principally on the local–to-zero model (3.1). Like (2.5), they involve functions of the periodogram $I(\lambda_j)$ at low frequencies such that $j = 1-m, ..., m-1$, with m satisfying rather similar conditions to those that would be required for (2.5) in Section 2. As a result, the $\hat{d}_j$ estimates converge more slowly than the $n^{\frac{1}{2}}$ parametric rate, but nevertheless the slow rate in (3.12) is easily justified.

With respect to $\hat{G}$, in view of (2.5), (3.1) and (3.2), it is natural to consider

$$\hat{G} = \sum_{j=1-m}^{m-1} w_{jm} \hat{h}(\lambda_j)^{-1} I(\lambda_j) \overline{\hat{h}(\lambda_j)}^{-1}, \tag{3.14}$$

where

$$\hat{h}(\lambda) = diag\left\{ e^{\frac{i\hat{d}_1\pi}{2}} \lambda^{-\hat{d}_1}, ..., e^{\frac{i\hat{d}_p\pi}{2}} \lambda^{-\hat{d}_p} \right\}.$$

Under conditions familiar from the semiparametric memory parameter estimation literature, (3.14) will satisfy (3.12). To make the procedure more fully automatic, rules for choice of m in (3.14), and of bandwidths in the $\hat{d}_a$ estimates, are required. These issues have been discussed in the literature (see e.g. Hurvich and Beltrao (1994), Robinson (1994a), Hurvich, Deo and Brodsky (1998), Henry and Robinson (1996), Hurvich and Deo (1999).

The "MAC" estimate (3.13), with $\hat{G}$ given by (3.14), is guaranteed non-negative definite. To see this, note first that (3.14) is non-negative definite. Thus it follows from Schur (1911, p.14) that it suffices to show that $Q(d_1, ..., d_p)$ is non-negative definite, for all $d_j \in (-\frac{1}{2}, \frac{1}{2})$, $j = 1, ..., p$. But from the previous development it is clear that

$$Q(d_1, ..., d_p) = \lim_{n\to\infty} \frac{1}{2\pi n} D_n \left[\gamma^*(0) + \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \{\gamma^*(j) + \gamma^*(-j)\} \right] D_n, \tag{3.15}$$

where $\gamma^*(j) = \int_{-\pi}^{\pi} f^*(\lambda) e^{ij\lambda} d\lambda$, such that $f^*(\lambda)$ satisfies (3.1) with $g_{ab} = 1$, all a, b (so G is now taken to be non-negative definite, with rank 1) and $\gamma^*(j)$ has the same asymptotic behaviour as $\gamma(j)$ with $g_{ab} = 1$, all a, b. But the term in square brackets in (3.15) is $n^{-1} \int_{-\pi}^{\pi} f^*(\lambda) \left| \sum_{t=1}^{n} e^{it\lambda} \right|^2 d\lambda$, which is clearl non-negative definite, for all n, since $f^*(\lambda)$ can be chosen non-negative definite, for all λ.

Even in the expectation that all d_j are zero, MAC estimates might be useful rivals to long autoregressive and weighted autocovariance (or periodogram) HAC estimates, these latter having the reputation of being appropriate in the presence of both (finite) peaks and (non-zero) troughs in $f(\lambda)$ at $\lambda = 0$.

The present topic was also discussed in case of scalar x_t by Beran (1989), Robinson (1994b). The first author, however, employed parametric memory parameter estimates, whereas the second author employed semiparametric averaged periodogram memory parameter estimates and noted the need for the rate condition in the first part of (3.12). The estimate (3.14) extends one of Robinson (1995) in the scalar case, in which $w_{jm} \equiv 1/(2m+1)$, and was employed in a different context by Robinson and Yajima (2002). Robinson (1994b) considered covariance matrix estimates in the case $x_t = (1, t, ..., t^{p-1})e_t$, where e_t is a scalar long memory process, this being relevant to inference on least squares estimates in polynomial-in-time linear regression with long memory or antipersistent disturbances. For the same kind of disturbances, Robinson (1997) considered covariance matrix estimates for a vector of scalar fixed-design nonparametric regression estimates at finitely many fixed points. The rates of convergence in these situations, and the forms of the limiting covariance matrices of the normalized statistics, differ from those found in the present paper, which is motivated by other situations in econometrics. In a fractional cointegration context, in which two distinct memory parameters are involved, Kim and Phillips (1999) proposed estimates of the long run covariance matrix, in which stationary fractional sequences are filtered in the time domain and the resulting $I(0)$ long run covariance matrix estimate rescaled.

Many econometric statistics are functionals of partial sums of vector variates, which themselves can be products of other nonlinear functions of underlying variates, for example generalized-method-of-moments estimates, including least squares estimates for linear regression models with stochastic regressors. Consider an estimate $\hat{\theta}$ of a vector-valued parameter θ of dimension no greater than p. Typically we can consider a linearization $\hat{\theta} - \theta = T_n \bar{x}$, where, when $\hat{\theta}$ is only implicitly-defined, this requires an initial consistency proof (which should itself allow for possible long memory or antipersistence), and T_n is a matrix-valued statistic. If, for some matrix E_n, $E_n T_n D_n$ converges in probability to a finite limit U of full row rank, we would deduce from (3.11) that $E_n(\hat{\theta} - \theta) \to_d \mathcal{N}(0, 2\pi U (G \circ Q(d_1, ..., d_p)) U')$ whence the previous discussion is relevant. However, in case the d_j are not all identical, lack of commutativity can obstruct this argument; the analysis could be preceded by a test of equality of the d_j, employing known limit distribution theory for semiparametric memory estimates. In this kind of setting, moreover, when a typical element of x_t is a nonlinear function of underlying variates such as a product of an explanatory variable and a disturbance, it is important to bear the following in mind. If some or all the underlying variates have long memory, it is still possible at one extreme that x_t can be $I(0)$, and on the other that $\bar{x}$ has a non-normal limit distribution (see Robinson, 1994c). Note also that disturbances will have to be replaced by residuals in order to produce proxies for x_t that can be used in the estimation of the d_j.

References

[1] Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817-858.

[2] Beran, J. (1989) A test of location for data with slowly decaying serial correlation. *Biometrika* 76, 261-269.

[3] Berk, K.N. (1974) Consistent autoregressive spectral estimates. *Annals of Statistics* 2, 489-502.

[4] Brillinger, D.R. (1975) *Time Series. Data Analysis and Theory.* San Francisco: Holden-Day.

[5] Brillinger, D.R. (1979) Confidence intervals for the cross-covariance function. *Selecta Statistica Canadiana* 5, 3-16.

[6] Grenander, U. & M. Rosenblatt (1957) *Statistical Analysis of Stationary Time Series.* New York: Wiley.

[7] Hannan, E.J. (1957) The variance of the mean of a stationary process. *Journal of the Royal Statistical Society* Series B 19, 282-285.

[8] Henry, M. & P.M. Robinson (1996) Bandwidth choice in Gaussian semiparametric estimation of long range dependence. In P.M. Robinson & M. Rosenblatt (eds.), *Athens Conference on Applied Probability and Time Series Analysis, Vol. II: Time Series Analysis*, pp.220-232. New York: Springer-Verlag.

[9] Hurvich, C.M. & K.I. Beltrao (1994) Automatic semiparametric estimation of the memory parameter of a long-memory time-series. *Journal of Time Series Analysis* 15, 285-302.

[10] Hurvich, C.M. & R. Deo (1999) Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long-memory time series. *Journal of Time Series Analysis* 20, 331-341.

[11] Hurvich, C.M., R. Deo & J. Brodsky (1998) The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long memory time series. *Journal of Time Series Analysis* 19, 19-46.

[12] Jowett, G.H. (1955) The comparison of means of sets of observations from sections of independent stochastic series. *Journal of the Royal Statistical Society* Series B 17, 208-227.

[13] Kim, C.S. & P.C.B. Phillips (1999) Fully modified estimation of fractional cointegration models. Preprint.

[14] Newey, W.K. & K.D. West (1987) A simple, positive semidefinite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 56, 203-208.

[15] Parzen, E. (1957) On consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics* 28, 329-348.

[16] Robinson, P.M. (1994a) Rates of convergence and optimal spectral bandwidth for long range dependence. *Probability Theory and Related Fields* 99, 443-473.

[17] Robinson, P.M. (1994b) Semiparametric analysis of long memory time series. *Annals of Statistics* 22, 515-539.

[18] Robinson, P.M. (1994c) Time series with strong dependence. Invited Paper, 1990 World Congress of the Econometric Society. In C.A. Sims (ed.), *Advances in Econometrics: Sixth World Congress*, Volume 1, 47-96. Cambridge: Cambridge University Press.

[19] Robinson, P.M. (1995) Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* 23, 1630-1661.

[20] Robinson, P.M. (1997) Large-sample inference for nonparametric regression with dependent errors. *Annals of Statistics* 28, 2054-2083.

[21] Robinson, P.M. (1998) Inference-Without-Smoothing in the Presence of Nonparametric Autocorrelation. *Econometrica* 66, 1163-1182.

[22] Robinson, P.M. & Y. Yajima (2002) Determination of cointegrating rank in fractional systems. *Journal of Econometrics* 106, 217-245.

[23] Schur, J. (1911) Bemarkungen zur Theorie der beschrämltem Bilinearforman mit unedlich vielers Veränderlichen. *Journal Reine Angewandte Matematik* 140, 1-28.

[24] Yong, C.-H. (1974) *Asymptotic Behaviour of Trigonometric Series*. Hong Kong: Chinese University of Hong Kong.