

Long Memory in Stock-Market Trading Volume



Ignacio N. LOBATO

Centro de Investigación Económica, Instituto Tecnológico Autónomo de México, México D.F. 10700, México (*ilobato@itam.mx*)

Carlos VELASCO

Departamento de Econometría y Estadística, Universidad Carlos III de Madrid, 28911 Leganés (Madrid), Spain (*cavelas@est-econ.uc3m.es*)

This article examines consistent estimation of the long-memory parameters of stock-market trading volume and volatility. The analysis is carried out in the frequency domain by tapering the data instead of detrending them. The main theoretical contribution of the article is to prove a central limit theorem for a multivariate two-step estimator of the memory parameters of a nonstationary vector process. Using robust semiparametric procedures, the long-memory properties of trading volume for the 30 stocks in the Dow Jones Industrial Average index are analyzed. Two empirical results are found. First, there is strong evidence that stock-market trading volume exhibits long memory. Second, although it is found that volatility and volume exhibit the same degree of long memory for most of the stocks, there is no evidence that both processes share the same long-memory component.

KEY WORDS: Detrending; Long-range dependence; Nonstationary processes; Semiparametric inference; Tapering; Volatility.

The analysis of the long-term dependence of economic time series is an outstanding statistical problem. Traditionally, to study persistence economists have emphasized the use of unit-root models if stationarity was not assumed and, more recently, the use of autoregressive models with roots close to 1 when covariance stationarity was assumed. In the last years, the notion of long memory has attracted the attention of economists trying to model and measure the persistence of stationary processes. For covariance stationary processes, long memory focuses on the behavior of the autocovariance sequence at long lags or on the behavior of the spectral density function in a neighborhood of zero frequency. Because many economic time series appear to be nonstationary in different degrees, there has been interest in extending the concept of long memory to nonstationary processes (e.g., see Hurvich and Ray 1995; Gil-Alaña and Robinson 1997; Robinson and Marinucci 1998; Velasco 1999a,b).

In both the stationary and the nonstationary cases, long memory characterizes the stochastic long-term dependence by one parameter d . In the stationary case, Robinson (1994a; 1995a,b) recently introduced consistent estimators of d for a broad class of long-memory processes. For nonstationary data, some of these estimators are still consistent under some conditions (see Velasco 1999a,b), but typically nonstationary data should be converted to stationary—for instance, by differencing—before applying these estimators.

In practice, the estimation of d is additionally complicated by the potential presence of deterministic trends. In this case, an obvious approach to estimate d would involve detrending the data to obtain a stationary sequence and then **applying any of Robinson's estimators. This procedure cannot guarantee the consistent estimation of d , however, unless the researcher possesses a priori information about the memory of the process and about the true form of the trend.**

In case this information is unavailable or imprecise, it is still possible to estimate d consistently in the frequency domain by means of tapering the data. Velasco (1999a,b) showed that tapering can be used to estimate d consistently without assuming if the process is stationary or not and without specifying and estimating any trend. The only requirement to apply the tapering procedure is to assume an upper bound for the memory of the process and for the order of the deterministic trend.

When the interest resides in vector observations rather than in scalars, model specification is additionally complicated. The generating mechanism of each economic variable can be radically different, and misspecification of the trend of a particular series may lead to the inconsistent estimation of the memory parameters of the whole vector. Again, as will be shown in this article, the use of tapering allows a consistent estimation of the memory parameters for a wide class of nonstationary vector processes.

The theoretical research of this article was motivated by an empirical example in which the two components of the vector process of interest apparently exhibit different stochastic properties. The two variables are volatility (measured by the absolute value of the return) and trading volume in the stock market. Even though volatility is considered to be stationary, whereas volume is treated as nonstationary (see Gallant, Rossi, and Tauchen 1992; Andersen 1996; Bollerslev and Jubinski 1999), some stochastic properties of both series, in particular the long-term dependence, could be very similar. Bollerslev and Jubinski (1999) considered a mixture-of-distribution hypothesis model, in which the latent information-arrival process ex-

hibits long memory, and derived that volatility and volume exhibit long-memory dependence of the same degree. Because volume is nonstationary, analyzing its long-memory properties is complicated. Bollerslev and Jubinski (1999) addressed this issue by linearly detrending the data. The resulting estimators of the long-memory parameters have unknown properties, however, as is illustrated in Section 2. Furthermore, it is a priori questionable that the non-deterministic part of volume is stationary and thus that its long-memory parameter belongs to the stationary region ($d < .5$). We employ a tapering procedure that overcomes these two difficulties.

This article makes two contributions. The first is to establish a multivariate central limit theorem (CLT) for a two-step semiparametric estimator of the long-memory parameters of a nonstationary vector process. This two-step estimator is based on a modification of the objective function considered by Lobato (1999) to allow for a tapering procedure as given by Velasco (1999b). The proposed statistic consistently estimates the long-memory parameters for a wide class of nonstationary processes with any degree of long-memory dependence (even when $d > .5$) or any order of polynomial trend if appropriate tapering is applied.

The second contribution of the article is to analyze the long-memory properties for the volume and return volatility processes of the 30 stocks that compose the Dow Jones Industrial Average (DJIA) index. We employ daily data for the period 1962–1994. Using the previous theory, we establish two empirical findings. First, stock-market volume exhibits long memory for the 30 stocks in the DJIA index. Second, for most stocks it can be concluded that return volatility and trading volume possess the same long-memory parameter, although apparently the long memory of the volatility and of the volume series cannot be explained by a common long-memory component.

The outline of the article is the following. In Section 1, the concept of long memory for stationary and nonstationary processes and the tapering procedure are introduced. Section 2 illustrates how tapering can be employed to analyze the long-memory properties of nonstationary series. In Section 3, the multivariate CLT for a two-step semiparametric estimator of the long-memory parameters of a possibly nonstationary vector process is established and a small Monte Carlo analysis of its finite-sample performance is reported. Section 4 contains the empirical analysis, and Section 5 concludes. The proof of the CLT is in the Appendix.

1. DEFINITIONS

1.1 Long Memory for Stationary Processes

A covariance stationary stochastic process exhibits long memory with memory parameter d when its spectral density function (SDF hereafter) $f(\lambda)$ satisfies

$$f(\lambda) \sim G\lambda^{-2d} \quad \text{as } \lambda \rightarrow 0^+, \quad (1)$$

where G is a finite positive constant and $d \in (-\frac{1}{2}, \frac{1}{2})$ (see Robinson 1994a and references therein). The parameter d governs the degree of memory of the series.

1.2 Long Memory for Nonstationary Processes

By analogy to the unit-root literature, the memory of a nonstationary stochastic process X_t , whose increments ΔX_t are stationary and satisfy (1) for some $d_{\Delta X} \in [-\frac{1}{2}, \frac{1}{2})$, is defined as $d_X = d_{\Delta X} + 1$, $d_X \in [\frac{1}{2}, \frac{3}{2})$. This definition generalizes the $I(d)$ processes terminology, in which an $I(1)$ process—namely, a process whose first-order differences have positive and bounded spectral density at the origin [an $I(0)$ process]—could be said to possess memory parameter $d_X = 1$.

For general nonstationary processes, Velasco (1999a,b) generalized the long-memory definition of Hurvich and Ray (1995). A nonstationary stochastic process $\{X_t\}$ is said to exhibit long memory with memory parameter $d_X = d + s$ if after being integer-differenced s times it is transformed into a stationary stochastic process with spectral density satisfying (1) for some $d \in [-\frac{1}{2}, \frac{1}{2})$. This definition of nonstationary integrated processes is not the only possible; for an alternative definition, see Robinson and Marinucci (1998). Formally, denote the successive differences of X_t as $U_t^{(r)} = \Delta^r X_t$, for $r = 1, 2, \dots, s$, and let $\lfloor z \rfloor$ be the largest integer less than or equal to z . Then the stochastic process X_t has memory parameter $d_X \geq -\frac{1}{2}$ if the stochastic process $U_t^{(s)} = \Delta^s X_t$, $s = \lfloor d_X + \frac{1}{2} \rfloor$, is covariance stationary with mean $\mu = 0$ and its SDF satisfies

$$f_{U^{(s)}}(\lambda) \sim G\lambda^{-2(d_X - s)} \quad \text{as } \lambda \rightarrow 0^+,$$

$$d_X - s \in \left[-\frac{1}{2}, \frac{1}{2}\right), \quad (2)$$

where G is a finite positive constant. Notice that Condition (2) is very general because it only refers to the behavior of the SDF in a neighborhood of zero frequency. In particular, it is not assumed that $f_{U^{(s)}}$ is the spectral density of a stationary and invertible fractional autoregressive integrated moving average process, and it is not ruled out that $f_{U^{(s)}}(\lambda)$ may have (integrable) poles or zeros at frequencies beyond the origin. Notice that for an observed stretch of data $\{X_1, X_2, \dots\}$, setting all initial conditions to 0 and integrating $U_t^{(s)}$, we can write, for $t \geq 1$,

$$X_t = \sum_{j_1=1}^t \sum_{j_2=1}^{j_1} \dots \sum_{j_s=1}^{j_{s-1}} U_{j_s}^{(s)}. \quad (3)$$

The previous definition of memory refers to a process that does not exhibit any deterministic part. If a process is composed of a purely stochastic part with memory d_X , such as the one considered on the right side of (3), and of some additional deterministic trend of order q , the memory of the process is defined to be d_X , the memory of the stochastic component.

The appearance of a polynomial trend can be explained, in first instance, because the mean of $U_t^{(s)} = \Delta^s X_t$ can be different from 0, $\mu \neq 0$. In this case, it is possible to show that a nonstationary stochastic process X_t of memory $d_X = d + s$ can be decomposed as the sum of a polynomial trend of order s and a stochastic process that is the s th-order integration of a long-memory covariance stationary

zero mean process with memory d . Thus, for random variables $U_o^{(r)}$, $r = 1, \dots, s$, that do not depend on time, after repeated integration it is obtained that

$$X_t = X_o + \sum_{r=1}^{s-1} U_o^{(r)} p_U^{(r)}(t) + \mu p_\mu^{(s)}(t) \quad (4)$$

$$+ \sum_{j_1=1}^t \sum_{j_2=1}^{j_1} \dots \sum_{j_s=1}^{j_{s-1}} U_{j_s}^{(s)}, \quad (5)$$

where $U_t^{(s)} = U_t^{(s)} - \mu$ has zero mean and the same spectral density as $U_t^{(s)}$, and $p_U^{(r)}(t)$ and $p_\mu^{(s)}(t)$ are polynomials in t of order r and s , respectively. The right side of (4) is a polynomial in time of order s given by the initial conditions, whereas expression (5) is purely stochastic with memory d_X .

In general, the observed process Y_t can be the sum of the purely nonstationary stochastic process X_t with memory d_X as defined in (3) or (5) and a polynomial trend of order q , T_t ,

$$Y_t = X_t + T_t, \quad (6)$$

where

$$T_t = \sum_{r=0}^q \alpha_r t^r$$

and both processes, Y_t and X_t , have the same memory parameter ($d_Y = d_X$) determined exclusively by the stochastic part.

Although for nonstationary process the SDF does not exist, it is possible to define a generalized or pseudo SDF that behaves like (1) also but adds to d the number of integer differences taken to achieve stationarity. Formally, define the *pseudo spectral density* $f_X(\lambda)$ of X_t as

$$f_X(\lambda) = |1 - \exp(i\lambda)|^{-2s} f_{U^{(s)}}(\lambda) \sim G\lambda^{-2d_X} \quad \text{as } \lambda \rightarrow 0^+, \quad (7)$$

where $f_{U^{(s)}}(\lambda)$ satisfies (2) and $f_X(\lambda)$ can behave as (1) for any $d_X \geq -\frac{1}{2}$, so it is not a spectral density for $d_X \geq \frac{1}{2}$. See Velasco (1999a) for details.

1.3 Frequency-Domain Analysis and Tapering

The basic statistics employed in the time series analysis in the frequency domain are the discrete Fourier transform (DFT) and the periodogram. The DFT of an observed sequence X_t , $t = 1, \dots, n$, where n denotes the sample size, at frequency λ is

$$w_X(\lambda) = (2\pi n)^{-1/2} \sum_{t=1}^n X_t \exp(i\lambda t). \quad (8)$$

For statistical and computational reasons, the DFT is evaluated at the Fourier frequencies $\lambda_j = 2\pi j/n$ (j integer). The periodogram of X_t at frequency λ_j , $I_X(\lambda_j) = |w_X(\lambda_j)|^2$, is the sample equivalent of the SDF for stationary processes,

and it is used to construct consistent parametric and non-parametric estimates of the SDF.

Unfortunately, statistics based on the periodogram suffer from a bias problem caused by leakage, which is especially accentuated for stochastic processes whose SDF's exhibit peaks at some frequencies. Tukey observed that the leakage problem could be attenuated if the spectral estimators were constructed with the data multiplied by a sequence of nonnegative weights rather than with the raw data (see Cooley and Tukey 1965). This sequence of weights, called a "data window" or a "taper," is constructed so that it takes values around 1 for the central part of the data, but it decays smoothly to 0 at both the beginning and the end of the sample. Consequently, the tapered DFT of an observed sequence X_t , $t = 1, \dots, n$, at frequency λ_j is

$$w_X^T(\lambda_j) = \left(2\pi \sum_{t=1}^n h_t^2\right)^{-1/2} \sum_{t=1}^n h_t X_t \exp(i\lambda_j t), \quad (9)$$

for any taper sequence $\{h_t\}_{t=1}^n$. Then, the tapered periodogram at frequency λ_j is $I_X^T(\lambda_j) = |w_X^T(\lambda_j)|^2$. Notice that the usual DFT $w_X(\lambda_j)$ is obtained by setting $h_t \equiv 1$, for all t .

Tapering has been employed in different frameworks by Zhurbenko (1979), Robinson (1986), and Dahlhaus (1988), among others, to carry out statistical inference for stochastic processes that exhibit behavior close to nonstationarity at certain frequencies. These processes are employed to describe the dynamic properties of data whose SDF estimates exhibit very strong peaks at zero frequency or at seasonal or cyclical frequencies. For these data, the usual integer or seasonal differencing may complicate statistical inference because this may lead to noninvertibility.

Several data tapers have been used in the literature (see Brillinger 1975, pp. 54–59). Zhurbenko (1979) used a general class of data weights $\{h_t^{KZ}\}$ suggested by Kolmogorov, based on p convolutions of the uniform density in $1, 2, \dots, n$. When $p = 1$, these are the uniform weights, $h_t \equiv 1$, so no tapering is carried out. When $p = 2$, the taper sequence is given by Bartlett's or triangular window. Although the weights $\{h_t^{KZ}\}$ generally have involved expressions for $p \geq 3$, they can be easily calculated in a computer using the properties of the Fourier transform. If $p \geq 3$, the shape of the taper is similar to the Gaussian probability density because the sum of three or more uniform random variables has an approximate normal distribution. Alekseev (1996) provided a broad discussion and some explicit formulas.

To implement the tapering procedure, the concept of the order p of a taper is fundamental. The order of a taper is a positive integer related to the smoothness of the taper. For a formal definition, see Velasco (1999a). In the next sections, the case $p = 1$ implies that no tapering is carried out. It can be shown that the asymptotic bias of the tapered periodogram for the SDF of stationary processes becomes smaller as the order p of the taper becomes higher. Consequently, the leakage problem can be controlled by using a taper of high enough order. This property extends to nonstationary processes, for which the pseudo SDF f_X defined

in (7) for $d_X \geq \frac{1}{2}$ is now the limit of the expectation of the tapered periodogram. In fact, for the case in which no deterministic trends are present, Velasco (1999a,b) proved that the semiparametric estimators of the memory d proposed by Robinson (1995a,b) are consistent for any d when tapers of order $p > d$ are employed. This result is especially significant because those estimators are consistent only for $d < 1$ when no taper is employed.

2. TAPERING AND LONG-MEMORY NONSTATIONARY PROCESSES

In this section we illustrate how tapering can be employed to estimate robustly the memory parameter of a possibly nonstationary stochastic process. Consider the model given by (6). To make inference about the memory parameter d_Y the standard approach attempts the estimation of d_Y in two stages. In the first stage, the data are transformed to eliminate the deterministic trend. This stage is motivated by the definition of the long-memory parameter that concerns exclusively the stochastic part of a process. The second stage is to estimate the long-memory parameter from the detrended data. Instead of this approach, in this section we propose to estimate d_Y directly from the raw data by means of tapering.

The two typical ways of detrending the data are, first, to estimate the deterministic trend T_t and then subtract it from the original series and, second, to difference the series. Both procedures have serious limitations because an exact knowledge of the true model seems essential to avoid the risks of inappropriate detrending and differencing (e.g., see Chan, Hayya, and Ord 1977; Nelson and Kang 1981, 1984; Durlauf and Phillips 1988).

For instance, it is ignored under what conditions d_Y can be estimated consistently from the residual series $Y_t - \hat{T}_t$ because the statistical properties of the fitted trend \hat{T}_t are likely to depend crucially on the memory parameter d_Y . If the exact shape of the trend is unknown, the first difficulty is to select the order q . For example, one approach would be to carry out a polynomial regression and choose q as the order of the last significant coefficient $\hat{\alpha}_r$, where $\hat{\alpha}_r$ are, for instance, the least squares estimates. Inference based on $\hat{\alpha}_r$ depends on the unknown d_Y , however; see Yajima (1988) for the stationary case and Deo and Hurvich (1998) for nonstationary series with linear trends. With nonparametric detrending methods, such as the one employed by Andersen (1996), similar difficulties arise, although Robinson (1997) proved that consistent estimation of d is possible for the stationary case in the presence of smooth trends.

The method of detrending by integer-differencing Y_t avoids some of the problems of trend estimation. As long as the trend-free differenced series has memory parameter in the interval $(-.5, .5)$, this parameter can be estimated using the consistent procedures of Robinson (1995a,b). There is no guarantee that this happens in practice, though. After differencing, q times the resulting series will be noninvertible when $q > s$ or will be nonstationary when $q < s$.

Instead of estimating the deterministic trend T_t or differencing the series, a general approach to estimate d_Y is to apply a tapering procedure to the raw data. From (4), (5), and (6), the tapered Fourier transform of the observed sequence Y_t is

$$w_Y^T(\lambda_j) = \frac{1}{\sqrt{2\pi \sum h_t^2}} \sum_{t=1}^n h_t \left(X_o + \sum_{r=1}^{s-1} U_o^{(r)} p_U^{(r)}(t) + \mu p_\mu^{(s)}(t) + \sum_{r=0}^q \alpha_r t^r \right) \exp(i\lambda_j t) \quad (10)$$

$$+ \frac{1}{\sqrt{2\pi \sum h_t^2}} \sum_{t=1}^n h_t \sum_{j_1=1}^t \sum_{j_2=1}^{j_1} \cdots \sum_{j_s=1}^{j_{s-1}} U_{j_s}^{(*)} \exp(i\lambda_j t). \quad (11)$$

The partial-sums term (11) reflects the accumulation of information about the memory parameter d_Y in the observed (nonstationary) time series Y_t , starting from $t = 1$. The right side of (10) is a nuisance component of the DFT, which comprises the information contained in $\{X_t\}_1^n$ from the past, X_o , and $U_o^{(r)}$, together with the information in $\{Y_t\}_1^n$ determined by the mean μ and the trend T_t .

Because the information about d_Y is exclusively contained in (11), to make inferences about d_Y the tapering approach proposes the use of sequences h_t so that (10) equals 0 for certain frequencies λ_j . The set of $\max(s+1, q+1)$ conditions

$$\sum_{t=1}^n h_t t^r \exp(i\lambda_j t) = 0 \quad \text{for } r = 0, \dots, \max(s, q) \quad (12)$$

removes the nuisance term (10) in the DFT of the observed sequence. Notice that for covariance stationary processes ($s, q = 0$) with unknown mean, $\alpha_0 + \mu$, (12) reduces to

$$\sum_{t=1}^n h_t \exp(i\lambda_j t) = 0,$$

which is satisfied by the usual DFT for any frequency $\lambda_j \neq 0 \pmod n$ and for the cosine bell taper at frequencies $\lambda_j \neq -1, 0, 1 \pmod n$. Therefore, the unknown mean of a covariance stationary process, $\alpha_0 + \mu$, appears in the DFT only at zero frequency. This is a well-known fact by researchers in the frequency domain in which estimators and test statistics are constructed avoiding the zero frequency for automatic mean correction. Equation (12) generalizes this property to higher-order trends. Therefore, the identity taper (or raw DFT) and the cosine bell taper should not be applied to series whose s 's are larger than 1 or possess any trend.

Condition (12) is satisfied by taper sequences of order $p > \max(q, s)$ at frequencies $\lambda_j, j = p, 2p, \dots, n - p$ (see Velasco 1999a). Therefore, frequency-domain inference about the parameter d_Y can be performed by choosing a taper of high enough order $p > \max(q, s)$ and defining

the relevant statistics over the set of frequencies $\lambda_j, j = p, 2p, \dots, n-p$. Hence, a conservative approach to estimate d_Y is to set a high value for the order of the taper. This strategy, however, presents a drawback because the higher the order of the taper, the higher the loss of efficiency in the estimation of d_Y (see Sec. 3). This inefficiency can be attenuated through initial differencing if it is known that $d_Y > .5$ or by recalling some tapering robustness to noninvertibility in other cases (see Velasco 1999a, theorem 9).

When $d_Y < .5$ and $q = 0$, there is no need for tapering, and hence the consistent estimators of d_Y proposed by Robinson (1994a, 1995a,b) can be employed. For this case, however, Velasco (1999a,b) showed that Robinson's estimators based on tapered data are still consistent and, furthermore, can be less biased in finite samples. The tapering procedure can also be very effective when the deterministic trends T_t are not polynomials but they are sufficiently smooth in t to be approximated by low-order polynomials. In these cases, it can be expected that the use of tapers with appropriate high enough orders will roughly remove the trends without explicit specification and estimation (see Robinson 1986).

3. A MULTIVARIATE TWO-STEP SEMIPARAMETRIC ESTIMATOR

3.1 A Nonstationary Long-Memory Vector Process

Consider a real-valued N -dimensional vector process $X_t, t = 0, \pm 1, \dots$, that has memory parameter vector $d^0 = (d_1^0, d_2^0, \dots, d_N^0)$, where $d_a^0 > -\frac{1}{2}$ for all a . Denote by $s_a^0 = [d_a^0 + \frac{1}{2}]$ the smallest number of integer differences necessary to make covariance stationary the a th element of X_t , and define the covariance stationary N -dimensional vector process $U_t = \text{diag}\{\Delta^{s_a^0}\}X_t$ with mean μ and autocovariance matrix at lag $j, \Gamma_j = E[(U_t - \mu)(U_{t+j} - \mu)']$. We assume that the spectral density matrix of $U_t, f_U(\lambda) = f(\lambda)$, exists and is implicitly defined by

$$\Gamma_j = \int_{-\pi}^{\pi} f(\lambda) e^{ij\lambda} d\lambda.$$

Denote by $f_{ab}(\lambda)$ the (a, b) th element of $f(\lambda)$. A natural extension of (2) to the vector case (see Lobato and Robinson 1998; Lobato 1999) is

$$f(\lambda) \sim \Lambda^0 R^0 \Lambda^0 \quad \text{as } \lambda \rightarrow 0^+, \quad (13)$$

where R^0 is a symmetric positive definite real matrix and

$$\Lambda^0 = \text{diag}\{\lambda^{s_a^0 - d_a^0}\}, \quad -\frac{1}{2} \leq s_a^0 - d_a^0 < \frac{1}{2}.$$

3.2 A Multivariate Two-Step Semiparametric Estimator

A consistent estimator of the long-memory parameter of a covariance stationary process can be motivated by examining Equation (1). Because (1) concerns only the behavior of the SDF in a neighborhood of 0 and does not constrain the short-term behavior in any way, the consistent approach proposes to estimate d using periodograms evaluated at frequencies on a shrinking neighborhood of the origin. This approach is implemented by defining a bandwidth parameter

m asymptotically negligible with respect to the sample size, n , so that only periodograms evaluated up to frequency $\lambda_m = 2\pi m/n$ are used. The idea was introduced by Geweke and Porter-Hudak (1983), formalized by Robinson (1994a), and further developed by Robinson (1995a,b).

In this article we propose a two-step estimator (TSE) based on an extension of the objective function considered by Lobato (1999) to allow for tapering as shown by Velasco (1999b) (see also Künsch 1987; Robinson 1995b). This function is the frequency-domain version of the Whittle log-likelihood function. Estimators of the long-memory parameters based on this function have several advantages over rival semiparametric estimators. For instance, they are more efficient, and there is no need to introduce additional user-chosen numbers to compute them. Furthermore, compared with a multivariate quasi maximum likelihood estimator (QMLE), a TSE has the same asymptotic efficiency and the advantage of being easier to compute because there is no need of carrying out a multidimensional maximization problem. The objective function considered is the following:

$$\begin{aligned} \mathcal{L}_p(R, d) &= \frac{p}{m} \sum_{j(p)}^m \{ \log |\Lambda_j R \Lambda_j| + \text{tr} [(\Lambda_j R \Lambda_j)^{-1} I^T(\lambda_j)] \}, \quad (14) \end{aligned}$$

where the symbol $\sum_{j(p)}^m$ means that the sum is performed over $j = p, 2p, \dots, m$ (we assume m/p to be integer), $\Lambda_j = \text{diag}\{\lambda_j^{-d_a}\}$, and $I^T(\lambda_j)$ is the tapered periodogram matrix $I^T(\lambda_j) = w^T(\lambda_j) w^T(\lambda_j)^*$, where $*$ represents transposition combined with conjugation and $w^T(\lambda_j)$ is given by (9).

Equation (14) is a modified version of the Whittle approximation of the Gaussian log-likelihood function considered by Lobato (1999). Notice the distinction between the true values d^0, R^0 and any admissible value d, R . Expression (14) can be simplified by concentrating out R to end up with the following concentrating objective function (see Lobato 1999):

$$\Upsilon_p(d) = \frac{-2p}{m} \sum_{a=1}^N d_a \sum_{j(p)}^m \log(\lambda_j) + \log |\hat{R}_p(d)|,$$

where

$$\hat{R}_p(d) = \frac{p}{m} \sum_{j(p)}^m \{ \Lambda_j^{-1} \text{Re}\{I^T(\lambda_j)\} \Lambda_j^{-1} \}. \quad (15)$$

The estimation procedure we propose is a TSE based on this objective function. The first step is to compute the univariate QMLE of Velasco (1999b) for every series (denote that vector by $\hat{d}^{(1)}$), and the second step is to compute the following expression:

$$\hat{d}^{(2)} = \hat{d}^{(1)} - \left(\frac{\partial^2 \Upsilon_p(d)}{\partial d \partial d'} \Big|_{\hat{d}^{(1)}} \right)^{-1} \left(\frac{\partial \Upsilon_p(d)}{\partial d} \Big|_{\hat{d}^{(1)}} \right). \quad (16)$$

Then the estimator of R^0 is, using (15),

$$\hat{R} = \hat{R}_p(\hat{d}^{(2)}). \quad (17)$$

3.3 A CLT for the TSE

The following assumptions are needed for the CLT.

Assumption 1. For $\beta \in (0, 2]$,

$$|f_{ab}(\lambda) - r_{ab}\lambda^{-d_a^0 - d_b^0}| = O(\lambda^{-d_a^0 - d_b^0 + \beta}) \quad \text{as } \lambda \rightarrow 0^+,$$

where r_{ab} is the (a, b) th element of R^0 .

Assumption 2. $U_t = \mu + \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}$ with $\sum_{j=0}^{\infty} \|A_j\| < \infty$, where $\|\cdot\|$ denotes the supremum norm and ε_t satisfies $E(\varepsilon_t | \mathcal{T}_{t-1}) = 0$, $E(\varepsilon_t \varepsilon_t' | \mathcal{T}_{t-1}) = I_N$, and $E(\varepsilon_a(t) \varepsilon_b(t) \varepsilon_c(t) | \mathcal{T}_{t-1}) = \mu_{abc}$, with $|\mu_{abc}| < \infty$ for $a, b, c = 1, \dots, N$; $E(\varepsilon_a(t) \varepsilon_b(t) \varepsilon_c(t) \varepsilon_d(t) | \mathcal{T}_{t-1}) = \mu_{abcd}$, where $|\mu_{abcd}| < \infty$ for $a, b, c, d = 1, \dots, N$, where \mathcal{T}_{t-1} is the σ field of events generated by $\{\varepsilon_s, s \leq t-1\}$.

Assumption 3. As $\lambda \rightarrow 0^+$,

$$\frac{dA_a(\lambda)}{d\lambda} = O(\lambda^{-1} \|A_a(\lambda)\|)$$

for $a = 1, \dots, N$, where $A_a(\lambda)$ is the a th row of $A(\lambda) = \sum_{j=0}^{\infty} A_j \exp(ij\lambda)$.

Assumption 4. As $n \rightarrow \infty$,

$$\frac{1}{m} + \frac{m^{1+2\beta} (\log m)^2}{n^{2\beta}} \rightarrow 0,$$

in addition, when $p > 1$, $m^{-1} \log n \log m \rightarrow 0$.

Assumption 5. The order p of the taper satisfies the following: When $p = 1$, $\max_a d_a^0 < \frac{3}{4}$, and when $p > 1$, $\max_a s_a^0 < p$.

Assumption 1 strengthens Specification (13) by imposing a rate of convergence of $f_{ab}(\lambda)$ to $r_{ab}\lambda^{-d_a^0 - d_b^0}$. This smoothness condition is typically imposed in spectral analysis (see Lobato 1999). One of the major technical contributions of this article is to prove the CLT when the tapering procedure is employed with $p > 1$ under Assumption 1. Previous research has both strengthened Assumption 1 (cf. Velasco 1999b, Assumption 8) and assumed a lower bound for β , such as $\beta > 1$. This extension was recommended by a referee who suggested the case of a stationary long-memory series $X_t(d_X < \frac{1}{2})$ observed with an added short-memory noise component for which $\beta \leq 2d_X < 1$. This case has empirical relevance because it happens in some long-memory stochastic volatility models. In addition, in certain bivariate models of volume and volatility, the short-memory component could affect the long-memory components of both volume or volatility (leverage effect) implying that $\beta < 1$.

Assumption 2 establishes that the process is linear with finite fourth moment. The assumption of linear fourth-order stationary processes has also been employed in the parametric literature (see Giraitis and Surgailis 1990; Hosoya 1997). Notice that the restriction of constant conditional innovations variances could be relaxed by assuming boundedness of the eighth moment as shown by Robinson and Henry (1999). Assumption 3 is a regularity condition similar to the ones imposed in the parametric case.

Assumption 4 [which is assumption A4' of Robinson (1995b) for the $p = 1$ case] imposes an upper bound in the rate of increase of m with n , necessary to control the bias from high frequencies. Notice that this upper bound is especially restrictive when β is small, suggesting that when

the long-memory series is observed with an added noise or when the leverage effect is important, the chosen m should be smaller. It also imposes a mild lower rate for m when tapering is applied.

Assumption 5 restricts the order of the taper to be higher than the maximum order of integration of the series that compose the vector process. This condition implies that the level of tapering is high enough for the tapered periodogram to be unbiased for the pseudo SDF and for the construction of consistent estimators of d (see Velasco 1999a,b).

When the components of the stochastic vector process contain deterministic polynomial trends, $T_t^{(a)}$, of order $q^{(a)}$ (≥ 0), the order of the taper has to satisfy the following assumption.

Assumption 6. The order of the taper p satisfies $p > \max_a q^{(a)}$.

Theorem 1. Under Assumptions 1, 2, 3, 4, and 5 (and 6 when deterministic trends are present)

$$\sqrt{m}(\hat{d}^{(2)} - d^0) \rightarrow_d N(0, p\Phi_p E^{-1}), \quad (18)$$

where $\hat{d}^{(2)}$ is given by (16), $E = 2(I_N + (R^0 \circ R^{0^{-1}}))$,

$$\Phi_p = \lim_{n \rightarrow \infty} \left(\sum_{t=1}^n h_t^2 \right)^{-2} \sum_{k=0, p, 2p, \dots}^{n-p} \left(\sum_{t=1}^n h_t^2 \cos t\lambda_k \right)^2,$$

and \circ denotes the Hadamard product of two matrices.

The proof is in the Appendix. Notice that, with respect to the CLT provided by Lobato (1999), the covariance matrix is increased by a factor of $p\Phi_p$. The factor Φ_p is due to the lack of orthogonality of the taper weights that implies correlation of adjacent tapered periodograms. The effect of Φ_p is very moderate, though. Notice that Φ_p takes the values of 1.05000, 1.00354, and 1.00086 for the Kolmogorov–Zhurbenko kernels with $p = 2, 3, 4$, respectively, implying increments of the variance of 5%, .35%, and .09% for each of the data tapers. The factor p is because only frequencies $\lambda_j, j = p, 2p, \dots, n-p$, are employed. Notice that only such frequencies can be employed with possibly non-stationary data; hence the factor p relates to the fact that there is no need to specify or to estimate the deterministic trends when tapering is applied.

Although the tapering procedure implies an increase in the limiting variance of the TSE, the finite-sample performance of estimators based on tapered data can be superior to those based on raw data. Hauser (1999), Velasco (1999a,b), and Velasco and Robinson (in press) found that the tapered versions of semiparametric and parametric estimators of d are more reliable than the procedures based on raw data.

To be able to use (18) for statistical inference, we need to obtain a consistent estimator for E . Considering the definition of E , the natural one is $\hat{E} = 2(I_N + (\hat{R} \circ \hat{R}^{-1}))$, where \hat{R} is given by (17). To prove the consistency of \hat{R} for R^0 , it is necessary to strengthen Assumption 4 to the following.

Assumption 4'. As $n \rightarrow \infty, m \sim \Psi n^\delta$, with Ψ a finite positive constant and $0 < \delta < 1$.

Table 1. Bias and Root Mean Squared Error of the TSE

		m								
		60			150			140		
		p								
		1	2	3	1	2	3	1	2	3
Bias	DGP1, d1	-.004	-.014	-.017	-.005	-.011	-.013	-.012	-.017	-.015
	d2	.003	-.004	-.013	.001	-.004	-.009	-.008	-.012	-.014
	DGP2, d1	-.004	-.014	-.017	-.005	-.011	-.012	-.011	-.017	-.015
	d2	-.010	-.002	-.012	.006	-.005	-.010	-.006	-.017	-.018
	DGP3, d1	-.004	-.015	-.018	-.005	-.011	-.012	-.012	-.017	-.015
	d2	.019	.003	-.008	.014	-.004	-.010	-.001	-.020	-.022
RMSE	DGP4, d1	-.004	-.014	-.017	-.005	-.011	-.013	-.012	-.017	-.015
	d2	.317	-.004	-.013	.221	-.004	-.009	.172	-.012	-.014
	DGP1, d1	.076	.121	.156	.045	.067	.089	.037	.053	.065
	d2	.075	.118	.155	.044	.066	.088	.035	.052	.064
	DGP2, d1	.076	.124	.158	.044	.068	.086	.037	.054	.065
	d2	.079	.118	.154	.047	.067	.085	.037	.054	.065
	DGP3, d1	.076	.126	.157	.044	.068	.087	.037	.054	.064
	d2	.081	.119	.154	.052	.067	.086	.040	.054	.066
	DGP4, d1	.076	.121	.156	.045	.067	.089	.037	.053	.065
	d2	.321	.118	.155	.223	.066	.088	.174	.052	.064

NOTE: The first three DGPs are independent ARFIMA's with (d1, d2) = (4, 4), (4, 6), and (4, 8). DGP4 is DGP1 with a linear trend added to the second series. Sample size n = 1,000.

Then, the following lemma can be established using theorem 5 of Robinson (1994a) in a way similar to that of Lobato (1999).

Lemma. Under Assumptions 1, 2, 3, 4', and 5 (and 6 when deterministic trends are present), $\hat{R} \rightarrow_p R^0$.

Consider as null hypothesis a linear set of r (r ≤ N) independent restrictions on $d^0, Hd^0 = z$, where H is r × N and z is r × 1. Then, the test statistic

$$T = \frac{m}{p} \Phi_p^{-1} (H \hat{d}^{(2)} - z)' (H \hat{E}^{-1} H')^{-1} (H \hat{d}^{(2)} - z) \quad (19)$$

is asymptotically distributed as a χ_r^2 distribution under the null hypothesis. An interesting case is testing for a common long-memory parameter for the vector process. In this case z is a vector of N - 1 zeros and $H = (I_{N-1} : 0) - (0 : I_{N-1})$, with dimension (N - 1 × N), where I_{N-1} is the identity matrix of order N - 1, and 0 is an N × 1 vector of zeros. In Section 4, we apply (19) to test for a common long-memory parameter for stock-market volatility and volume.

To establish a CLT for \hat{R} , Assumption 4' needs to be strengthened.

Assumption 4''. As $n \rightarrow \infty, m \sim \Psi n^\delta$, with Ψ a finite positive constant and $0 < \delta < 2\beta/(1 + 2\beta)$.

Theorem 2. Under Assumptions 1, 2, 3, 4'', and 5 (and 6 when deterministic trends are present),

$$\sqrt{m}(\text{vec}(\hat{R}(\hat{d})) - \text{vec}(R)) \rightarrow_d N(0, p\Phi_p F), \quad (20)$$

where F is a square matrix of dimension $N' = N(N + 1)/2$ and typical element $F_{ab,rs} = \frac{1}{2}(R_{ar}R_{bs} + R_{as}R_{br})$.

The proof is in the Appendix, Section A.4. The components of $\text{vec}(\hat{R}(\hat{d}))$ can be considered as standard prewhitened nonparametric spectral estimates at zero frequency. Hence, they can be used to estimate the squared coherency between series a and b at zero frequency, $H_{ab}(0)^2$, by $\widehat{H}_{ab}(0)^2 = \widehat{R}_{ab}^2 / (\widehat{R}_{aa}\widehat{R}_{bb})$. Then, using (20) and the delta method, it can be proved that, for $0 < H_{ab}(0)^2 < 1$ (see Brillinger 1975, p. 257),

$$\begin{aligned} \sqrt{m}(\widehat{H}_{ab}(0)^2 - H_{ab}(0)^2) \\ \rightarrow_d N(0, 2p\Phi_p H_{ab}(0)^2(1 - H_{ab}(0)^2)^2), \quad (21) \end{aligned}$$

for $a, b = 1, \dots, N$.

We finish this section with a small finite-sample Monte Carlo exercise reported in Tables 1 and 2. The finite-sample bias and root mean squared error (RMSE) of the TSE re-

Table 2. Proportion of Rejections of the Wald Test of Common Long-Memory Parameter Based on 5% Asymptotic Critical Value

		m								
		60			150			240		
		p								
		1	2	3	1	2	3	1	2	3
DGP1		.099	.130	.153	.073	.095	.105	.070	.077	.089
DGP2		.631	.414	.331	.937	.692	.528	.991	.850	.680
DGP3		.987	.843	.693	1.000	.995	.959	1.000	1.000	.996
DGP4		.943	.130	.153	.981	.095	.105	.985	.077	.089

NOTE: DGP's and sample size are the same as in Table 1.

ported in Table 1 are for four different data-generating processes (DGP's). The first three DGP's are Gaussian independent autoregressive fractionally integrated moving averages with $(d_1, d_2) = (.4, .4), (.4, .6),$ and $(.4, .8)$. In the fourth DGP, we have added a linear trend (that takes the values $.005t, t = 1, \dots, n$) to the second series for the first DGP. We have considered one sample size ($n = 1,000$), three values for m ($m = 60, 150, 240$), and three values for p ($p = 1, 2, 3$). The number of replications has been 5,000. S-PLUS and FORTRAN 90 codes with all the procedures employed in the article are available from the authors. Table 1 indicates that biases are small and, typically, negative. As could be expected, however, for the fourth DGP the bias for the second estimate is very large when $p = 1$ is employed. Notice that Assumption 5 does not hold for the third DGP and $p = 1$, explaining the slightly larger bias of estimates for d_2 . RMSE decreases with m and increases with p , as expected. In additional experiments that are not reported, we have observed that the finite-sample behavior of the TSE is very similar to the QMLE. QMLE is computationally more demanding, which is another reason to prefer TSE.

In Table 2, we examine the performance of the T test given in (19) for $H = (1, -1)$ and $z = 0$ (the bivariate Wald test for a common d) for the four DGP's considered previously. This table illustrates the effect of choosing p incorrectly. Notice that, for the fourth DGP, Assumption 6 requires that p is chosen greater than 1. Table 2 shows that, for this DGP when no tapering is carried out ($p = 1$), the true null hypothesis is rejected more than 90% of the times. In addition, it can be observed that, for the chosen values of n and m , the rejection proportions (RP's) under the null hypothesis increase with p and decrease with m in agreement with Theorem 1 (they are about 8% instead of 5%), whereas the RP's decrease with p and increase with m under the alternative.

4. LONG MEMORY IN TRADING VOLUME

In this section we analyze the long-memory properties for the daily trading volume and return volatility processes of the 30 stocks that compose the DJIN index for the period July 1962 to December 1994.

Because stock-market trading volume is nonstationary, several detrending procedures for the volume data have been considered in the empirical finance literature. For instance, Gallant et al. (1992) fitted a quadratic polynomial trend, Andersen (1996) estimated nonparametrically the trend of the logarithm of the volume series using both equally and unequally weighted moving averages, and Bollerslev and Jubinski (1999) fitted a linear trend. There is a lack of rigorous statistical theory, however, on the effects of detrending for the inference on the long-memory parameters of nonstationary long-memory processes. Hence, the determination of a detrending mechanism that would allow for inference on the long-memory parameter of stock-market volume is still an unsolved problem.

The most prominent economic theory regarding the behavior of stock-market trading volume and its relation to return volatility appears to be the information-counting

model or mixture-of-distribution hypothesis (MDH) model. There are several MDH models based on different theoretical backgrounds (see Clark 1973; Epps and Epps 1976; Tauchen and Pitts 1983; Lamoureux and Lastrapes 1994; Andersen 1996; Bollerslev and Jubinski 1999), and essentially the MDH posits a joint dependence of returns volatility and volume on an underlying information-flow process. Hence, according to the MDH, both the return volatility and the trading-volume processes should share the same stochastic properties.

Because it has been repeatedly shown that a main feature of return volatility (for exchange-rate returns and for stock-market returns) is the presence of long memory (e.g., see Ding, Granger, and Engle 1993; Bollerslev and Mikkelsen 1996; Baillie, Bollerslev, and Mikkelsen 1996; Lobato and Savin 1998), it is of interest to test if stock-market volume exhibits long memory and, if this is the case, to test also if return volatility and volume share the same long-memory parameter d . These two implications were derived by Bollerslev and Jubinski (1999)—see their equations (5) and (6)—but their testing procedure has two drawbacks. First, when applied to data linearly (or nonlinearly) detrended, the estimators of the long-memory parameters have unknown properties. Second, it is unknown a priori that the long-memory parameter of volume lies in the stationary region ($d < .5$), and in fact they reported some cases for which $d > .5$. In this section, we use a tapering procedure to overcome both difficulties.

Table 3 lists the 30 stocks in the DJIA index with their ticks and the periods covered by the data. The period covers July 1962 to December 1994 (hence, the number of observations is 8,180) for all stocks but 3. These were the data employed by Lobato and Savin (1998) in which strong evidence of long memory was found in squared and absolute returns (as measures of the return volatility process) using a semiparametric Lagrange multiplier (LM) test for weak dependence.

To illustrate a series, Figure 1 presents trading volume (corrected for stock splits) for IBM. Because trending heteroscedasticity is clearly present—see also figure 2A of Andersen (1996)—we analyze the logarithm of stock trading volume. This way of stabilizing the variance is a standard practice in the empirical finance literature. Figure 2 presents the daily log-trading volume for IBM (compare with Andersen's fig. 2B). The IBM log-trading volume exhibits a strongly nonlinear trend that might be represented by a quadratic or a cubic polynomial. To study the long-memory properties of log-trading volume, we could set the order of the taper at least to $p = 3$ or $p = 4$ and apply the tools considered in the previous sections. This approach, as we have seen, is a robust way of estimating the long-memory properties of processes whose nonstationary trends can be approximated by quadratic or cubic polynomials.

In Table 4 we present the results of a modified version of the univariate version of the LM test for weak dependence of Lobato and Robinson (1998) applied to log-trading volume corrected for stock splits for the 30 stocks in the DJIA. The modification consists of using the tapered periodogram and defining the statistic over the frequencies

Table 3. List of the 30 Dow Jones Industrial Average Companies With Beginning and Ending Dates

Stock	Name	Beginning date	Ending date
ATT	AT&T Corp.	07-02-1962	12-30-1994
ALD	Allied Signal Inc.	07-02-1962	12-30-1994
AA	Aluminum Company of America	07-02-1962	12-30-1994
AXP	American Express Co.	05-18-1977	12-30-1994
BS	Bethlehem Steel Corp.	07-02-1962	12-30-1994
BA	Boeing Co.	07-02-1962	12-30-1994
CAT	Caterpillar Inc.	07-02-1962	12-30-1994
CHV	Chevron Corp.	07-02-1962	12-30-1994
KO	Coca Cola Corp.	07-02-1962	12-30-1994
DIS	Walt Disney Co.	07-02-1962	12-30-1994
DD	DuPont E.I. De Nemours & Co.	07-02-1962	12-30-1994
EK	Eastman Kodak Co.	07-02-1962	12-30-1994
XON	Exxon Corp.	07-02-1962	12-30-1994
GE	General Electric Corp.	07-02-1962	12-30-1994
GM	General Motors Corp.	07-02-1962	12-30-1994
GT	Goodyear Tire & Rubber Co.	07-02-1962	12-30-1994
IBM	IBM	07-02-1962	12-30-1994
IP	International Paper Co.	07-02-1962	12-30-1994
MCD	McDonalds Corp.	07-05-1966	12-30-1994
MRK	Merck & Co., Inc.	07-02-1962	12-30-1994
3M	Minnesota Mining & Mfg. Co.	07-02-1962	12-30-1994
JPM	Morgan, J. P. & Co. Inc.	04-01-1969	12-30-1994
MO	Philip Morris Co. Inc.	07-02-1962	12-30-1994
PG	Procter and Gamble Co.	07-02-1962	12-30-1994
S	Sears Roebuck & Co.	07-02-1962	12-30-1994
TX	Texaco Inc.	07-02-1962	12-30-1994
UK	Union Carbide Corp.	07-02-1962	12-30-1994
UTX	United Technology Corp.	07-02-1962	12-30-1994
WX	Westinghouse Electric Corp.	07-02-1962	12-30-1994
Z	Woolworth Corp.	07-02-1962	12-30-1994

$j = p, 2p, \dots, m$. Hence, the test statistic employed is

$$LM = (m/p) \left(\frac{\sum_{j(p)}^m \nu_j I^T(\lambda_j)}{\sum_{j(p)}^m I^T(\lambda_j)} \right)^2,$$

where

$$\nu_j = \log j - (p/m) \sum_{j(p)}^m \log j. \quad (22)$$

The null hypothesis that the log-trading volume process is weak dependent ($H_0 : d^0 = 0$) is rejected in favor of the alternative that the log-trading volume process exhibits long

memory ($H_a : d^0 \neq 0$) when LM is significantly large relative to the χ_1^2 distribution. Notice that, to facilitate their examination, Table 4 reports the p values rather than the values of the test statistics. We present the results for the Kolmogorov-Zhurbenko taper with $p = 3$ and a grid of values for m from 60 to 300. Results for the $p = 4$ case are qualitatively similar. From (18) it is obvious that inference is more accurate as m is larger. For big values of m , however, the statistics employed would reflect the medium- and short-term behavior of the process that would bias our inference. This trade-off between bias and variance was addressed, for instance, by Robinson (1994b) and Henry and Robinson (1996) for the univariate case. Following the pro-

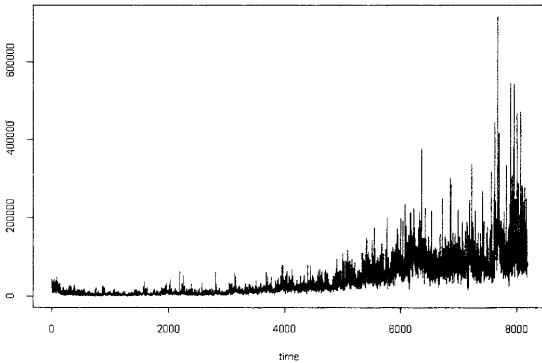


Figure 1. Trading Volume for IBM Over the Period July 1962—December 1994.

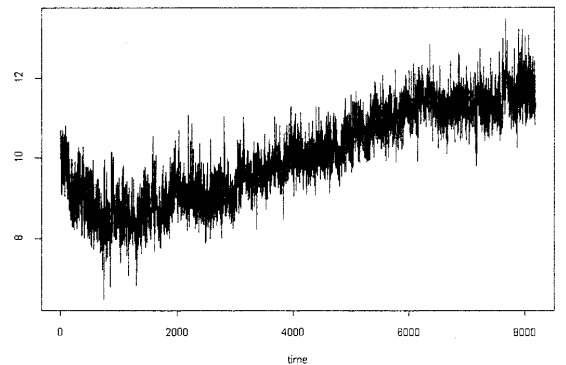


Figure 2. Log-Trading Volume for IBM Over the Period July 1962—December 1994.

Table 4. *P* Values of the Univariate LM Test for Long Memory for Volume for the Kolmogorov–Zhurbenko Taper of Order $p = 3$

Stock	m										
	60	84	108	132	156	180	204	228	252	276	300
ATT	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
ALD	.070	.028	.010	.001	.000	.000	.000	.000	.000	.000	.000
AA	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
AXP	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
BS	.095	.019	.009	.001	.000	.000	.000	.000	.000	.000	.000
BA	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
CAT	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
CHV	.019	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
KO	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
DIS	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
DD	.025	.008	.001	.000	.000	.000	.000	.000	.000	.000	.000
EK	.140	.005	.003	.000	.000	.000	.000	.000	.000	.000	.000
XON	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
GE	.021	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
GM	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
GT	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
IBM	.269	.104	.012	.007	.001	.000	.000	.000	.000	.000	.000
IP	.001	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
MCD	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
MRK	.026	.002	.000	.000	.000	.000	.000	.000	.000	.000	.000
3M	.213	.033	.003	.000	.000	.000	.000	.000	.000	.000	.000
JPM	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
MO	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
PG	.057	.015	.001	.000	.000	.000	.000	.000	.000	.000	.000
S	.493	.074	.005	.001	.000	.000	.000	.000	.000	.000	.000
TX	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
UK	.013	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
UTX	.013	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
WX	.035	.005	.000	.000	.000	.000	.000	.000	.000	.000	.000
Z	.016	.005	.000	.000	.000	.000	.000	.000	.000	.000	.000

NOTE: The null hypothesis is $d_{\text{volume}} = 0$.

cedure of Lobato and Savin (1998), we prefer to report the results for a grid of m rather than the results for just some optimal m . Furthermore, notice that even when $m = 300$ the shortest cycle included in the estimation is 27.3 days, which is marginally higher than the monthly trading day, and hence the potential bias caused by the monthly seasonal peak is avoided in all the calculations. Inspection of Table 4 indicates that log-trading volume exhibits long-term dependence. This agrees with Bollerslev and Jubinski's (1999) results for individual stocks in the S&P 100 index.

The null hypothesis of weak dependence can also be tested using Wald-type tests based on (19). We prefer to report the results of the LM test because, as pointed out by Lobato and Savin (1998, p. 283), the LM test tends to reject less often than the Wald test. Results (that are not reported) of univariate and bivariate Wald-type tests based on (19) confirm the evidence from the LM test.

Because both stock-market volume and return volatility exhibit long memory, we proceed to test whether both processes share the same long-memory parameter. The test statistic employed, based on (19), is

$$T = \frac{m}{p} \Phi_p^{-1} (H \hat{d}^{(2)})' (H \hat{E}^{-1} H')^{-1} (H \hat{d}^{(2)}), \quad (23)$$

where $H = (1, -1)$. As mentioned in Section 3, the test statistic T asymptotically follows a χ_1^2 distribution under the null even if the analyzed series possess poly-

nomial or stochastic trends of order less than p . We present the results for the same values for m and p in Table 5.

Table 5 shows that for most of the stocks and most values of m the null hypothesis of a common long-memory parameter cannot be rejected, the exceptions being just six stocks (ATT, AXP, BA, IP, JPM, and Z). Because for most of the values of m the null hypothesis of equal long-memory parameters cannot be rejected, Table 6 presents the estimated values of the common d . From this table, it is clear that most of the estimates range between .3 and .5, although there are some estimates above .5. For completeness, we report the TSE for volume (upper row) and volatility (lower row) for those cases in which the hypothesis of a common long-memory parameter is rejected. The asymptotic standard errors associated with these estimates vary slightly across series. The typical standard error for $m = 60$ is .090, for $m = 180$ it is .050, and for $m = 300$ it is about .038. Notice that, due to the particular form of the matrix E , both elements in the main diagonal are the same, and hence the standard errors for the TSE for volume and volatility are the same.

Notice that the null hypothesis of a common long-memory parameter could have been rejected for those six stocks because the trend of log-trading volume was not well captured by a quadratic polynomial (as occurs for the $p = 1$ case in which the null of common d is always rejected). In case the trend in log-volume trading

Table 5. *P* Values of the Wald Test for Common Long-Memory Parameter for Volatility (measured by absolute returns) and Volume for the Kolmogorov–Zhurbenko Taper of Order $p = 3$

Stock	<i>m</i>										
	60	84	108	132	156	180	204	228	252	276	300
ATT	.276	.469	.055	.000	.000	.000	.000	.000	.000	.000	.000
ALD	.671	.573	.503	.647	.842	.169	.165	.072	.032	.050	.040
AA	.692	.418	.523	.653	.866	1.00	.705	.959	.977	.664	.775
AXP	.031	.553	.334	.001	.000	.000	.000	.000	.000	.000	.000
BS	.470	.590	.719	.771	.770	.458	.129	.049	.011	.002	.000
BA	.276	.012	.003	.097	.030	.097	.049	.007	.003	.000	.000
CAT	.664	.052	.369	.275	.706	.464	.088	.085	.070	.155	.052
CHV	.403	.773	.920	.446	.672	.496	.332	.237	.147	.234	.167
KO	.975	.116	.210	.516	.969	.878	.725	.540	.304	.423	.666
DIS	.933	.480	.975	.920	.195	.213	.554	.329	.218	.090	.045
DD	.672	.079	.298	.636	.690	.855	.661	.960	.734	.778	.225
EK	.795	.911	.120	.724	.958	.868	.831	.982	.626	.904	.677
XON	.793	.280	.303	.296	.214	.138	.050	.064	.010	.005	.003
GE	.282	.006	.034	.070	.031	.427	.024	.053	.112	.223	.343
GM	.644	.166	.093	.631	.991	.625	.526	.476	.176	.307	.245
GT	.684	.087	.615	.326	.361	.215	.128	.071	.142	.045	.029
IBM	.024	.003	.038	.074	.150	.348	.516	.580	.736	.371	.696
IP	.033	.518	.319	.092	.037	.039	.013	.007	.015	.001	.000
MCD	.019	.580	.385	.393	.549	.142	.129	.058	.011	.012	.006
MRK	.410	.904	.304	.438	.376	.537	.852	.692	.402	.490	.574
3M	.201	.827	.865	.718	.216	.336	.257	.052	.005	.003	.013
JPM	.000	.002	.025	.003	.000	.000	.000	.000	.000	.000	.000
MO	.996	.959	.688	.929	.727	.996	.985	.817	.804	.399	.208
PG	.086	.262	.248	.377	.306	.676	.965	.895	.818	.886	.634
S	.362	.444	.732	.415	.658	.458	.255	.177	.132	.187	.095
TX	.078	.247	.317	.626	.679	.644	.453	.410	.441	.787	.818
UK	.935	.948	.358	.214	.208	.018	.080	.168	.209	.045	.042
UTX	.560	.913	.932	.998	.523	.301	.203	.121	.080	.156	.109
WX	.190	.223	.220	.293	.514	.756	.524	.233	.195	.188	.177
Z	.117	.050	.027	.004	.006	.000	.000	.000	.000	.000	.000

NOTE: The null hypothesis is that $\alpha_{volatility} = \alpha_{volume}$.

could be approximated by a polynomial of order third, fourth, or fifth, we should employ a taper of order $p = 4$, $p = 5$, or $p = 6$, respectively. Using these higher-order tapers, it appears that in all cases except AXP the trend in volume is properly captured by a high-order polynomial and the common long-memory parameters are in the interval .3–.5. These results agree with Bollerslev and Jubinski's (1999) results for the stocks in the S&P 100 index.

Because it appears that both volume and volatility possess the same long-memory parameter, it is of interest to examine if both processes are driven by the same long-memory component—that is, if volume and volatility are fractionally cointegrated. Even though the issue of fractional cointegration has received much attention lately (see Robinson and Marinucci 1998 or Ray and Tsay 1998), relevant asymptotic distribution theory has proved elusive.

One way of checking whether volume and volatility are fractionally cointegrated is to test the necessary condition that the coherency between both series is 1 at zero frequency. In Table 7 we present the estimated squared coherency at zero frequency for the $p = 3$ case in which the squared coherency is estimated by

$$\widehat{H}_{ab}(0)^2 = \frac{\widehat{R}_{ab}^2}{\widehat{R}_{aa}\widehat{R}_{bb}} = \frac{(\sum_{j(p)}^m \lambda_j^{\hat{d}_a + \hat{d}_b} \text{Re} I_{ab}(\lambda_j))^2}{\sum_{j(p)}^m \lambda_j^{2\hat{d}_a} I_{aa}(\lambda_j) \sum_{j(p)}^m \lambda_j^{2\hat{d}_b} I_{bb}(\lambda_j)}$$

Typically the estimated squared coherency is in the range .1–.3, and only in one case is marginally higher than .4. Using (21) we can have an idea about the typical standard error of $\widehat{H}_{ab}(0)^2$. When $\widehat{H}_{ab}(0)^2 \approx \frac{1}{3}$ and $p = 3$, for $m = 60$, $\text{s.e.}(\widehat{H}_{ab}(0)^2) \approx .122$, and for $m = 150$, $\text{s.e.}(\widehat{H}_{ab}(0)^2) \approx .077$. Hence, Table 7 clearly indicates the absence of fractional cointegration between volume and volatility. Results when the squared coherency at zero frequency is estimated by imposing a common \hat{d} are similar. In all cases the estimated squared coherency is below .4.

Finally, we note that, because the semiparametric tools employed in this article are robust to the seasonal behavior of the vector process considered, the empirical results reported in the article are not affected by some data filtering common in the empirical finance literature. For instance, filtering the data by excluding the Christmas period (December 23rd to the first trading day of the following year), as Andersen (1996) proposed, would not affect our results.

5. CONCLUSIONS

In this article new methods for carrying out statistical inference on the long memory parameters of stock-market trading volume and volatility have been examined. We have avoided detrending procedures and analyzed consistent es-

Table 6. Two-Step Estimator (TSE) of the Common Long-Memory Parameter d for Volatility (measured by absolute returns) and Volume (reported when the corresponding p value in Table 5 is above .05) or TSE for Volume (upper row) and Volatility (lower row) (reported when the corresponding p value in Table 5 is below .05) for the Kolmogorov–Zhurbenko Taper of Order $p = 3$

Stock	m										
	60	84	108	132	156	180	204	228	252	276	300
ATT	.655	.654	.635	.414	.403	.382	.388	.333	.348	.314	.321
				.782	.771	.728	.690	.660	.622	.601	.609
ALD	.338	.346	.325	.314	.321	.294	.289	.311	.274	.321	.274
									.411		.401
AA	.454	.358	.368	.336	.363	.375	.380	.378	.361	.338	.298
AXP	.809	.815	.841	.582	.434	.433	.074	.464	.477	.388	.414
				.912	.987	1.00	1.61	.933	.913	1.03	.960
BS	.251	.268	.244	.242	.252	.270	.288	.252	.234	.224	.202
								.394	.402	.419	.448
BA	.419	.368	.338	.418	.308	.442	.352	.334	.298	.263	.243
		.660	.662		.502		.498	.521	.501	.502	.493
CAT	.570	.517	.437	.455	.474	.443	.398	.399	.410	.405	.389
CHV	.403	.773	.920	.446	.672	.496	.332	.237	.147	.234	.167
KO	.493	.441	.409	.396	.429	.405	.372	.352	.354	.348	.353
DIS	.553	.507	.512	.491	.431	.401	.390	.394	.385	.373	.324
											.450
DD	.347	.367	.353	.327	.320	.332	.336	.311	.278	.285	.267
EK	.350	.459	.402	.367	.357	.329	.328	.351	.333	.327	.329
XON	.500	.554	.476	.406	.403	.382	.386	.395	.274	.223	.222
									.458	.414	.420
GE	.527	.598	.574	.430	.530	.425	.511	.420	.418	.414	.412
		.252	.333		.321		.318				
GM	.553	.507	.511	.442	.471	.460	.455	.424	.393	.360	.364
GT	.463	.472	.452	.453	.413	.409	.390	.383	.381	.284	.283
										.417	.428
IBM	.632	.650	.548	.373	.365	.362	.366	.356	.352	.337	.343
	.329	.323	.339								
IP	.471	.393	.374	.295	.224	.214	.224	.231	.250	.225	.205
					.401	.367	.401	.434	.425	.432	.433
MCD	.274	.463	.497	.379	.391	.386	.401	.399	.369	.347	.345
	.606										
MRK	.219	.284	.285	.374	.372	.361	.349	.313	.277	.268	.281
3M	.287	.290	.319	.324	.349	.323	.320	.344	.274	.264	.244
									.458	.452	.400
JPM	.293	.383	.438	.392	.334	.295	.272	.251	.234	.232	.233
	.880	.782	.702	.704	.720	.674	.670	.660	.677	.634	.632
MO	.494	.485	.424	.394	.416	.407	.392	.375	.367	.360	.354
PG	.376	.329	.370	.402	.374	.334	.363	.347	.342	.346	.347
S	.292	.452	.482	.372	.424	.371	.345	.354	.355	.369	.367
TX	.508	.558	.489	.463	.461	.428	.430	.445	.428	.408	.411
UK	.390	.383	.403	.381	.373	.381	.315	.388	.387	.290	.291
							.446			.421	.419
UTX	.462	.451	.421	.397	.362	.368	.369	.353	.367	.353	.348
WX	.347	.328	.374	.335	.325	.296	.314	.333	.296	.302	.291
Z	.360	.259	.173	.134	.142	.100	.103	.095	.112	.111	.104
			.382	.381	.380	.427	.448	.411	.404	.403	.427

timates based on tapered data. Application of these procedures indicates that for the stocks in the DJIA index the logarithm of the trading volume exhibits long memory and, additionally, that it shares the same long-memory parameter with the volatility process for most of the stocks. The estimated coherency between log-volume and absolute returns at zero frequency, however, is very small.

Finally, we note that tapering is a statistical tool that can be very useful in many economic problems. For instance, the comparison of trend stationarity and difference stationarity (an issue that attracted great interest during the last decade) is a problem for which the tapering procedure is especially suitable.

ACKNOWLEDGMENTS

We thank the editor, associate editor, and two referees for valuable comments and useful suggestions that led to a substantial improvement of the article. Velasco's research was funded by Spanish Dirección General de Enseñanza Superior, ref. no. PB98-0025.

APPENDIX: CLT FOR TSE

As shown by Lobato (1999), the TSE has the same asymptotic distribution as the QMLE, and in fact it is given by (18) if

$$\sqrt{m} \frac{\partial \Upsilon_p(d)}{\partial d} \Big|_{d_0} \rightarrow_d N(0, p\Phi_p E)$$

Table 7. Estimated Squared Coherency at Zero Frequency for the Kolmogorv–Zhurbenko Taper of Order $p = 3$

Stock	m										
	60	84	108	132	156	180	204	228	252	276	300
ATT	.000	.015	.010	.032	.032	.029	.029	.029	.058	.044	.052
ALD	.220	.325	.360	.369	.314	.251	.231	.240	.240	.202	.184
AA	.000	.017	.062	.121	.110	.140	.163	.154	.173	.177	.124
AXP	.190	.137	.119	.118	.155	.145	.042	.101	.099	.102	.074
BS	.086	.127	.219	.234	.190	.196	.182	.168	.181	.197	.198
BA	.185	.225	.177	.193	.205	.197	.182	.197	.241	.213	.250
CAT	.100	.165	.125	.137	.125	.137	.109	.076	.073	.071	.098
CHV	.165	.147	.125	.130	.151	.166	.196	.199	.218	.239	.212
KO	.367	.356	.333	.268	.236	.250	.269	.276	.252	.259	.254
DIS	.145	.244	.192	.265	.234	.223	.205	.206	.236	.241	.238
DD	.110	.122	.128	.205	.203	.188	.175	.178	.191	.193	.201
EK	.207	.245	.254	.306	.304	.318	.286	.282	.302	.293	.288
XON	.085	.119	.111	.108	.126	.110	.145	.133	.133	.131	.125
GE	.026	.085	.086	.088	.087	.082	.097	.110	.141	.113	.105
GM	.150	.207	.234	.205	.228	.212	.248	.259	.257	.229	.218
GT	.042	.086	.046	.062	.056	.045	.049	.050	.054	.058	.048
IBM	.270	.321	.253	.216	.227	.235	.222	.220	.268	.222	.225
IP	.192	.243	.273	.233	.226	.260	.239	.213	.209	.209	.206
MCD	.169	.112	.172	.133	.103	.101	.127	.135	.161	.133	.129
MRK	.258	.264	.187	.160	.162	.184	.218	.202	.246	.244	.249
3M	.256	.288	.259	.222	.171	.190	.170	.171	.157	.199	.178
JPM	.055	.059	.045	.032	.027	.083	.097	.092	.114	.099	.096
MO	.048	.024	.095	.132	.171	.180	.146	.119	.165	.187	.179
PG	.110	.174	.126	.150	.157	.142	.147	.163	.179	.171	.166
S	.185	.255	.286	.179	.202	.163	.180	.198	.210	.237	.226
TX	.319	.401	.375	.355	.349	.321	.324	.312	.333	.296	.269
UK	.072	.165	.228	.238	.209	.228	.185	.181	.169	.155	.143
UTX	.040	.086	.124	.145	.137	.125	.135	.138	.159	.147	.145
WX	.120	.101	.091	.133	.147	.139	.150	.146	.160	.142	.161
Z	.234	.365	.359	.287	.236	.236	.228	.207	.180	.185	.189

and

$$\frac{\partial^2 \Upsilon_p(d)}{\partial d \partial d'} \Big|_{\tilde{d}} \rightarrow_p E$$

for all \tilde{d} such that $\|\tilde{d} - d^0\| \leq \|\tilde{d}^{(1)} - d^0\|$,

where E is positive definite. Both statements are proved in Sections A.1 and A.2, respectively. Section A.3 contains important auxiliary results. The matrix E is positive definite as shown by Lobato (1999). Section A.4 contains a sketch of the proof of Theorem 2.

A.1 CLT FOR SCORE

In this section we prove that, for any $N \times 1$ vector η ,

$$\sqrt{m/p} \eta' \frac{\partial \Upsilon_p(d)}{\partial d} \Big|_{d^0} \rightarrow_d N(0, \Phi_p \eta' E \eta).$$

Proof. The proof is similar to the proof of Lobato (1999). We just stress the differences. First, notice that

$$\begin{aligned} \frac{\partial \Upsilon_p(d)}{\partial d_a} &= -\frac{2p}{m} \sum_{j(p)}^m \log j \\ &+ \text{tr} \left[\hat{R}(d)^{-1} \frac{p}{m} \sum_{j(p)}^m (\log j) \Lambda_j^{-1} \right] \end{aligned}$$

$$\times \left\{ i_a I^T(\lambda_j) + I^T(\lambda_j) i_a \right\} \Lambda_j^{-1} \Bigg\}, \quad (\text{A.1})$$

where the matrix i_a is an $N \times N$ matrix with every element equal to 0 except the a th diagonal, which is 1. Then, calling η_a to the a th element of η and using that $\hat{R}(d^0)^{-1} = R^{0^{-1}} + o_p(1)$ [choose $s = m$ in Expression (A.10) in Lemma 1 in Section A.3 and divide by m],

$$\begin{aligned} \sqrt{m/p} \eta' \frac{\partial \Upsilon_p(d)}{\partial d} \Big|_{d^0} &= \frac{2}{\sqrt{m/p}} \sum_{a=1}^N \eta_a \sum_{j(p)}^m \nu_j \\ &\times (r^a \Lambda_j^{0^{-1}} \text{Re}\{I_a^T(\lambda_j)\} \lambda_j^{d_a^0} - 1) + o_p(1), \quad (\text{A.2}) \end{aligned}$$

where ν_j is as given in (22), r^a is the a th row of $R^{0^{-1}}$, and $I_a^T(\lambda_j)$ is the a th column of $I^T(\lambda_j)$. Furthermore, (A.2) is asymptotically equivalent to

$$\begin{aligned} \frac{2}{\sqrt{m/p}} \sum_{a=1}^N \eta_a \sum_{j(p)}^m \nu_j \\ \times (r^a \Lambda_j^{0^{-1}} \text{Re}\{I_a^T(\lambda_j) - A(\lambda_j) J(\lambda_j) A_a^*(\lambda_j)\} \lambda_j^{d_a^0}) \quad (\text{A.3}) \end{aligned}$$

$$\begin{aligned} + \frac{2}{\sqrt{m/p}} \sum_{a=1}^N \eta_a \sum_{j(p)}^m \nu_j \\ \times (r^a \Lambda_j^{0^{-1}} \text{Re}\{A(\lambda_j) J(\lambda_j) A_a^*(\lambda_j)\} \lambda_j^{d_a^0} - 1), \quad (\text{A.4}) \end{aligned}$$

where $J(\lambda) = (2\pi \sum h_t^2)^{-1} |\sum_{t=1}^n h_t \varepsilon_t e^{it\lambda}|^2$. Now using (A.8), proved in Section A.3, and summation by parts, (A.3) is

$$2(m/p)^{-1/2} \sum_{a=1}^N \eta_a \sum_{j(p)}^m \nu_j \sum_{b=1}^N r^{ba} \lambda_j^{d_b^0} \times (\text{Re}\{I_{ba}^T(\lambda_j) - A_b(\lambda_j)J(\lambda_j)A_a^*(\lambda_j)\}) \lambda_j^{d_a^0} = o_p(1),$$

where r^{ab} is the (a, b) th element of $R^{0^{-1}}$. Equation (A.4) is

$$\begin{aligned} & \frac{2}{\sqrt{m/p}} \sum_{a=1}^N \eta_a \sum_{j(p)}^m \nu_j \\ & \times \left(r^a \Lambda_j^{0^{-1}} \text{Re} \left\{ A(\lambda_j) \frac{1}{2\pi \sum h_t^2} \sum_{t=1}^n h_t^2 \varepsilon_t \varepsilon_t' A_a^*(\lambda_j) \right\} \lambda_j^{d_a^0} - 1 \right) \\ & + \frac{2}{\sqrt{m/p}} \sum_{a=1}^N \eta_a \sum_{j(p)}^m \nu_j r^a \Lambda_j^{0^{-1}} \\ & \times \text{Re} \left\{ A(\lambda_j) \frac{1}{2\pi \sum h_t^2} \sum_{t \neq s} h_t h_s \varepsilon_t \varepsilon_s' \right. \\ & \left. \times \exp(i(t-s)\lambda_j) A_a^*(\lambda_j) \right\} \lambda_j^{d_a^0}. \end{aligned}$$

Now the first part is negligible using lemma 1 in appendix 4 of Lobato (1999) and that $\max_t |h_t| \leq 1$, whereas the second part can be written as $\sum_{t=1}^n z_t$, where $z_t = h_t \varepsilon_t' \sum_{s=1}^{t-1} \Theta_{t-s} h_s \varepsilon_s$ with

$$\Theta_t = \left(\pi \sqrt{m/p} \sum_{j(p)}^m h_t^2 \right)^{-1} \sum_{j(p)}^m \nu_j \Omega_j \cos(t\lambda_j)$$

and $\Omega_j = \sum_{a=1}^N \eta_a \text{Re}\{A(\lambda_j)' \Lambda_j^{0^{-1}} r^{a'} \overline{A_a(\lambda_j)} + A_a'(\lambda_j) r^a \Lambda_j^{0^{-1}} \overline{A(\lambda_j)}\} \lambda_j^{d_a^0}$, where the overline indicates conjugation. z_t is a martingale difference, and we can apply a standard CLT for a martingale difference sequence (see Hall and Heyde 1980, sec. 3.2). This entails proving the following:

1. $\sum_{t=1}^n E(z_t^2 | \mathcal{T}_{t-1}) - p \Phi_p \sum_{a=1}^N \sum_{b=1}^N \eta_a \eta_b E_{ab} \rightarrow p \cdot 0$.
2. $\sum_{t=1}^n E(z_t^2 I(|z_t| > \rho)) \rightarrow 0$ for all $\rho > 0$.

Proof of 1. We have that

$$\begin{aligned} \sum_{t=1}^n E(z_t^2 | \mathcal{T}_{t-1}) &= \sum_{t=2}^n \sum_{s=1}^{t-1} h_s^2 \varepsilon_s' \Theta_{t-s} \Theta_{t-s} \varepsilon_s \\ &+ \sum_{t=2}^n \sum_{s \neq s'}^{t-1} h_s h_{s'} \varepsilon_s' \Theta_{t-s} \Theta_{t-s'} \varepsilon_{s'}. \end{aligned}$$

The second term on the right is $o_p(1)$ using lemma 2 in appendix 4 of Lobato (1999) and the fact that $\max_t |h_t| < \infty$. The first one using lemma 3 in appendix 4 of Lobato

(1999) has mean

$$\begin{aligned} & \frac{p}{\pi^2 m (\sum h_t^2)^2} \sum_{j(p)}^m \sum_{j'(p)}^m \nu_j \nu_{j'} \\ & \times \text{tr}[\Omega_j' \Omega_{j'}] \sum_{t=1}^{n-1} \sum_{s=1}^{n-t} h_t^2 h_{s+t}^2 \cos(s\lambda_j) \cos(s\lambda_{j'}), \end{aligned}$$

and using Lemma 2 in Section A.3 we complete the proof for 1.

Proof of 2. We prove the sufficient condition $\sum_{t=1}^n E(z_t^4) \rightarrow 0$. Because

$$\begin{aligned} \sum_{t=1}^n E(z_t^4) &= \sum_{t=1}^n E \left[\sum_{s=1}^{t-1} h_t^4 h_s \varepsilon_s' \Theta_{t-s} \varepsilon_t \varepsilon_t' \sum_{r=1}^{t-1} h_r \Theta_{t-r} \varepsilon_r \right. \\ & \left. \times \sum_{u=1}^{t-1} h_u \varepsilon_u' \Theta_{t-u} \varepsilon_t \varepsilon_t' \sum_{v=1}^{t-1} h_v \Theta_{t-v} \varepsilon_v \right] \\ &\leq C \left\{ \sum_{t=1}^n \text{tr} \left[\sum_{s=1}^{t-1} \Theta_{t-s} \Theta_{t-s} \Theta_{t-s}' \Theta_{t-s} \right] \right. \\ & \left. + \sum_{t=1}^n \text{tr} \left[\sum_{s=1}^{t-1} \Theta_{t-s}' \sum_{r=1}^{t-1} \Theta_{t-r} \Theta_{t-r}' \Theta_{t-s} \right] \right\} \end{aligned}$$

for some finite positive constant C , then $\sum_{t=1}^n E(z_t^4) = O(n^{-1}(\log m)^4)$ as in appendix 1 of Lobato (1999).

A.2 CONVERGENCE OF HESSIAN

In this section we prove that

$$\frac{\partial^2 \Upsilon_p(d)}{\partial d_a \partial d_b} \Big|_{\bar{d}} \rightarrow_p E_{ab}$$

for all \bar{d} such that $\|\bar{d} - d^0\| \leq \|\hat{d}^{(1)} - d^0\|$. (A.5)

Proof. The proof is similar to that of Lobato (1999). Define $d_* = \max_a d_a^0$, and for $k = 0, 1, 2$, define

$$\hat{F}_k(d) = pm^{-1} \sum_{j(p)}^m (\log j)^k \text{diag}\{\lambda_j^{d_a}\} \text{Re}\{I^T(\lambda_j)\} \text{diag}\{\lambda_j^{d_a}\}$$

and

$$\hat{G}_k(d) = pm^{-1} \sum_{j(p)}^m (\log j)^k \text{diag}\{j^{d_a}\} \text{Re}\{I^T(\lambda_j)\} \text{diag}\{j^{d_a}\}.$$

We also employ that $pm^{-1} \sum_{j(p)}^m \log^2 j - (pm^{-1} \sum_{j(p)}^m \log j)^2 \rightarrow 1$ as $m \rightarrow \infty$. What we need to show is that

$$\left\| \hat{F}_k(d^0) - R^0 \frac{p}{m} \sum_{j(p)}^m \log^k j \right\| = o_p(1). \quad (\text{A.6})$$

By summation by parts, the left side of (A.6) is

$$\left\| \frac{p}{m} \sum_{j(p)}^{m-p} ((\log j)^k - (\log(j+p))^k) \sum_{l(p)}^j (\text{diag}\{\lambda_l^{d_a}\}) \right\|$$

$$\begin{aligned} & \times \operatorname{Re}\{I^T(\lambda_l)\} \operatorname{diag}\{\lambda_l^{d_a}\} - R^0) + \frac{p}{m} (\log m)^k \\ & \times \sum_{j(p)}^m (\operatorname{diag}\{\lambda_j^{d_a}\} \operatorname{Re}\{I^T(\lambda_j)\} \operatorname{diag}\{\lambda_j^{d_a}\} - R^0) \Big\| . \end{aligned}$$

The orders of magnitude of

$$\left\| \sum_{j(p)}^m (\operatorname{diag}\{\lambda_j^{d_a}\} \operatorname{Re}\{I^T(\lambda_j)\} \operatorname{diag}\{\lambda_j^{d_a}\} - R^0) \right\|$$

are as follows:

1. For $p = 1$, and $d_* < 1$, adapting Lobato's (1999) procedures when $d_* \geq .5$ as in theorem 3 of Velasco (1999b),

$$\begin{aligned} & O_p(n^{-\beta} m^{\beta+1} + m^{1/(5-4d_*)} (\log m)^{2/(5-4d_*)} \\ & + m^{2d_*-1} \log m + n^{-1/2} m^{(d_*+1)/2} (\log n)^{5/4} \\ & + n^{-1/4} m^{d_*} (\log m)^{1/2} + m^{1/2}) \end{aligned}$$

2. For $p > 1$, $d_* < p$, using (A.10) in Section A.3,

$$O_p(n^{-\beta} m^{\beta+1} + \log n + m^{1/2})$$

Then, using these orders of magnitude, Assumption 4, and that $|(\log j)^k - (\log(j+p))^k| \leq k j^{-1} (\log(j+p))^{k-1}$, the orders of magnitude of the left side of (A.6) are as follows:

1. For $p = 1$, $d_* < 1$,

$$\begin{aligned} & O_p([n^{-\beta} m^{\beta} + m^{d(d_*-1)/(5-4d_*)} (\log m)^{2/(5-4d_*)} \\ & + m^{2(d_*-1)} \log m + n^{-1/2} m^{(d_*-1)/2} (\log n)^{5/4} \\ & + n^{-1/4} m^{d_*-1} (\log m)^{1/2} + m^{-1/2}] \log m) = o_p(1) \end{aligned}$$

2. For $p > 1$, $d_* < p$,

$$O_p([n^{-\beta} m^{\beta} + m^{-1} \log n + m^{-1/2}] \log m) = o_p(1)$$

A.3 AUXILIARY RESULTS

Lemma 1. Under the conditions of Theorem 1,

$$\sum_{j(p)}^s (f_{ab}(\lambda_j) \lambda_j^{d_a^0 + d_b^0} - r_{ab}) = O\left(\frac{s^{\beta+1}}{n^\beta}\right), \quad (\text{A.7})$$

$$\begin{aligned} & \sum_{j(p)}^s \operatorname{Re}\{I_{ab}^T(\lambda_j) - A_a(\lambda_j) J^T(\lambda_j) A_b^*(\lambda_j)\} \lambda_j^{d_a^0 + d_b^0} \\ & = O_p(\log^2 n + [s^{d_*-p+1} + n^{-.5} s^{2(d_*-p+1)}] \log s)^{1/2}, \quad (\text{A.8}) \end{aligned}$$

$$\begin{aligned} & \sum_{j(p)}^s (\operatorname{Re}\{A_a(\lambda_j) J^T(\lambda_j) A_b^*(\lambda_j)\} - f_{ab}(\lambda_j)) \lambda_j^{d_a^0 + d_b^0} \\ & = O_p(s^{1/2}), \quad (\text{A.9}) \end{aligned}$$

and hence

$$\begin{aligned} & \sum_{j(p)}^s (\operatorname{Re}\{I_{ab}(\lambda_j)\} \lambda_j^{d_a^0 + d_b^0} - r_{ab}) \\ & = O_p\left(\frac{s^{\beta+1}}{n^\beta} + \log n + s^{1/2}\right). \quad (\text{A.10}) \end{aligned}$$

Everything, except (A.8) holds as in the work of Lobato (1999), just replacing the raw periodogram by the tapered periodogram and the running index in the summation. Now, the proof of (A.8) is as shown by Lobato (1999) up to the point at which the expectation of the square of the left side of (A.8) is divided into two parts, the terms with fourth-order cumulants and the terms without.

The terms without the cumulants have two parts. For the first part, using theorem 6 of Velasco (1999a) and $d_* + .5 < p$, their order of magnitude is $O(\sum_{j(p)}^s \{j^{-1} + \gamma_{j,j}\}) = O(\log s)$, where $\gamma_{j,k} = (jk)^{d_*-p} \log j$, $j < k$. To analyze the second part, notice that in theorem 6 of Velasco (1999a), in the expression for the covariances of $w_x^T(\lambda_j)$, the terms $|j-k|^{-p}$ are due to the bound for $|H_{j,k}|$, where $H_{j,k} = (2\pi \sum_{t=1}^n h_t^2)^{-1} \int_{-\pi}^{\pi} D^T(\lambda_j - \lambda) D^T(\lambda - \lambda_k) d\lambda$ and $D^T(\lambda_j) = \sum_{t=1}^n h_t \exp(it\lambda_j)$. Therefore, we obtain that $E[v_a^T(\lambda_j) v_a^T(\lambda_k)] = H_{j,k} + O(j^{-1}|j-k|^{-p} \log n + j^{-1}|j-k|^{1-p} + \gamma_{j,k})$, where $v_a^T(\lambda_j) = w_a^T(\lambda_j)/f_{aa}^{1/2}(\lambda_j)$. Notice that the $H_{j,k}$ are identically 0 for nontapered series, $p = 1$, $j \neq k \pmod{n}$, but may not be for $p > 1$. In (4.26) of Robinson (1995b), however, all the terms $|H_{j,k}|^2$ and $|H_{j,-k}|^2$ cancel out by symmetry, obtaining that the order of magnitude of the second term without fourth-order cumulants is

$$\begin{aligned} & O\left(\sum_{j(p)}^s \sum_{k>j}^s \{j^{-2}|j-k|^{-2p} \log^2 n + j^{-2}|j-k|^{2-2p} + \gamma_{j,k}^2\}\right) \\ & + O\left(\sum_{j(p)}^s \sum_{k>j}^s \{j^{-1}|j-k|^{-p} \log n \right. \\ & \left. + j^{-1}|j-k|^{1-p} + \gamma_{j,k}\} |j-k|^{-p}\right) \\ & = O(\log^2 n + s^{d_*-p+1} \log s). \end{aligned}$$

For the terms with fourth-order cumulants, we find the same expressions for tapered data as given by Robinson (1995b) and Lobato (1999), defining for the a th series,

$$\begin{aligned} P_j(a, k_p) &= \int_{-\pi}^{\pi} |A_I^{k_p}(\lambda)(1 - e^{i\lambda})^{-s_a^0}|^2 \\ & - A_I^{k_p}(\lambda_j)(1 - e^{i\lambda_j})^{-s_a^0} K^T(\lambda - \lambda_j) d\lambda, \quad (\text{A.11}) \end{aligned}$$

where $K^T(\lambda) = (2\pi \sum_1^n h_t^2)^{-1} |D^T(\lambda)|^2$. Then applying

Lemma 3 provided at the end of this section, its order of magnitude is $O(\sum_{j(p)}^s \{[\gamma_{j,j} + j^{-2}]^{3/2} + n^{-.5}[j^{-2} + \gamma_{j,j}]\}) =$

$O(1)$ for the first part, and for the second part a typical term has order of magnitude

$$O \left(\begin{array}{c} (\lambda_j \lambda_k)^{d_a^0 + d_b^0} P_j^{1/2}(a, k_1) P_j^{1/2}(b, k_2) P_k^{1/2}(a, k_3) P_k^{1/2}(b, k_4) \\ + P_j^{1/2}(a, k_1) P_j^{1/2}(b, k_2) P_k^{1/2}(b, k_3) \lambda_j^{d_a^0 + d_b^0} \lambda_k^{d_b^0} + n^{-1/2} P_j^{1/2}(a, k_1) P_j^{1/2}(b, k_2) \lambda_j^{d_a^0} \lambda_k^{d_b^0} \end{array} \right).$$

Using the results in Lemma 3, this part is

$$O \left(\begin{array}{c} \sum_{j(p)}^s \sum_{k>j}^s \{ [j^{-2} + \gamma_{j,j}] [k^{-1} + \gamma_{k,k}^{1/2}] \\ + n^{-.5} [j^{-1} + \gamma_{j,j}^{1/2}] [k^{-1} + \gamma_{k,k}^{1/2}] \} \end{array} \right) \\ = O(\log s + s^{d_* - p + 1} \log^{1/2} s + n^{-.5} s^{2(d_* - p + 1)} \log s).$$

Lemma 2. Under the conditions of Theorem 1,

$$\frac{(\sum h_t^2)^{-2}}{\pi^2 m} \sum_{j(p)}^m \sum_{k(p)}^m \nu_j \nu_{k'} \\ \times \text{tr}[\Omega_j' \Omega_k] \sum_{t=1}^{n-1} \sum_{s=1}^{n-t} h_t^2 h_{s+t}^2 \cos(s\lambda_j) \cos(s\lambda_k) \\ = \Phi_p \sum_{a=1}^N \sum_{b=1}^N \eta_a \eta_b E_{ab} + o(1). \quad (\text{A.12})$$

Proof. Proceeding as in lemma 6 of Velasco (1999b), (A.12) is

$$\frac{(\sum h_t^2)^{-2}}{4\pi^2 m} \sum_{j(p)}^m \sum_{k(p)}^m \nu_j \nu_{j'} \text{tr}[\Omega_j' \Omega_k] \\ \times \left[\sum_{t=1}^n h_t^2 \cos t(\lambda_j - \lambda_k) \right]^2 + O(m^{-1}(\log m)^2) \quad (\text{A.13})$$

because $\|\Omega_j\| = O(1)$. Now, using the properties of the data taper transforms and the differentiability of Ω_j and of ν_j similarly to the proof of lemma 7 of Velasco and Robinson (in press), (A.13) is

$$\frac{\Phi_p}{4\pi^2 m} \sum_{j(p)}^m \nu_j^2 \text{tr}[\Omega_j' \Omega_j] + o(1).$$

The rest of the proof follows from appendix 1 of Lobato (1999).

Lemma 3. Under the conditions of Theorem 1, $P_j(a, k_p) = O(f_{aa}(\lambda_j)[j^{-2} + \gamma_{j,j}])$, for $a = 1, \dots, r$, $k_p = 1, 2, 3, 4$, $j = p, 2p, \dots, n - p$, $p > 1$.

Proof. Consider without loss of generality $k_p = 1$ and $a = 1$, and set $P_j = P_j(1, k_1)$. The result in the lemma follows as in lemma 3 of Robinson (1995b) or Velasco (1999b), by decomposing the integral that defines P_j in (A.11) in the following intervals. First, choose $\varepsilon > 0$, fixed, such that Ass-

umption 1 holds for $|\lambda| < \varepsilon$, as in the proof of theorem 2 of Robinson (1995a); then

$$\left| \int_{-\varepsilon}^{-\varepsilon} \right| = O \left(n^{1-2p} \left[\int_{\varepsilon}^{\pi} f_{11}(\lambda) d\lambda + 1 \right] \right) \\ = O(f_{11}(\lambda_j) j^{1-2p}),$$

using that $|D^T(\lambda)| \leq C \min\{n, n^{1-p}|\lambda|^{-p}\}$ for a taper of order $p \geq 1$. Second,

$$\left| \int_{-\varepsilon}^{-\lambda_j/2} \right| \\ = O \left(n^{1-2p} \left[\int_{\lambda_j}^{\pi} \lambda^{-2(d+p)} d\lambda + f_{11}(\lambda_j) \int_{\lambda_j}^{\pi} \lambda^{-2p} d\lambda \right] \right) \\ = O(n^{1-2p} [\lambda_j^{1-2(d+p)} + f_{11}(\lambda_j) \lambda_j^{1-2p}]) \\ = O(f_{11}(\lambda_j) j^{1-2p}).$$

Third,

$$\left| \int_{-\lambda_j/2}^{\lambda_j/2} \right| = O \left(\int_{-\lambda_j/2}^{\lambda_j/2} f_{11}(\lambda) K^T(\lambda - \lambda_j) d\lambda \right. \\ \left. + f_{11}(\lambda_j) |\lambda_j| \sup_{-\lambda_j/2 \leq \lambda \leq \lambda_j/2} K^T(\lambda - \lambda_j) \right) \\ = O(f_{11}(\lambda_j) [\gamma_{j,j} + n^{1-2p} \lambda_j^{1-2p}]) \\ = O(f_{11}(\lambda_j) [\gamma_{j,j} + j^{1-2p}]),$$

using the same method as that of Velasco (1999a, pp. 356–357) because $f_{11}(\lambda)$ might not be integrable. Finally, using Assumption 3,

$$\left| \int_{\lambda_j/2}^{2\lambda_j} \right| = O \left(\sup_{\lambda_j/2 \leq \lambda \leq 2\lambda_j} \left\| \frac{d}{d\lambda} A_1(\lambda) \right\|^2 \right. \\ \left. \times \int_{\lambda_j/2}^{2\lambda_j} |\lambda - \lambda_j|^2 K^T(\lambda - \lambda_j) d\lambda \right) \\ = O(f_{11}(\lambda_j) j^{-2}),$$

as $\int_0^{2\lambda_j} |\lambda|^2 K^T(\lambda) d\lambda = O(n^{-2})$, $p > 1$.

A.4 CLT FOR COHERENCE ESTIMATES

We need to prove under the conditions of Theorem 2 that for any $N' \times 1$ vector η

$$\sqrt{m}(\text{vec}(\hat{R}(\hat{d})) - \text{vec}(R)) \rightarrow_d N(0, p\Phi_p \eta' F \eta).$$

Proof. We follow as in Section A.1. First notice that

$$\begin{aligned} & \sqrt{m/pn}'(\text{vec}(\hat{R}(\hat{d})) - \text{vec}(\hat{R}(d_0))) \\ &= \frac{1}{\sqrt{m/p}} \sum_{a=1}^{N'} \sum_{b=1}^{N'} \eta_{ab} \sum_{j(p)}^m (\lambda_j^{d_a^o + d_b^o} \text{Re}[I_{ab}^T(\lambda_j)] - R_{ab}) \\ & \quad \times (\lambda_j^{\hat{d}_a + \hat{d}_b - d_a^o - d_b^o} - 1) = o_p(1), \end{aligned}$$

using that

$$\begin{aligned} \sup_{p \leq j \leq m} |\lambda_j^{\hat{d}_a + \hat{d}_b - d_a^o - d_b^o} - 1| &= O_p(\log n \sup_a |\hat{d}_a - d_a^o|) \\ &= O_p(m^{-1/2} \log n). \end{aligned}$$

Then

$$\begin{aligned} & \sqrt{m/pn}'(\text{vec}(\hat{R}(d_0)) - \text{vec}(R)) \\ &= \frac{1}{\sqrt{m/p}} \sum_{a=1}^{N'} \sum_{b=1}^{N'} \eta_{ab} \sum_{j(p)}^m \lambda_j^{d_a^o + d_b^o} \\ & \quad \times \text{Re}[I_{ab}^T(\lambda_j) - A_a(\lambda_j)J(\lambda_j)A_b^*(\lambda_j)] \quad (\text{A.14}) \end{aligned}$$

$$\begin{aligned} & + \frac{1}{\sqrt{m/p}} \sum_{a=1}^{N'} \sum_{b=1}^{N'} \eta_{ab} \sum_{j(p)}^m \\ & \quad \times (\lambda_j^{d_a^o + d_b^o} \text{Re}[A_a(\lambda_j)J(\lambda_j)A_b^*(\lambda_j)] - R_{ab}), \quad (\text{A.15}) \end{aligned}$$

where (A.14) is $o_p(1)$ using (A.8) proved in Section A.3 and (A.15) is

$$\begin{aligned} & \frac{1}{\sqrt{m/p}} \sum_{a=1}^{N'} \sum_{b=1}^{N'} \eta_{ab} \sum_{j(p)}^m \\ & \quad \times \left(\lambda_j^{d_a^o + d_b^o} \text{Re} \left\{ A_a(\lambda_j) \frac{1}{2\pi \sum h_t^2} \cdot \right. \right. \\ & \quad \quad \left. \left. \times \sum_{t=1}^n h_t^2 \varepsilon_t \varepsilon_t' A_b^*(\lambda_j) \right\} - R_{ab} \right) \\ & + \frac{1}{\sqrt{m/p}} \sum_{a=1}^{N'} \sum_{b=1}^{N'} \eta_{ab} \sum_{j(p)}^m \\ & \quad \times \left(\lambda_j^{d_a^o + d_b^o} \text{Re} \left\{ A_a(\lambda_j) \frac{1}{2\pi \sum h_t^2} \sum_{t \neq s}^n \right. \right. \\ & \quad \quad \left. \left. \times h_t h_s \varepsilon_t \varepsilon_s' \exp(i(t-s)\lambda_j) A_b^*(\lambda_j) \right\} \right). \end{aligned}$$

Now the first part is negligible using lemma 1 in appendix 4 of Lobato (1999) and the second part can be written as $\sum_{t=1}^n \tilde{z}_t$, where $\tilde{z}_t = h_t \varepsilon_t' \sum_{s=1}^{t-1} \tilde{\Theta}_{t-s} h_s \varepsilon_s$ with

$$\tilde{\Theta}_t = \left(2\pi \sqrt{m/p} \sum h_t^2 \right)^{-1} \sum_{j(p)}^m \tilde{\Omega}_j \cos(t\lambda_j)$$

and $\tilde{\Omega}_j = \sum_{a=1}^{N'} \sum_{b=1}^{N'} \eta_{ab} \lambda_j^{d_a^o + d_b^o} [A_a'(\lambda_j) \overline{A_b(\lambda_j)} + A_b'(\lambda_j) \overline{A_a(\lambda_j)}]$. Therefore \tilde{z}_t is a martingale difference, and we can apply a standard CLT for a martingale difference sequence as in Section A.1. This entails proving the following:

1. $\sum_{t=1}^n E(|\tilde{z}_t|^2 | \mathcal{T}_{t-1}) - p \Phi_p \sum_{a=1}^{N'} \sum_{b=1}^{N'} \sum_{r=1}^{N'} \sum_{s=1}^{N'} \eta_{ab} \eta_{rs} E_{ab,rs}^* \rightarrow_p 0$.
2. $\sum_{t=1}^n E(|\tilde{z}_t|^2 I(|z_t| > \rho)) \rightarrow 0$ for all $\rho > 0$.

Both follow as in Section A.1; the main difference is that now, $D = d_a^o + d_b^o + d_r^o + d_s^o$,

$$\begin{aligned} & \text{tr}[\tilde{\Omega}_j' \tilde{\Omega}_j] \\ &= \sum_{a,b,r,s=1}^{N'} \eta_{ab} \eta_{rs} \lambda_j^D \text{tr} \{ [A_a^*(\lambda_j) A_r(\lambda_j) + A_r^*(\lambda_j) A_s(\lambda_j)] \\ & \quad \times [A_a^*(\lambda_j) A_b(\lambda_j) + A_b^*(\lambda_j) A_a(\lambda_j)] \}, \end{aligned}$$

and this is $(2\pi)^2$ times

$$\begin{aligned} & \sum_{a,b,r,s=1}^{N'} \eta_{ab} \eta_{rs} \lambda_j^D \{ f_{bs}(\lambda_j) f_{ra}(\lambda_j) + f_{sb}(\lambda_j) f_{ar}(\lambda_j) \\ & \quad + f_{br}(\lambda_j) f_{sa}(\lambda_j) + f_{rb}(\lambda_j) f_{as}(\lambda_j) \} \\ &= 2 \sum_{a,b,r,s=1}^{N'} \eta_{ab} \eta_{rs} \lambda_j^D \\ & \quad \times \{ \text{Re}[f_{bs}(\lambda_j) f_{ra}(\lambda_j)] + \text{Re}[f_{rb}(\lambda_j) f_{as}(\lambda_j)] \} \\ &= 2 \sum_{a,b,r,s=1}^{N'} \eta_{ab} \eta_{rs} \{ R_{bs} R_{ra} + R_{rb} R_{as} \} + o(1), \end{aligned}$$

as $n \rightarrow \infty$, following the expression for F .

[Received February 1999. Revised January 2000.]

REFERENCES

- Alekseev, V. G. (1996), "Jackson- and Jackson-Vallée Poussin-type Kernels and Their Probability Applications," *Theory of Probability and its Applications*, 41, 137–143.
- Andersen, T. G. (1996), "Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility," *Journal of Finance*, 60, 169–204.
- Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996), "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 74, 3–30.
- Bollerslev, T., and Jubinski, D. (1999), "Equity Trading Volume and Volatility: Latent Information Arrivals and Common Long-run Dependencies," *Journal of Business & Economic Statistics*, 17, 9–21.
- Bollerslev, T., and Mikkelsen, H. O. (1996), "Modeling and Pricing Long Memory in Stock Market Volatility," *Journal of Econometrics*, 73, 151–184.
- Brillinger, D. R. (1975), *Time Series, Data Analysis and Theory*, San Francisco: Holden-Day.
- Chan, H. K., Hayya, J. C., and Ord, K. J. (1977), "A Note on Trend Removal Methods: The Case of Polynomial Trend Versus Variate Differencing," *Econometrica*, 45, 737–744.
- Clark, P. K. (1973), "A Subordinated Stochastic Process Model With Finite Variance for Speculative Prices," *Econometrica*, 41, 135–155.

- Cooley, J. W., and Tukey, J. W. (1965), "An Algorithm for the Machine Calculation of Complex Fourier Series," *Mathematics of Computation*, 19, 267–301.
- Dahlhaus, R. (1988), "Small Sample Effects in Time Series Analysis: A New Asymptotic Theory and a New Estimate," *The Annals of Statistics*, 16, 808–841.
- Deo, R., and Hurvich, C. (1998), "Linear Trend With Fractionally Integrated Errors," *Journal of Time Series Analysis*, 19, 379–397.
- Ding, Z., Granger, C. W. J., and Engle, R. F. (1993), "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83–106.
- Durlauf, N., and Phillips, P. C. B. (1988), "Trends Versus Random Walks in Time Series Analysis," *Econometrica*, 56, 1333–1354.
- Epps, T. W., and Epps, M. L. (1976), "The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture-of-Distributions Hypothesis," *Econometrica*, 44, 305–321.
- Gallant, A. F., Rossi, P. E., and Tauchen, G. E. (1992), "Stock Prices and Volume," *Review of Financial Studies*, 5, 199–242.
- Gewu, J., and Porter-Hudak, S. (1983), "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis*, 4, 221–238.
- Gil Alaña, L. A., and Robinson, P. M. (1997), "Testing of Unit Root and Other Non-stationary Hypotheses in Macroeconomic Time Series," *Journal of Econometrics*, 80, 241–268.
- 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," *Probability Theory and Related Fields*, 86, 87–104.
- Hall, P., and Heyde, C. C. (1980), *Martingale Limit Theory and its Application*, New York: Academic Press.
- Hauser, M. A. (1999), "Maximum Likelihood Estimates for ARMA and ARFIMA Models: A Monte Carlo Study," *Journal of Statistical Planning and Inference*, 80, 229–255.
- Henry, M., and Robinson, P. M. (1996), *Bandwidth Choice in Gaussian Semiparametric Estimation of Long Range Dependence* (Lecture Note in Statistics 115), New York: Springer-Verlag.
- Hosoya, Y. (1997), "A Limit Theory for Long-Range Dependence and Statistical Inference on Related Models," *The Annals of Statistics*, 25, 105–137.
- Hurvich, C. M., and Ray, B. K. (1995), "Estimation of the Memory Parameter for Nonstationary or Non-invertible Fractionally Integrated Processes," *Journal of Time Series Analysis*, 16, 17–42.
- Künsch, H. R. (1987), "Statistical Aspects of Self-Similar Processes," in *Proceedings of the First World Congress of the Bernoulli Society*, Tashkent: VNU Science Press, pp. 67–74.
- Lamoureux, C. G., and Lastrapes, W. D. (1994), "Endogenous Trading Volume and Momentum in Stock-return Volatility," *Journal of Business & Economic Statistics*, 12, 253–260.
- Lobato, I. N. (1999), "A Semiparametric Two Step Estimator in a Multivariate Long Memory Model," *Journal of Econometrics*, 90, 129–153.
- Lobato, I. N., and Robinson, P. M. (1998), "A Nonparametric Test for $I(0)$," *Review of Economic Studies*, 65, 475–495.
- Lobato, I. N., and Savin, N. E. (1998), "Real and Spurious Long-Memory Properties of Stock-Market Data" (with discussion), *Journal of Business & Economic Statistics*, 16, 261–283.
- Nelson, C. R., and Kang, H. (1981), "Spurious Periodicity in Inappropriately Detrended Series," *Econometrica*, 49, 741–751.
- (1984), "Pitfalls in the Use of Time as an Explanatory Variable in Regression," *Journal of Business & Economic Statistics*, 2, 73–82.
- Ray, B. K., and Tsay, R. S. (1998), "Identifying Common Long-range Dependence in a Vector Time Series," unpublished manuscript, University of Chicago, Graduate School of Business.
- Robinson, P. M. (1986), "On the Errors-in-Variables Problem for Time Series," *Journal of Multivariate Analysis*, 19, 240–250.
- (1994a), "Semiparametric Analysis of Long-Memory Time Series," *The Annals of Statistics*, 22, 515–539.
- (1994b), "Rates of Convergence and Optimal Spectral Bandwidth for Long Range Dependence," *Probability Theory and Related Fields*, 99, 443–473.
- (1995a), "Log-Periodogram Regression of Time Series with Long Range Dependence," *The Annals of Statistics*, 23, 1048–1072.
- (1995b), "Gaussian Semiparametric Estimation of Long Range Dependence," *The Annals of Statistics*, 23, 1630–1661.
- (1997), "Large Sample Inference for Nonparametric Regression With Dependent Errors," *The Annals of Statistics*, 25, 2054–2083.
- Robinson, P. M., and Henry, M. (1999), "Long and Short Memory Conditional Heteroskedasticity in Estimating the Memory Parameter of Levels," *Econometric Theory*, 15, 299–336.
- Robinson, P. M., and Marinucci, D. (1998), "Semiparametric Frequency Domain Analysis of Fractional Cointegration," STICERD working paper, EM/98/350, London School of Economics.
- Tauchen, G. E., and Pitts, M. (1983), "The Price Variability-Volume Relationship on Speculative Markets," *Econometrica*, 51, 485–505.
- Velasco, C. (1999a), "Non-stationary Log-periodogram Regression," *Journal of Econometrics*, 91, 325–371.
- (1999b), "Gaussian Semiparametric Estimation of Non-stationary Time Series," *Journal of Time Series Analysis*, 20, 87–127.
- Velasco, C., and Robinson, P. M. (in press), "Whittle Pseudo-Maximum Likelihood Estimation of Non-stationary Time Series," *Journal of the American Statistical Association*, 95.
- Yajima, Y. (1988), "On Estimation of a Regression Model With Long Memory Stationary Error," *The Annals of Statistics*, 16, 191–807.
- Zhurbenko, I. G. (1979), "On the Efficiency of Estimates of a Spectral Density," *Scandinavian Journal of Statistics*, 6, 49–56.