Optimal Time Interval Selection in Long-Run Correlation Estimation

Pedro H. Albuquerque

College of Business Administration
Texas A&M International University

JEL Classification: C14
Keywords: Long-Run Correlation, Bartlett, Lag Selection, Time Interval, Alignment, Newey-West, Andrews-Monahan

Abstract

This paper presents an asymptotically optimal time interval selection criterion for the long-run correlation block estimator (Bartlett kernel estimator) based on the Newey-West and Andrews-Monahan approaches. An alignment criterion that enhances finite-sample performance is also proposed. The procedure offers an optimal yet unobtrusive alternative to the common practice in finance and economics of arbitrarily choosing time intervals or lags in correlation studies. A Monte Carlo experiment using parameters derived from Dow Jones returns data confirms that the procedures are MSE-superior to typical alternatives such as aggregation over arbitrary time intervals, parametric VAR estimation, and Newey-West covariance matrix estimation with automatic lag selection.

This paper is based on a chapter of my Ph.D. Dissertation at the University of Wisconsin – Madison. I am grateful to F. Araújo, S. Durlauf, J. Faria, B. Hansen, M. Hirukawa, M. Horvath, Y. Kitamura, B. Pötscher, B. Tabak, K. D. West, and two anonymous referees for many helpful comments and suggestions. I am nevertheless solely responsible for remaining errors.

Correspondence address: 5201 University Blvd., Laredo, TX, 78041-1920, USA. E-mail: pedrohalbuquerque@yahoo.com. Phone: 956-326-2510. Fax: 956-326-2494.

1 Introduction

Correlation measures are frequently used in economics and finance to characterize the relations between pairs of time series, many times as a prelude to more detailed empirical analyses. Estimation procedures however are usually nonoptimal. For example, in studies of the relation between stock returns, a typical approach is to apply a simple correlation estimator to weekly or monthly aggregate returns, despite the fact that data is usually available at higher frequencies.[1] Another common procedure is the use of a VAR to estimate correlation measures, where the number of lags is chosen arbitrarily or using an information criterion.[2]

Furthermore, time series correlation studies are many times concerned with permanent relations. For example, studies of stock market returns may want to characterize only the long-run relationship, filtering out the effects of reversible components. Long-run estimates tend to be heuristically approximated through the choice of return horizons or lags that are considered long enough to capture permanent effects. There are however two problems with this common practice. When the procedure involves time aggregation of high-frequency data, there may be unnecessary loss of information. Additionally, there is usually no concern for the optimality of time interval or lag choices.

For instance, a researcher may have access to daily stock return data, and yet choose to use aggregate monthly return data in order to estimate return correlations. The procedure though is inefficient, since the aggregation of daily data into monthly data can be improved on by using an estimator based on overlapping monthly returns defined on a daily basis, in other words, a *block estimator*. Additionally, the choice of a monthly time interval

[1] Examples can be found in Campbell and Ammer (1993) and Ammer and Mei (1996).
[2] See for example King and Watson (1994) and Forbes and Rigobon (2002).

is arbitrary and not necessarily optimal, particularly when the explicit goal of the analyst is the characterization of the relation between permanent components. For example, a monthly time interval may imply a return horizon that is or too short, and thus unable to rule out all temporary effects, or too long, and consequently less efficient.

One can propose therefore a better procedure that has the additional advantage of not significantly departing from common practice in time series correlation studies. The procedure employs a block estimator and an optimal time interval selection criterion, or, formally, it uses a nonparametric consistent long-run correlation estimator based on the k-lag difference correlation estimator (Bartlett kernel estimator), combined with automatic lag selection and alignment criteria. Notice that, in the case of the block estimator, the choice of a lag (kernel size k) is equivalent to the choice of a time interval. This property is exactly what makes the procedure simple, convenient and familiar to practitioners.

The methodological approach follows Andrews and Monahan (1992) and Newey and West (1994). The automatic lag selection criterion is based on the minimization of the asymptotic MSE, leading to a time interval choice that is long enough to minimize the estimator bias and short enough to minimize the estimator variance.

It may appear to some that an optimal criterion for long-run correlation estimation is unnecessary, since the Newey-West procedure can be used to optimally estimate long-run covariance matrices and, therefore, to calculate long-run correlations. However, due to the way it is constructed, the Newey-West procedure does not guarantee the optimal estimation of correlation, as will be discussed later.

A Monte Carlo experiment using a VMA(5) and GARCH(1,1) parameters derived from Dow Jones returns data is used to evaluate the effectiveness of the lag selection and alignment criteria. The proposed estimator proves to be adequate and MSE-superior to commonly employed alternatives.

2 Long-Run Correlation

Long-run correlation can be defined using the concept of complex coherency from spectral analysis.[3] Following Priestley (1981), given two $I(0)$ (stationary) variables x_t and y_t, which are the first differences of two $I(1)$ (nonstationary) variables X_t and Y_t, the coherency at frequency ω can be interpreted as the correlation between the random coefficients of the spectral components of x_t and y_t at frequency ω:

$$C(\omega) = \frac{s_{xy}(\omega)}{\sqrt{s_{xx}(\omega)s_{yy}(\omega)}}\,,\ |C(\omega)| \le 1\,,\ |s_{xx}(\omega)| > 0\,,\ |s_{yy}(\omega)| > 0\,,$$

where $s_{xx}(\omega)$ and $s_{yy}(\omega)$ are the spectra and $s_{xy}(\omega)$ is the cross-spectrum, or, in other words, the Fourier transforms of autocovariances and cross-covariances, given by

$$s_{xx}(\omega) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\gamma_{xx}(n)e^{-i\omega n}\,,\ s_{yy}(\omega) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\gamma_{yy}(n)e^{-i\omega n}\,,\ s_{xy}(\omega) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\gamma_{xy}(n)e^{-i\omega n}\,,$$

$$\gamma_{xx}(n) = E[(x_t - \mu_x)(x_{t-n} - \mu_x)],\ \gamma_{yy}(n) = E[(y_t - \mu_y)(y_{t-n} - \mu_y)]\,,$$

$$\gamma_{xy}(n) = E[(x_t - \mu_x)(y_{t-n} - \mu_y)]\,,\quad \mu_x = E[x_t],\quad \mu_y = E[y_t],$$

where γ represents the autocovariances or the cross-covariances of x_t and y_t.[4]

The time-domain concept of *long-run correlation*, which applies to pairs of $I(0)$ variables, is equivalent to the frequency-domain concept of *complex coherency at frequency zero* ($\omega = 0$),[5]

---

[3] See for example Granger & Rees (1968), Granger and Engle (1983), and McCallum (1984).

[4] See Anderson (1971), Koopmans (1974), Fuller (1976), Priestley (1981), Granger and Watson (1984), and Brockwell and Davis (1991).

[5] Notice that if the levels X_t and Y_t were $I(0)$ then the first differences x_t and y_t would be overdifferenced and the long-run correlation parameter λ would not be defined, since s_{xx} and s_{yy} would be zero. In other words, overdifferencing has the effect of a high-pass filter, eliminating all long-run information from a variable.

$$\lambda \equiv \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}}, \quad -1 \le \lambda \le 1, \; s_{xx} > 0, \; s_{yy} > 0, \tag{2.1}$$

where
$$s_{xy} = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\gamma_{xy}(n), \; s_{xx} = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\gamma_{xx}(n), \; s_{yy} = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\gamma_{yy}(n). \tag{2.2}$$

Granger and Weiss (1983) and Engle and Granger (1987) noted that, under certain cointegration literature assumptions, two $I(1)$ variables cointegrate if and only if their first differences have squared long-run correlation equal to one. On the other hand, zero long-run correlation between first differences will only imply the absence of structural long-run relation between two $I(1)$ variables under certain identification restrictions.[6]

3 Nonparametric Estimation of Long-Run Correlation

This section presents a nonparametric estimator of long-run correlation based on the block estimator (k-lag difference) approach of Bartlett (1950), Cochrane (1988) and Cochrane and Sbordone (1988). Cochrane for example presented a nonparametric statistic for unit root processes called the variance ratio, based on the k-lag difference variance of a series. This paper goes a few steps further in this line of research by developing automatic time interval selection and alignment criteria for the long-run correlation block estimator.

---

[6] See for example Fisher and Seater (1993). Notice also that the fact that two variables do not cointegrate does not imply the absence of structural long-run relationship; see for example McCallum (1993) for a discussion. Moreover, under certain identification assumptions, a nonzero long-run correlation value will indicate the existence of structural long-run relationship, even in the absence of cointegration; see Fisher and Seater (1993) for an example using VARMA models.

3.1 Estimator

Consider two random variables x_t and y_t with summable covariances and autocovariances. Given a sample of size T, $1 \le t \le T$, an analog estimator of the long-run correlation defined in (2.1) is a kernel estimator

$$\hat{\lambda}_k = \frac{\hat{s}_{xy}(k)}{\sqrt{\hat{s}_{xx}(k)\hat{s}_{yy}(k)}}, \tag{3.1}$$

where the cross-spectrum and spectrum estimators are

$$\hat{s}_{xy}(k) = \frac{1}{2\pi}\sum_{n=-T+1}^{T-1}\kappa(n,k)\hat{\gamma}_{xy}(n), \qquad \hat{s}_{xx}(k) = \frac{1}{2\pi}\sum_{n=-T+1}^{T-1}\kappa(n,k)\hat{\gamma}_{xx}(n),$$

$$\hat{s}_{yy}(k) = \frac{1}{2\pi}\sum_{n=-T+1}^{T-1}\kappa(n,k)\hat{\gamma}_{yy}(n),$$

the cross-covariance and autocovariance estimators are

$$\hat{\gamma}_{xy}(n) = \begin{cases} \dfrac{1}{T}\displaystyle\sum_{t=1}^{T+n}(x_t - \bar{\mu}_x)(y_{t-n} - \bar{\mu}_y), & n < 0, \\ \dfrac{1}{T}\displaystyle\sum_{t=1}^{T-n}(x_{t+n} - \bar{\mu}_x)(y_t - \bar{\mu}_y), & n \ge 0, \end{cases}$$

$$\hat{\gamma}_{xx}(n) = \frac{1}{T}\sum_{t=1}^{T-|n|}(x_{t+|n|} - \bar{\mu}_x)(x_t - \bar{\mu}_x), \quad \bar{\mu}_x = \frac{1}{T}\sum_{t=1}^{T} x_t, \tag{3.2}$$

$$\hat{\gamma}_{yy}(n) = \frac{1}{T}\sum_{t=1}^{T-|n|}(y_{t+|n|} - \bar{\mu}_y)(y_t - \bar{\mu}_y), \quad \bar{\mu}_y = \frac{1}{T}\sum_{t=1}^{T} y_t, \tag{3.3}$$

and $\kappa(n,k)$ is a kernel with bandwidth k, $1 \le k \le T-1$.[7]

Different kernels can be used, each one having advantages and disadvantages, as discussed in Newey and West (1987, 1994) and Andrews (1991). The Newey-West (Bartlett) kernel approach is chosen in this paper, due to its block estimator equivalence, as discussed in what follows.[8]

---

[7] See Priestley (1981, pg. 432).

[8] Andrews (1991) proved that the QS kernel is optimal with respect to an asymptotic truncated MSE criterion among kernels that generate positive semi-definite covariance estimates, such as the Bartlett kernel. This result however should not be seen as definitive evidence against the use of the Bartlett kernel, since the latter may perform better than the

The Bartlett kernel is defined as

$$\kappa(n,k)=\begin{cases}1-|n|/k\,, & |n|<k,\\ 0, & |n|\geq k.\end{cases}$$

Consider now the expression for the covariance block estimator (k-lag difference estimator) divided by k:[9]

$$\frac{\hat{\sigma}_{XY}(k)}{k}=\frac{1}{k}\sum_{t=k}^{T}\frac{\left[\left(1-L^k\right)X_t-k\bar{\mu}_x\right]\left[\left(1-L^k\right)Y_t-k\bar{\mu}_y\right]}{T-k}\,, \tag{3.4}$$

$$\left(1-L^k\right)X_t=X_t-X_{t-k}\,,\quad \left(1-L^k\right)Y_t=Y_t-Y_{t-k}\,,$$

where X_t and Y_t are the integrated versions of x_t and y_t, given by

$$X_t=X_0+\sum_{i=1}^{t}x_i\,,\; Y_t=Y_0+\sum_{j=1}^{t}y_j\,,\text{ for } 1\leq t\leq T\,.$$

As shown in Appendix 1,[10] the block estimator is asymptotically equivalent to 2π times the Bartlett kernel estimator:

$$\frac{\hat{\sigma}_{XY}(k)}{k}=\frac{T}{T-k}2\pi\hat{s}_{xy}(k)+O_p\left(\frac{k}{T}\right),$$

and the correlation block estimator can be used therefore to estimate long-run correlation instead of equation (3.1), holding the same asymptotic properties:

$$\hat{\lambda}_k=\frac{\hat{\sigma}_{XY}(k)}{\sqrt{\hat{\sigma}_{XX}(k)\,\hat{\sigma}_{YY}(k)}}\,.$$

The block estimator has convenient features: it is just a simple correlation estimator applied to changes measured over time intervals of size

---

QS kernel with finite samples, as discussed in Andrews (1991) and as shown in Newey and West (1994). Newey and West argue for example that the Bartlett kernel performs better when processes are characterized by autocorrelations that “die out slowly” – a case frequently found in economic data, see for example Cochrane (1988). Moreover, as discussed in Hannan (1970, pg. 287), the QS kernel, differently from the Bartlett kernel, assumes negative values, leading to long-run correlation estimates that are not bounded (in absolute value) by unity, a property that, besides being undesirable by itself, complicates the derivation of the statistical properties of the estimator.

[9] As in Cochrane (1988), the substitution of k times the average of the first differences for the average of the k-lag differences is used as a finite-sample enhancement.

[10] See also Cochrane (1988).

k, and even though it is easy to calculate and interpret due to its straightforward time-domain representation, its asymptotic properties can be derived using frequency-domain methods.

Practitioners have used similar yet less efficient procedures by arbitrarily selecting time intervals or lags and by aggregating high-frequency data into longer intervals, perhaps unaware of the statistical implication of their choices. The estimator proposed here can be seen therefore as an optimal yet unobtrusive alternative to commonly employed methods.

3.2 Consistency

As shown in Andrews (1991), "automatic bandwidth kernel estimators are consistent with nonstationary as well as fourth order stationary random variables." What follows is a summary of his results.

The Bartlett kernel estimator of the spectral matrix at frequency zero will converge in probability to its true value when the bandwidth, as a function of the sample size, has the properties

$$\lim_{T\to\infty} k(T) = \infty \text{ and } \lim_{T\to\infty} \frac{k(T)^2}{T} = 0, \qquad (3.5)$$

and given the following assumptions:[11]

(a) the random variables x_t and y_t have summable covariance matrix Euclidian norms and summable fourth order cumulants, or, in other words, they are unconditionally weakly stationary (what allows for autocorrelation and conditional heteroscedasticity, but does not allow for unconditional heteroscedasticity);

(b) $\sqrt{T}(\bar{\mu} - \mu) = O_p(1)$ in equations (3.2) and (3.3); and

(c) the random variables have summable variances.

Moreover, if

[11] A few other assumptions in Andrews will automatically hold in the context of this paper, so they are not presented.

$$\lim_{T\to\infty}\frac{k(T)^{2q+1}}{T}=\xi\in(0,\infty) \quad \text{and} \quad \left\|\begin{matrix} s_{xx}^{(q)} & s_{xy}^{(q)} \\ s_{xy}^{(q)} & s_{yy}^{(q)} \end{matrix}\right\|<\infty$$

for some $q\in(0,\infty)$, where the smoothness parameters are given by

$$s_{xx}^{(q)}=\frac{1}{2\pi}\sum_{n=-\infty}^{\infty}|n|^q\gamma_{xx}(n),\quad s_{yy}^{(q)}=\frac{1}{2\pi}\sum_{n=-\infty}^{\infty}|n|^q\gamma_{yy}(n), \text{ and } s_{xy}^{(q)}=\frac{1}{2\pi}\sum_{n=-\infty}^{\infty}|n|^q\gamma_{xy}(n),$$

then it will be enough that

$$\lim_{T\to\infty}k(T)=\infty \text{ and } \lim_{T\to\infty}\frac{k(T)}{T}=0\,. \tag{3.6}$$

The results above can be extended to the unconditionally heteroscedastic case by assuming summable (across lags) supremum (across time) of covariance matrix Euclidian norms and summable (across lags) supremum (across time) of absolute fourth order cumulants.

Condition (3.5) will hold for automatic lag selection procedures (like the one discussed in the next subsection) as long as the estimated lag selection parameter converges in probability to some nonzero and finite value. Additionally, condition (3.6) will hold for those procedures as long as the estimated lag selection parameter converges in probability to its true value at a rate $T^{1/2}$ or faster.

These consistency properties also apply to the Bartlett kernel estimator of long-run correlation since, under the conditions for (3.5), and using a Taylor expansion around s_{xx}, s_{yy}, and s_{xy}, the long-run correlation estimator can be expressed as

$$\hat{\lambda}(k)=\lambda+\mathbf{D}_\lambda\begin{bmatrix}\hat{s}_{xx}(k)-s_{xx}\\ \hat{s}_{yy}(k)-s_{yy}\\ \hat{s}_{xy}(k)-s_{yy}\end{bmatrix}+O_p\left(\left(\frac{k^2}{T}\right)^{\frac{1}{2}}\right),$$

where

$$\mathbf{D}_\lambda=\begin{bmatrix}\frac{\partial\lambda}{\partial s_{xx}} & \frac{\partial\lambda}{\partial s_{yy}} & \frac{\partial\lambda}{\partial s_{xy}}\end{bmatrix}=\begin{bmatrix}\frac{-\lambda}{2s_{xx}} & \frac{-\lambda}{2s_{yy}} & \frac{1}{\sqrt{s_{xx}s_{yy}}}\end{bmatrix},$$

or, in other words, the long-run correlation estimator is asymptotically equivalent to a linear combination of spectral matrix estimators at frequency zero and therefore has the same consistency properties of these estimators.

3.3 Lag Selection

According to Andrews (1991), the following MSE-minimization procedure will be valid as long as the assumptions of the previous subsection hold and the random variables are eighth order stationary. Using the same approach of Newey and West (1994), an optimal lag selection criterion minimizes the asymptotic MSE (AMSE) of the long-run correlation estimator by exploiting the trade-off between the asymptotic variance (Avar), obtained in Appendix 2,

$$\text{Avar}\left(\hat{\lambda}_k\right) = \frac{2}{3}\frac{k}{T}\left(1-\lambda^2\right)^2, \tag{3.7}$$

and the square of the asymptotic bias (Abias), obtained in Appendix 3,

$$\text{Abias}\left(\hat{\lambda}_k\right)^2 = \frac{\psi^2}{k^2}, \tag{3.8}$$

where

$$\psi = \frac{s_{xy}^{(1)}}{\sqrt{s_{xx}s_{yy}}} - \frac{\lambda}{2}\left(\frac{s_{xx}^{(1)}}{s_{xx}} + \frac{s_{yy}^{(1)}}{s_{yy}}\right),$$

$$s_{xx}^{(1)} = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}|n|\gamma_{xx}(n),\ s_{yy}^{(1)} = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}|n|\gamma_{yy}(n), \tag{3.9}$$

and

$$s_{xy}^{(1)} = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}|n|\gamma_{xy}(n). \tag{3.10}$$

The following proposition is proved in Appendix 4:

Proposition 1: *when* $\lambda^2 < 1$*, the lag selection criterion that minimizes the asymptotic mean squared error of the Bartlett kernel long-run correlation estimator is*

$$k = \left\lceil 1.4422 \left[\left(\frac{\psi}{1-\lambda^2} \right)^2 T \right]^{\frac{1}{3}} \right\rceil \tag{3.11}$$

where $\lceil \cdot \rceil$ *is the integer ceiling function.*

Notice that Proposition 1 is not a particular case of the Newey and West (1994) automatic lag selection criterion. The Newey-West procedure is based on the minimization of the MSE of a weighted function of the estimated spectral matrix $\hat{\mathbf{S}}(k)$:

$$\min E\left[\mathbf{w}'\left(\hat{\mathbf{S}}(k) - \mathbf{S}\right)\mathbf{w}\right]^2,$$

where **w** is a weight vector. Proposition 1, on the other hand, is based on the minimization of the MSE of the long-run correlation function:

$$\min E\left[\hat{\lambda}_k - \lambda\right]^2.$$

These two MSE cannot be rendered equivalent, since there is no choice of **w** that can simultaneously solve $\mathbf{w}\hat{\mathbf{S}}(k)\mathbf{w} = \hat{\lambda}_k$ and $\mathbf{wSw} = \lambda$. The Newey-West lag selection procedure therefore cannot be used to optimally estimate long-run correlation, despite the fact that it can be used to optimally estimate long-run covariance matrices. This result will be confirmed in section 4 through a Monte Carlo simulation using, among other procedures, a standard Newey-West estimator with a weight vector of ones.

3.4 Alignment

In some cases, a time shift of the kernel (alignment) may improve the finite-sample properties of the cross-spectrum estimator.[12] For example, suppose that the practitioner wants to study the relationship between money and inflation using monthly data. As an educated guess, he believes that money leads inflation by 18 months. In this case, the guessed optimal

[12] See Priestley (1981, pg. 710).

alignment parameter would be equal to 18, and a search range for the parameter should include this guess. Notice that not using an optimal alignment criterion is not different from assuming a search range that goes from zero to zero.

Define therefore the alignment parameter a as follows.

Definition 1: *the alignment parameter minimizes the following weighted sum of absolute covariances:*

$$a = \arg\min \sum_{n=a_{\min}}^{a_{\max}} |n-a| \cdot |\gamma_{xy}(n)|,\ |a_{\min}| < \infty,\ |a_{\max}| < \infty,\ a_{\min} \le a_{\max}. \quad (3.12)$$

The parameters a, $a_{\min}$ and $a_{\max}$ are integer numbers. The alignment parameter is used to relocate the cross-spectrum kernel such that the highest kernel weights are applied to the cross-covariances with the highest absolute values. This technique does not change the asymptotic properties of the estimator, but has the potential to improve finite-sample performance, as Monte Carlo simulations will prove in section 4.

3.5 Lag Selection and Alignment in Practice: The Newey-West Approach

The alignment and the lag selection criteria depend on prior knowledge of spectral parameters. In practice, however, the parameters are not known. As in Newey and West (1994), one solution is to estimate the parameters using a truncated kernel estimator and a thumb rule for bandwidth selection. The parameter estimates are then plugged into equations (3.11) and (3.12). Notice that the data can be prewhitened before applying the procedure, and the results can be recolored afterwards.[13]

---

[13] The use of prewhitening may improve the estimator performance in some cases – see for example Andrews and Monahan (1992) and Newey and West (1994) for details and implementation. It precludes however the direct application of the correlation block estimator to the data, in which case the procedure may lose attractiveness to practitioners that wish to use the time interval choice as part of the analysis of the data. Moreover, inference may become less straightforward.

The truncated kernel cannot be used here however, because it may generate negative spectral estimates. Consider therefore the following analog Bartlett kernel estimators of (3.9), (3.10) and (2.2):

$$\hat{s}_{xy}^{(1)}(m,a)=\frac{1}{2\pi}\sum_{n=-m}^{m}\kappa(n,m)|n|\hat{\gamma}_{xy}(a+n), \qquad \hat{s}_{xx}^{(1)}(m)=\frac{1}{2\pi}\sum_{n=-m}^{m}\kappa(n,m)|n|\hat{\gamma}_{xx}(n),$$

$$\hat{s}_{yy}^{(1)}(m)=\frac{1}{2\pi}\sum_{n=-m}^{m}\kappa(n,m)|n|\hat{\gamma}_{yy}(n), \qquad \hat{s}_{xx}(m)=\frac{1}{2\pi}\sum_{n=-m}^{m}\kappa(n,m)\hat{\gamma}_{xx}(n),$$

$$\hat{s}_{yy}(m)=\frac{1}{2\pi}\sum_{n=-m}^{m}\kappa(n,m)\hat{\gamma}_{yy}(n), \quad \text{with} \quad m=\left\lceil \zeta(T/100)^{1/5}\right\rceil,$$

where, following Hirukawa (2005), and as shown in Appendix 5, the thumb rule for the plug-in bandwidth m has an optimal growth rate of $O(T^{1/5})$, and where ζ is a positive constant. Monte Carlo experiments presented in previous articles, such as Newey and West (1994), typically suggest values of ζ ranging from 3 to 12.

The alignment parameter a is found using an analog representation of (3.12),

$$a=\underset{a_{\min}\le a\le a_{\max}}{\arg\min}\sum_{n=a_{\min}}^{a_{\max}}|n-a|\cdot\left|\hat{\gamma}_{xy}(n)\right|, \tag{3.13}$$

where the working assumption is that all covariances outside the search interval ($a_{\min}$ to $a_{\max}$) are zero.

The following parameters are then calculated conditionally on m and a:

$$\hat{\lambda}_{m,a}=\frac{\hat{\sigma}_{XY}(m,a)}{\sqrt{\hat{\sigma}_{XX}(m)\,\hat{\sigma}_{YY}(m)}},$$

$$\hat{\sigma}_{XY}(m,a)=\begin{cases}\displaystyle\sum_{t=m+a}^{T}\frac{\left[(1-L^m)X_t-m\bar{\mu}_x\right]\left[(1-L^m)Y_{t-a}-m\bar{\mu}_y\right]}{T-m-a}, & a\ge 0,\\[2ex] \displaystyle\sum_{t=m-a}^{T}\frac{\left[(1-L^m)X_{t+a}-m\bar{\mu}_x\right]\left[(1-L^m)Y_t-m\bar{\mu}_y\right]}{T-m+a} & a<0,\end{cases},$$

$$\text{and } \hat{\psi}_{m,a}=\frac{\hat{s}_{xy}^{(1)}(m,a)}{\sqrt{\hat{s}_{xx}(m)\hat{s}_{yy}(m)}}-\frac{\hat{\lambda}_{m,a}}{2}\left(\frac{\hat{s}_{xx}^{(1)}(m)}{\hat{s}_{xx}(m)}+\frac{\hat{s}_{yy}^{(1)}(m)}{\hat{s}_{yy}(m)}\right),$$

which, when plugged in equation (3.11), lead to

$$k = \left[1.4422 \left[\left(\frac{\hat{\psi}_{m,a}}{1 - \hat{\lambda}^2_{m,a}} \right)^2 T \right]^{\frac{1}{3}} \right],$$

and finally to[14]

$$\hat{\lambda}_{k,a} = \frac{\hat{\sigma}_{XY}(k,a)}{\sqrt{\hat{\sigma}_{XX}(k)\,\hat{\sigma}_{YY}(k)}}. \tag{3.14}$$

3.6 Lag Selection and Alignment in Practice: The Andrews-Monahan approach

Andrews (1991) shows how parametric estimates can be used to obtain the first-step spectral parameters, while Andrews and Monahan (1992) explain how the procedure can be improved by using prewhitening. Consider for example a generalization of Andrews and Monahan (1992) Monte Carlo procedure, where here a VAR(*p*) is used instead of a VAR(1) for prewhitening, and a VAR(1) is used instead of an AR(1) for the first-step spectral parameter estimation. Prewhitening employs a VAR(*p*) estimated using LS:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \tilde{\mathbf{B}}_0 + \sum_{n=1}^{p} \tilde{\mathbf{B}}_n \begin{bmatrix} x_{t-n} \\ y_{t-n} \end{bmatrix} + \begin{bmatrix} \tilde{x}_t^w \\ \tilde{y}_t^w \end{bmatrix},$$

where $\tilde{x}_t^w$ and $\tilde{y}_t^w$ are the prewhitened series. A VAR(1),

$$\begin{bmatrix} \tilde{x}_t^w \\ \tilde{y}_t^w \end{bmatrix} = \tilde{\mathbf{C}}_0 + \tilde{\mathbf{C}}_1 \begin{bmatrix} \tilde{x}_{t-1}^w \\ \tilde{y}_{t-1}^w \end{bmatrix} + \begin{bmatrix} \tilde{\varepsilon}_t \\ \tilde{\mu}_t \end{bmatrix},$$

with an estimated innovation covariance matrix $\tilde{\mathbf{\Sigma}}$, can now be used to estimate the first-step parameters, leading to the following spectral matrix estimates:

$$\tilde{\mathbf{S}} = \begin{bmatrix} \tilde{s}_{xx} & \tilde{s}_{xy} \\ \tilde{s}_{xy} & \tilde{s}_{yy} \end{bmatrix} = \frac{1}{2\pi} \left(\mathbf{I} - \tilde{\mathbf{C}}_1 \quad \right)^{-1} \tilde{\mathbf{\Sigma}} \left(\mathbf{I} - \tilde{\mathbf{C}}_1{}' \right)^{-1},$$

[14] Code that calculates this automatic lag selection and alignment procedure is available at the author's home page (http://www.pedrohalbuquerque.net).

and smoothness matrix estimates:

$$\widetilde{\mathbf{S}}^{(1)} = \begin{bmatrix} \widetilde{s}_{xx}^{(1)} & \widetilde{s}_{xy}^{(1)} \\ \widetilde{s}_{xy}^{(1)} & \widetilde{s}_{yy}^{(1)} \end{bmatrix} = \frac{1}{2\pi}\left(\widetilde{\mathbf{H}} + \widetilde{\mathbf{H}}'\right), \qquad \widetilde{\mathbf{H}} = \left(\mathbf{I} - \widetilde{\mathbf{C}}_1\right)^{-2} \widetilde{\mathbf{C}}_1 \sum_{n=0}^{\infty} \widetilde{\mathbf{C}}_1^n \widetilde{\Sigma}\left(\widetilde{\mathbf{C}}_1{}'\right)^n ,$$

where, in practice, the summation is truncated when j reaches a value that is deemed large enough. The optimal lag based on the Andrews-Monahan approach is therefore given by

$$k = \left\lceil 1.4422\left[\left(\frac{\widetilde{\psi}}{1-\widetilde{\lambda}^2}\right)^2 T\right]^{\frac{1}{3}} \right\rceil ,$$

where

$$\widetilde{\lambda} = \frac{\widetilde{s}_{xy}}{\sqrt{\widetilde{s}_{xx}\widetilde{s}_{yy}}} \quad \text{and} \quad \widetilde{\psi} = \frac{\widetilde{s}_{xy}^{(1)}}{\sqrt{\widetilde{s}_{xx}\widetilde{s}_{yy}}} - \frac{\widetilde{\lambda}}{2}\left(\frac{\widetilde{s}_{xx}^{(1)}}{\widetilde{s}_{xx}} + \frac{\widetilde{s}_{yy}^{(1)}}{\widetilde{s}_{yy}}\right).$$

The long-run covariance matrix estimate can now be recolored:

$$\begin{bmatrix} \widetilde{\sigma}_{XX}(k) & \widetilde{\sigma}_{XY}(k) \\ \widetilde{\sigma}_{XY}(k) & \widetilde{\sigma}_{YY}(k) \end{bmatrix} = \widetilde{\mathbf{Q}} \cdot \begin{bmatrix} \widetilde{\sigma}_{X^wX^w}(k) & \widetilde{\sigma}_{X^wY^w}(k) \\ \widetilde{\sigma}_{X^wY^w}(k) & \widetilde{\sigma}_{Y^wY^w}(k) \end{bmatrix} \cdot \widetilde{\mathbf{Q}}',$$

where

$$\widetilde{\sigma}_{X^wY^w}(k) = \sum_{t=k}^{T} \frac{\left(1-L^k\right)\widetilde{X}_t^w \left(1-L^k\right)\widetilde{Y}_t^w}{T-k}, \qquad \widetilde{\mathbf{Q}} = \left(\mathbf{I} - \sum_{n=1}^{p} \widetilde{\mathbf{B}}_n\right)^{-1},$$

and where $\widetilde{X}_t^w$ and $\widetilde{Y}_t^w$ are the integrated versions of $\widetilde{x}_t^w$ and $\widetilde{y}_t^w$, leading to:

$$\widetilde{\lambda}_k = \frac{\widetilde{\sigma}_{XY}(k)}{\sqrt{\widetilde{\sigma}_{XX}(k)\widetilde{\sigma}_{YY}(k)}}. \tag{3.15}$$

4 Monte Carlo Simulations

Appendix 6 shows the results of Monte Carlo simulations with 10,000 iterations each, based on parameters estimated from Dow Jones Industrial Average returns data. Table 1 presents combinations of sample sizes T equal to 100, 400 and 1600 (column 2) and long-run correlation values λ equal to 0.0, 0.4 and 0.8 (column 3).

4.1 Experiment

Series x_t and y_t are generated using VMA(5) and GARCH(1,1) processes with independent and identically t-distributed innovations ν_t and ξ_t:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 1 & \alpha\left(1-\theta\sum_{n=1}^{5}\frac{L^n}{5}\right) \\ \alpha\left(1-\theta\sum_{n=1}^{5}\frac{L^n}{5}\right)L^3 & L^3 \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \mu_t \end{bmatrix},$$

$$\varepsilon_t = \delta_0 + \nu_t\sqrt{g_t}\,, \qquad g_t = \beta_0 + \beta_1(\varepsilon_{t-1}-\delta_0)^2 + \beta_2 g_{t-1}\,, \qquad \mathrm{var}(\nu_t) = 1\,,$$

$$\mu_t = \delta_0 + \xi_t\sqrt{h_t}\,, \qquad h_t = \beta_0 + \beta_1(\mu_{t-1}-\delta_0)^2 + \beta_2 h_{t-1}\,, \qquad \mathrm{var}(\xi_t) = 1\,,$$

where parameters δ_0, β_0, β_1, and β_2 and the number of degrees of freedom (DOF) of the t-distributed innovations were estimated from Dow Jones returns data ranging from January 9, 1990, to August 1, 2001, corresponding to 3018 daily observations. The estimated parameters are:

$$\delta_0 = \underset{(0.000133)}{0.000648}\,, \quad \beta_0 = \underset{(2.26\mathrm{E}-07)}{6.42\mathrm{E}-07}\,, \quad \beta_1 = \underset{(0.007509)}{0.050154}\,,$$

$$\beta_2 = \underset{(0.008042)}{0.944037}\,, \quad \text{and} \quad \mathrm{DOF} = \underset{(0.564310)}{5.605809}\,,$$

where the values between parentheses represent standard errors.

The values of α and θ determine the value of λ according to the equation

$$\lambda = \frac{2\alpha(1-\theta)}{1+\alpha^2(1-\theta)^2}\,.$$

The experiment creates pairs of series that emulate the Dow Jones statistical process, allowing however for different levels of long-run correlation. A lag of three periods is applied to series y_t in order to test the effectiveness of the alignment criterion. Parameter θ assumes values of 0.0, 0.5 and 0.8, as shown in column 1, and is used to evaluate the sensitivity of the estimation procedures to the presence of small moving average terms at longer lags. This type of process is commonly found in economic and financial data and tends to pose problems to some estimators, as discussed for example in Cochrane (1988), Schwert (1989) and Newey and West (1994).

4.2 Benchmarks

Four common estimation procedures are taken as benchmarks that represent current and common practices and are presented in columns 4 to 9 of Table 1:

a) aggregation of daily data over time intervals of 5 and 20 days, roughly representing correlation estimates based on weekly and monthly aggregates (columns 4 and 5, "5 Days" and "20 Days," respectively);

b) VAR estimation with order selection based on the Akaike information criterion (AIC) and on the Schwarz Bayesian criterion (SBC), with long-run correlation estimates calculated using the spectral matrix procedure described on page 836 of Andrews (1991) (columns 6 and 7, "AIC" and "SBC," respectively);

c) block estimator using the Schwert (1989) lag selection thumb rule $k = \lceil 4(T/100)^{1/4} \rceil$ (column 8, "k_{TR}"); and

d) block estimator of the covariance matrix using the Newey-West automatic lag selection, with standard weight vector of ones and without prewhitening (column 9, "k_{NW}").

These benchmarks are compared to the following new estimators proposed in this article:

a) the block estimator with automatic lag selection criterion and without alignment $(a = 0)$ based on the Newey-West approach, given by equation (3.14) (columns 10, 11 and 12, "k_2", "k_4" and "k_{12}," respectively);

b) the block estimator with automatic lag selection and alignment criteria based on the Newey-West approach, given by equations (3.13) and (3.14) (columns 13, 14 and 15, "$k_{2,a}$", "$k_{4,a}$" and "$k_{12,a}$," respectively); and

c) the block estimator with prewhitening and automatic lag selection criterion using the Andrews-Monahan approach, given by equation (3.15) (columns 16, 17 and 18, "$k_{A,1}$", "$k_{A,AIC}$" and "$k_{A,SBC}$," respectively).

The estimators based on the Newey-West approach use the following first-step thumb rule:

$$m = \left\lceil \zeta (T/100)^{1/5} \right\rceil ,$$

with values of ζ equal to 2, 4 and 12. The prewhitening step of the estimators based on the Andrews-Monahan approach uses VAR orders equal to one or selected according to the Akaike information criterion (AIC) and the Schwarz Bayesian criterion (SBC).

4.3 Results

Table 1 depicts the MSE values for each estimator and different combinations of θ, λ and T, where

$$\text{MSE} = \left(\bar{\hat{\lambda}} - \lambda\right)^2 + \hat{\sigma}_\lambda^2 ,\ \bar{\hat{\lambda}} - \lambda = \frac{1}{10000}\sum_{n=1}^{10000} \hat{\lambda}_n - \lambda \text{ and } \hat{\sigma}_\lambda^2 = \frac{1}{10000}\sum_{n=1}^{10000}\left(\hat{\lambda}_n - \bar{\hat{\lambda}}\right)^2 .$$

The results in columns 4 and 5 in Table 1 ("5 Days" and "20 Days") indicate that the common practice of aggregating high-frequency data into longer time intervals may lead to poor MSE statistics. As expected, estimators based on aggregate weekly data (column 4) perform poorly for high values of λ. Estimators based on aggregate monthly data (column 5), on the other hand, perform poorly for small samples, due to the wasteful use of the available information. The simulations indicate that this practice should be avoided.

Column 6 ("AIC") shows that the parametric VAR estimator that uses the Akaike information criterion for order selection performs poorly for small sample sizes. Column 7 ("SBC"), in contrast, reveals that the VAR estimator that use the Schwartz Bayesian criterion for order selection tends to perform better than other benchmarks, but not as well as this article's proposed estimator. For example, for θ=0.8, λ=0.8, and T=400, the VAR-SBC estimator has a MSE of 0.175, which is 3.6 times higher than the MSE of 0.048 of the proposed estimator in column 14, and 5.3 times higher than the MSE of 0.033 of the proposed estimator in column 15.

Column 8 ("k_{TR}") gives the performance of the block estimator using the Schwert (1989) lag selection rule. This thumb rule performs poorly for high values of λ, as expected. Column 9 ("k_{NW}") presents the results for the covariance matrix estimator using the Newey-West (1994) automatic lag selection criterion. As in the case of the Schwert lag selection rule, this estimator performs poorly for high values of λ. The Newey-West criterion should not be used therefore for long-run correlation estimation through the estimation of the covariance matrix, in agreement with subsection 3.3.

Columns 10, 11 and 12 ("k_2", "k_4" and "k_{12}") show the results for the proposed block estimator using the lag selection criterion based on the Newey-West approach and without alignment (a=0), as shown in subsection 3.5. Notice the clear improvement over the two previous benchmark block estimators (columns 8 and 9, "k_{TR}" and "k_{NW}"), particularly when parameter ζ is equal to 12. The VAR-SBC estimator (column 7) however outperforms the proposed estimator without alignment in a majority of cases, even when ζ is equal to 12.

The proposed block estimator with automatic lag selection and alignment criteria based on the Newey-West approach, as described in subsection 3.5, outperforms all other estimators, as shown in column 13, 14 and 15 ("$k_{2,a}$", "$k_{4,a}$" and "$k_{12,a}$"). The MSE values for this estimator are uniformly among the lowest for all combinations of parameters, particular when ζ is equal to 4 and 12. A choice of ζ equal to 4 tends to produce the best results when the values of θ are small, while a choice of ζ equal to 12 tends to be beneficial when the values of θ are large.

Finally, the results for the proposed block estimator based on the Andrews-Monahan approach described in subsection 3.6 are shown in columns 16, 17 and 18 ("$k_{A,1}$", "$k_{A,AIC}$", and "$k_{A,SBC}$"). This estimator performs at par with the VAR-AIC and VAR-SBC, with the exception of the VAR(1) case ("$k_{a,1}$"), which performs poorly. The latter is the result of omitted variable bias due to the misspecification of the model used in the prewhitening step. Notice also that this variant of the estimator performs well when $\lambda=0$ not for its own merits but simply because, in this particular case, there is no misspecification and the model is parsimonious.

The Monte Carlo simulations reveal therefore that the joint use of the lag selection and alignment criteria based on the Newey-West approach is effective when processes contain small moving average terms at longer lags and possible time misalignments, without producing significant drawbacks. The proposed procedure leads to significant improvements over methods currently employed, suggesting that the estimator presented in this article should be considered a useful addition to the practitioner's toolbox.

5 Conclusions

Long-run correlation estimators have many applications in finance and economics, for example, in the study of stock returns and in the measurement of the relations and lags between monetary and real variables.

This paper used the approaches of Andrews and Monahan (1992) and Newey and West (1994) to develop automatic lag selection criteria for a nonparametric consistent long-run correlation estimator based on the block estimator (*k*-lag difference correlation estimator). In addition, an alignment criterion that potentially enhances finite sample performance was presented.

A Monte Carlo experiment showed that the lag selection and the alignment criteria presented here are effective and superior to commonly employed methods, such as aggregation over arbitrary time intervals, parametric VAR estimation, and Newey-West automatic lag selection of the covariance matrix.

The optimal yet unobtrusive long-run correlation estimator presented in this article intends to reduce the gap between econometric theory and practice by offering not only an asymptotically optimal alternative to current practices, but also a formal statistical framework for researchers dealing with time-series correlation studies.

References

Ammer, J. & J. Mei (1996). “Measuring International Economic Linkages with International Stock Market Data.” *The Journal of Finance*, 51, 5, 1743-1763.

Andrews, D. W. K. (1991). “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation.” *Econometrica*, 59, 3, 817-858.

Andrews, D. W. K. & J. C. Monahan (1992). “An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator.” *Econometrica*, 60, 4, 953-966.

Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. New York: John Wiley & Sons.

Bartlett, M. S. (1950). “Periodogram Analysis and Continuous Spectra.” *Biometrika*, 37, 1/2, 1-16.

Brockwell, P. J. & R. A. Davis (1991). *Time Series: Theory and Methods*. New York: Springer.

Campbell, J. Y. & J. Ammer (1993). “What Moves the Stock and Bond Markets? A Variance Decomposition for Long-Term Asset Returns” *The Journal of Finance*, 48, 1, 3-37.

Cochrane, J. H. (1988). “How Big is the Random Walk in GNP?” *Journal of Political Economy*, 96, 893-920.

Cochrane, J. H. & A. M. Sbordone (1988). “Multivariate Estimates of the Permanent Components of GNP and Stock Prices.” *Journal of Economic Dynamics and Control*, 12, 255-296.

Engle, R. F. & C. W. J. Granger (1987). “Co-Integration and Error Correction: Representation, Estimation, and Testing.” *Econometrica*, 55, 251-276.

Fisher, M. E. & J. J. Seater (1993). “Long-Run Neutrality and Superneutrality in an ARIMA Framework.” *American Economic Review*, 83, 402-415.

Forbes, K. & R. Rigobon (2002). “No Contagion, Only Interdependence: Measuring Stock Market Comovements, *The Journal of Finance*, 57, 5, 2223-2261.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*. New York: John Wiley & Sons.

Granger, C. W. J. & R. Engle (1983). “Applications of Spectral Analysis in Econometrics.” In D. R. Brillinger and P. R. Krishnaiah (Eds.), *Handbook of Statistics*, Vol. 3. Amsterdam: Elsevier.

Granger, C. W. J. & H. J. B. Rees (1968). “Spectral Analysis of the Term Structure of the Interest Rates.” *Review of Economic Studies*, 35, 67-76.

Granger, C. W. J. & M. W. Watson (1984). “Time Series and Spectral Methods in Econometrics.” In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*, Vol. II. New York: North-Holland.

Granger, C. W. J. & A. A. Weiss (1983). “Time Series Analysis of Error-Correction Models.” In *Studies in Econometrics, Time Series, and Multivariate Statistics*. New York: Academic Press.

Hannan, E. J. (1970). *Multiple Time Series*. New York: John Wiley and Sons.

Hirukawa, M. (2005). *A Two-Stage Plug-In Bandwidth Selection and Its Implementation for Covariance Estimation*. Working Paper 04005, Department of Economics. Montreal: Concordia University.

King, R. G., & M. W. Watson (1994). “The Post-War U.S. Phillips Curve: A Revisionist Econometric History.” In A. H. Meltzer and C. I. Plosser, (Eds.), *Carnegie-Rochester Conference Series on Public Policy*, 41. Amsterdam: North-Holland.

Koopmans, L. H. (1974). *The Spectral Analysis of Time Series*. San Diego: Academic Press.

McCallum, B. T. (1984). “On Low-Frequency Estimates of Long-Run Relationships in Macroeconomics.” *Journal of Monetary Economics*, 14, 3-14.

McCallum, B. T. (1993). “Unit Roots in Macroeconomic Time Series: Some Critical Issues.” *Federal Reserve Bank of Richmond Economic Quarterly*, 79, 13–43.

Newey, W. K. & K. D. West (1987). “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica*, 55, 703-708.

Newey, W. K. & K. D. West (1994). “Automatic Lag Selection in Covariance Matrix Estimation.” *Review of Economic Studies*, 61, 631-653.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*. London, Academic Press.

Schwert, G. W. (1989). "Tests for Unit Roots: A Monte Carlo Investigation." *Journal of Business & Economic Statistics*, 7, 2, 147-159.

Appendix 1

The k-lag difference covariance estimator of x_t and y_t is:

$$\hat{\sigma}_{XY}(k) = \frac{1}{T-k}\sum_{t=k}^{T}\left[\left(1-L^k\right)X_t - k\bar{\mu}_x\right]\left[\left(1-L^k\right)Y_t - k\bar{\mu}_y\right],$$

where $\left(1-L^k\right)X_t = X_t - X_{t-k}$ and $\left(1-L^k\right)Y_t = Y_t - Y_{t-k}$, such that

$$\hat{\sigma}_{XY}(k) = \frac{1}{T-k}\sum_{t=k}^{T}\left[\left(\sum_{i=1}^{k}x'_{t-k+i}\right)\left(\sum_{j=1}^{k}y'_{t-k+j}\right)\right],$$

where $x'_t = x_t - \bar{\mu}_x$ and $y'_t = y_t - \bar{\mu}_y$, or

$$\hat{\sigma}_{XY}(k) = \frac{1}{T-k}\sum_{t=k}^{T}\left[\sum_{l=1}^{k-1}\left(\sum_{i=1}^{l}x'_{t-k+i}y'_{t-l+i} + \sum_{j=1}^{l}x'_{t-l+j}y'_{t-k+j}\right) + \sum_{m=1}^{k}x'_{t-k+m}y'_{t-k+m}\right]$$

$$= \frac{1}{T-k}\left(\sum_{t=k}^{T}x'_{t-k+1}y'_t + 2\sum_{t=k-1}^{T}x'_{t-k+2}y'_t + \ldots + (k-1)\sum_{t=2}^{T}x'_{t-1}y'_t + k\sum_{t=1}^{T}x'_t y'_t \right.$$

$$\left. + (k-1)\sum_{t=2}^{T}x'_t y'_{t-1} + \ldots + 2\sum_{t=k-1}^{T}x'_t y'_{t-k+2} + \sum_{t=k}^{T}x'_t y'_{t-k+1}\right) + R(T,k),$$

where

$$R(T,k) = -\frac{1}{T-k}\left[\sum_{i=2}^{k-1}\sum_{j=1}^{i-1}(i-j)\left(x'_j y'_{k-i+j} + x'_{T-k+i-j+1}y'_{T-j+1}\right.\right.$$

$$\left.\left. + x'_{k-i+j}y'_j + x'_{T-j+1}y'_{T-k+i-j+1}\right) + \sum_{j=1}^{k-1}(k-j)\left(x'_j y'_j + x'_{T-j+1}y'_{T-j+1}\right)\right],$$

or

$$\frac{\hat{\sigma}_{XY}(k)}{k} = \frac{T}{T-k}2\pi\hat{s}_{xy}(k) + \frac{R(T,k)}{k}.$$

Given the assumption of summable autocovariances and covariances, it follows that $R(T,k)/k = O_p(k/T)$, therefore, the block covariance estimator is asymptotically equivalent to 2π times the Bartlett kernel estimator of the cross-spectrum at frequency zero:

$$\frac{\hat{\sigma}_{XY}(k)}{k} = \frac{T}{T-k} 2\pi \hat{s}_{xy}(k) + O_p\left(\frac{k}{T}\right).$$

Appendix 2

Hannan (1970, pg. 280), Priestley (1981, pg. 699) and Brockwell and Davis (1991, pg. 446) show that, under the consistency assumptions of Subsection 3.2, in particular,

$$\lim_{T\to\infty} k(T) = \infty \text{ and } \lim_{T\to\infty} k(T)/T = 0,$$

the asymptotic covariance (Acov) between two spectra or cross-spectra Bartlett kernel estimators at frequency zero is

$$\text{Acov}(\hat{s}_{ab}(k), \hat{s}_{cd}(k)) = \frac{2}{3}\frac{k}{T}(s_{ac}s_{bd} + s_{ad}s_{bc}). \tag{A2.1}$$

The long-run correlation and its estimator are defined as

$$\lambda = \frac{s_{xy}}{\sqrt{s_{xx}\ s_{yy}}}, \tag{A2.2}$$

$$\hat{\lambda}_k = \frac{\hat{s}_{xy}(k)}{\sqrt{\hat{s}_{xx}(k)\,\hat{s}_{yy}(k)}}. \tag{A2.3}$$

According to (A2.1), the components of (A2.3) have an asymptotic covariance matrix given by

$$\text{Acov}(\hat{\mathbf{s}}(k)) = \text{cov}\begin{bmatrix} \hat{s}_{xx}(k) \\ \hat{s}_{yy}(k) \\ \hat{s}_{xy}(k) \end{bmatrix} = \frac{2}{3}\frac{k}{T}\begin{bmatrix} 2s_{xx}^2 & 2s_{xy}^2 & 2s_{xx}s_{xy} \\ 2s_{xy}^2 & 2s_{yy}^2 & 2s_{yy}s_{xy} \\ 2s_{xx}s_{xy} & 2s_{yy}s_{xy} & s_{xx}s_{yy} + s_{xy}^2 \end{bmatrix}. \tag{A2.4}$$

As in Hannan (1970, pg. 287), a Taylor expansion of (A2.2) around $E[\hat{s}_{xx}(k)]$, $E[\hat{s}_{yy}(k)]$, and $E[\hat{s}_{xy}(k)]$ leads to

$$\hat{\lambda}(k) = \frac{E[\hat{s}_{xy}(k)]}{\sqrt{E[\hat{s}_{xx}(k)]E[\hat{s}_{yy}(k)]}} + \mathbf{D}_\lambda \begin{bmatrix} \hat{s}_{xx}(k) - E[\hat{s}_{xx}(k)] \\ \hat{s}_{yy}(k) - E[\hat{s}_{yy}(k)] \\ \hat{s}_{xy}(k) - E[\hat{s}_{xy}(k)] \end{bmatrix} + O_p\left(\left(\frac{k}{T}\right)^{\frac{1}{2}}\right), \quad \text{(A2.5)}$$

where

$$\mathbf{D}_\lambda = \begin{bmatrix} \frac{\partial\lambda}{\partial s_{xx}} & \frac{\partial\lambda}{\partial s_{yy}} & \frac{\partial\lambda}{\partial s_{xy}} \end{bmatrix} = \begin{bmatrix} \frac{-\lambda}{2s_{xx}} & \frac{-\lambda}{2s_{yy}} & \frac{1}{\sqrt{s_{xx}s_{yy}}} \end{bmatrix}.$$

From (A2.5), and since

$$E[\hat{\lambda}_k] - E[\hat{s}_{xy}(k)] \Big/ \sqrt{E[\hat{s}_{xx}(k)]E[\hat{s}_{yy}(k)]} = O(k/T),$$

one can conclude that the asymptotic variance of the long-run correlation estimator is given by

$$\text{Avar}(\hat{\lambda}_k) = \mathbf{D}_\lambda \cdot \text{Acov}(\hat{\mathbf{s}}(k)) \cdot \mathbf{D}'_\lambda,$$

such that

$$\text{Avar}(\hat{\lambda}_k) = \frac{2}{3}\frac{k}{T}(1-\lambda^2)^2. \quad \text{(A2.6)}$$

Appendix 3

The long-run correlation asymptotic bias (Abias) is given by

$$\text{Abias}(\hat{\lambda}_k) = E[\hat{\lambda}_k - \lambda] = E\left[\frac{\hat{s}_{xy}(k)}{\sqrt{\hat{s}_{xx}(k)\hat{s}_{yy}(k)}} - \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}}\right].$$

Following Hannan (1970, pg. 283), and under the same consistency assumptions of Appendix 2, the asymptotic bias vector is:

$$\text{Abias}(\hat{\mathbf{s}}) = \text{bias}\begin{bmatrix} \hat{s}_{xx}(k) & \hat{s}_{yy}(k) & \hat{s}_{xy}(k) \end{bmatrix}' = \frac{-1}{k}\begin{bmatrix} s_{xx}^{(1)} & s_{yy}^{(1)} & s_{xy}^{(1)} \end{bmatrix}',$$

where $s_{xx}^{(1)} = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}|n|\gamma_{xx}(n)$, $s_{yy}^{(1)} = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}|n|\gamma_{yy}(n)$, and $s_{xy}^{(1)} = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}|n|\gamma_{xy}(n)$.

A Taylor expansion of (A2.2) around s_{xx}, s_{yy}, and s_{xy}, with $T \to \infty$, leads to

$$E\left[\hat{\lambda}(k)-\lambda\right]=\mathbf{D}_{\lambda}\begin{bmatrix}E\left[\hat{s}_{xx}(k)-s_{xx}\right]\\ E\left[\hat{s}_{yy}(k)-s_{yy}\right]\\ E\left[\hat{s}_{xy}(k)-s_{xy}\right]\end{bmatrix}+O\left(\frac{1}{k^2}\right).$$

Therefore, the asymptotic bias of the long-run correlation estimator is

$$\text{Abias}\left(\hat{\lambda}_k\right)=\mathbf{D}_{\lambda}\cdot\text{Abias}\left[\hat{\mathbf{s}}(k)\right]\ ,$$

where vector $\mathbf{D}_{\lambda}$ is defined as in Appendix 2, implying that

$$\text{Abias}\left(\hat{\lambda}_k\right)=-\frac{\psi}{k},\quad \psi=\frac{s_{xy}^{(1)}}{\sqrt{s_{xx}s_{yy}}}-\frac{\lambda}{2}\left(\frac{s_{xx}^{(1)}}{s_{xx}}+\frac{s_{yy}^{(1)}}{s_{yy}}\right). \qquad \text{(A3.1)}$$

Appendix 4

Problem: to find the optimal lag selection k that minimizes the asymptotic mean square error (AMSE) of the long-run correlation estimator,

$$\min_k \text{AMSE}\left(\hat{\lambda}_k\right)=\left[\text{Abias}\left(\hat{\lambda}_k\right)\right]^2+\text{Avar}\left(\hat{\lambda}_k\right). \qquad \text{(A4.1)}$$

From (A4.1), (A2.6) and (A3.1), and under the same consistency assumptions of Appendix 2, it is straightforward to see that the AMSE is given by

$$\text{AMSE}\left(\hat{\lambda}_k\right)=\frac{1}{k^2}\psi^2+\frac{k}{T}\frac{2}{3}\left(1-\lambda^2\right)^2, \qquad \text{(A4.2)}$$

and consequently the optimal k that minimizes the AMSE is

$$k=\left\lceil 1.4422\left[\left(\frac{\psi}{1-\lambda^2}\right)^2 T\right]^{\frac{1}{3}}\right\rceil,$$

where $\lceil\cdot\rceil$ represents the integer ceiling function.

Appendix 5

Problem:

$$\min_m \text{AMSE}\left(\hat{\eta}_m\right)=\left[\text{Abias}\left(\hat{\eta}_m\right)\right]^2+\text{Avar}\left(\hat{\eta}_m\right), \qquad \text{(A5.1)}$$

where

$$\hat{\eta}_m = \frac{\hat{\psi}_m}{1-\hat{\lambda}_m^2}, \text{ and } \eta = \frac{\psi}{1-\lambda^2}. \tag{A5.2}$$

Following Hirukawa (2005) and Priestley (1981, pg. 325), the asymptotic covariance matrix of the spectral components of $\hat{\eta}_m$ is given by

$$\text{Acov}(\hat{\mathbf{s}}(m)) = \frac{4m^3}{3T}\begin{bmatrix} \frac{s_{xx}^2}{m^2} & \frac{s_{xy}^2}{m^2} & \frac{s_{xx}s_{xy}}{m^2} & \frac{s_{xx}^2}{4m} & \frac{s_{xy}^2}{4m} & \frac{s_{xx}s_{xy}}{4m} \\ \frac{s_{xy}^2}{m^2} & \frac{s_{yy}^2}{m^2} & \frac{s_{yy}s_{xy}}{m^2} & \frac{s_{xy}^2}{4m} & \frac{s_{yy}^2}{4m} & \frac{s_{yy}s_{xy}}{4m} \\ \frac{s_{xx}s_{xy}}{m^2} & \frac{s_{yy}s_{xy}}{m^2} & \frac{s_{xx}s_{yy}+s_{xy}^2}{2m^2} & \frac{s_{xx}s_{xy}}{4m} & \frac{s_{yy}s_{xy}}{4m} & \frac{s_{xx}s_{yy}+s_{xy}^2}{8m} \\ \frac{s_{xx}^2}{4m} & \frac{s_{xy}^2}{4m} & \frac{s_{xx}s_{xy}}{4m} & \frac{s_{xx}^2}{10} & \frac{s_{xy}^2}{10} & \frac{s_{xx}s_{xy}}{10} \\ \frac{s_{xy}^2}{4m} & \frac{s_{yy}^2}{4m} & \frac{s_{yy}s_{xy}}{4m} & \frac{s_{xy}^2}{10} & \frac{s_{yy}^2}{10} & \frac{s_{yy}s_{xy}}{10} \\ \frac{s_{xx}s_{xy}}{4m} & \frac{s_{yy}s_{xy}}{4m} & \frac{s_{xx}s_{yy}+s_{xy}^2}{8m} & \frac{s_{xx}s_{xy}}{10} & \frac{s_{yy}s_{xy}}{10} & \frac{s_{xx}s_{yy}+s_{xy}^2}{20} \end{bmatrix},$$

where

$$\hat{\mathbf{s}}(m) = \left[\hat{s}_{xx}(m) \;\; \hat{s}_{yy}(m) \;\; \hat{s}_{xy}(m) \;\; \hat{s}_{xx}^{(1)}(m) \;\; \hat{s}_{yy}^{(1)}(m) \;\; \hat{s}_{xy}^{(1)}(m)\right]',$$

and the asymptotic bias vector by

$$\text{Abias}(\hat{\mathbf{s}}(m)) = \frac{-1}{m}\left[s_{xx}^{(1)} \;\; s_{yy}^{(1)} \;\; s_{xy}^{(1)} \;\; s_{xx}^{(1)2} \;\; s_{yy}^{(1)2} \;\; s_{xy}^{(1)2}\right]'.$$

As in Appendix 2, a Taylor expansion of (A5.2) around $E[\hat{s}_{xx}(m)]$, $E[\hat{s}_{yy}(m)]$, $E[\hat{s}_{xy}(m)]$, $E[\hat{s}_{xx}^{(1)}(m)]$, $E[\hat{s}_{yy}^{(1)}(m)]$, and $E[\hat{s}_{xy}^{(1)}(m)]$ leads to

$$\hat{\eta}_m = \hat{\eta}_m^E + \mathbf{D}_\eta \begin{bmatrix} \hat{s}_{xx}(m) - E[\hat{s}_{xx}(m)] \\ \hat{s}_{yy}(m) - E[\hat{s}_{yy}(m)] \\ \hat{s}_{xy}(m) - E[\hat{s}_{xy}(m)] \\ \hat{s}_{xx}^{(1)}(m) - E[\hat{s}_{xx}^{(1)}(m)] \\ \hat{s}_{yy}^{(1)}(m) - E[\hat{s}_{yy}^{(1)}(m)] \\ \hat{s}_{xy}^{(1)}(m) - E[\hat{s}_{xy}^{(1)}(m)] \end{bmatrix} + O_p\left(\left(\frac{m^3}{T}\right)^{\frac{1}{2}}\right), \tag{A5.3}$$

where

$$\mathbf{D}_\eta = \left[\frac{\partial \eta}{\partial s_{xx}} \quad \frac{\partial \eta}{\partial s_{yy}} \quad \frac{\partial \eta}{\partial s_{xy}} \quad \frac{\partial \eta}{\partial s_{xx}^{(1)}} \quad \frac{\partial \eta}{\partial s_{xy}^{(1)}} \quad \frac{\partial \eta}{\partial s_{xy}^{(1)}} \right],$$

$$\hat{\eta}_m^E = \frac{1}{1-\hat{\lambda}_m^E} \left[\frac{E\left[s_{xy}^{(1)}(m)\right]}{\sqrt{E\left[s_{xx}(m)\right]E\left[s_{yy}(m)\right]}} - \frac{\hat{\lambda}_m^E}{2} \left(\frac{E\left[s_{xx}^{(1)}(m)\right]}{E\left[s_{xx}(m)\right]} + \frac{E\left[s_{yy}^{(1)}(m)\right]}{E\left[s_{yy}(m)\right]} \right) \right],$$

and

$$\hat{\lambda}_m^E = \frac{E\left[\hat{s}_{xy}(m)\right]}{\sqrt{E\left[\hat{s}_{xx}(m)\right]E\left[\hat{s}_{yy}(m)\right]}}.$$

From (A5.3), and since

$$E\left[\hat{\eta}_m\right] - \hat{\eta}_m^E = O\left(m^3/T\right),$$

one can conclude that the asymptotic variance of $\hat{\eta}_m$ is given by

$$\text{Avar}\left(\hat{\eta}_m\right) = \mathbf{D}_\eta \cdot \text{Acov}\left(\hat{\mathbf{s}}(m)\right) \cdot \mathbf{D}'_\eta,$$

and in this case, as $m \to \infty$ and $m/T \to 0$, the problem of calculating the asymptotic variance of $\hat{\eta}_m$ reduces to

$$\text{Avar}\left(\hat{\eta}_m\right) = \frac{4}{30}\frac{m^3}{T} \cdot \mathbf{d}_\eta \cdot \begin{bmatrix} s_{xx}^2 & s_{xy}^2 & s_{xx}s_{xy} \\ s_{xy}^2 & s_{yy}^2 & s_{yy}s_{xy} \\ s_{xx}s_{xy} & s_{yy}s_{xy} & \left(s_{xx}s_{yy} + s_{xy}^2\right)/2 \end{bmatrix} \cdot \mathbf{d}'_\eta,$$

where

$$\mathbf{d}_\eta = \left[\frac{\partial \eta}{\partial s_{xx}^{(1)}} \quad \frac{\partial \eta}{\partial s_{xy}^{(1)}} \quad \frac{\partial \eta}{\partial s_{xy}^{(1)}} \right] = \frac{1}{1-\lambda^2} \left[\frac{\lambda}{2s_{xx}} \quad \frac{\lambda}{2s_{yy}} \quad \frac{1}{\sqrt{s_{xx}s_{yy}}} \right],$$

such that

$$\text{Avar}\left(\hat{\eta}_m\right) = \frac{1}{15}\frac{m^3}{T}. \tag{A5.4}$$

Now, for the asymptotic bias, a Taylor expansion of (A5.2) around s_{xx}, s_{yy}, s_{xy}, $s_{xx}^{(1)}$, $s_{yy}^{(1)}$, and $s_{xy}^{(1)}$ leads to

$$E\left[\hat{\eta}_m - \eta\right] = \mathbf{D}_\eta \begin{bmatrix} E\left[\hat{s}_{xx}(m) - s_{xx}\right] \\ E\left[\hat{s}_{yy}(m) - s_{yy}\right] \\ E\left[\hat{s}_{xy}(m) - s_{xy}\right] \\ E\left[\hat{s}_{xx}^{(1)}(m) - s_{xx}^{(1)}\right] \\ E\left[\hat{s}_{yy}^{(1)}(m) - s_{yy}^{(1)}\right] \\ E\left[\hat{s}_{xy}^{(1)}(m) - s_{xy}^{(1)}\right] \end{bmatrix} + O\left(\frac{1}{m^2}\right),$$

and therefore the asymptotic bias of $\hat{\eta}_m$ is

$$\text{Abias}(\hat{\eta}_m) = \mathbf{D}_\eta \cdot \text{Abias}(\hat{\mathbf{s}}(m)),$$

and the equation for the asymptotic bias can be summarized as:

$$\text{Abias}(\hat{\eta}_m) = \alpha/m, \tag{A5.5}$$

where α represents a trivial yet lengthy combination of parameters.

Given equations (A5.4) and (A5.5), the solution to problem (A5.1) is

$$m = \left\lceil 10\alpha^2 T \right\rceil^{1/5}.$$

Appendix 6

Table 1 - Monte Carlo - MSE of $\hat{\lambda}$ [(a)]

θ	λ	T	Aggregate[(b)] 5 Days	Aggregate[(b)] 20 Days	VAR(p)[(c)] AIC	VAR(p)[(c)] SBC	k_{TR}[(d)]	k_{NW}[(e)]	k_2[(f)]	k_4[(g)]	k_{12}[(h)]	$k_{2,a}$[(i)]	$k_{4,a}$[(j)]	$k_{12,a}$[(k)]	$k_{A,1}$[(l)]	$k_{A,AIC}$[(m)]	$k_{A,SBC}$[(n)]
0.0	0.0	100	0.053	0.253	0.164	0.031	0.028	0.022	0.016	0.027	0.074	0.026	0.037	0.082	0.031	0.164	0.031
0.0	0.0	400	0.013	0.052	0.019	0.008	0.010	0.007	0.005	0.009	0.022	0.008	0.011	0.024	0.008	0.019	0.008
0.0	0.0	1600	0.003	0.013	0.006	0.002	0.003	0.002	0.002	0.002	0.007	0.002	0.003	0.007	0.002	0.006	0.002
0.0	0.4	100	0.107	0.214	0.185	0.155	0.121	0.165	0.181	0.129	0.093	0.086	0.064	0.076	0.198	0.186	0.156
0.0	0.4	400	0.067	0.046	0.024	0.017	0.046	0.074	0.166	0.034	0.037	0.010	0.011	0.020	0.171	0.024	0.017
0.0	0.4	1600	0.058	0.013	0.007	0.004	0.023	0.023	0.024	0.011	0.014	0.002	0.003	0.006	0.163	0.007	0.004
0.0	0.8	100	0.294	0.147	0.074	0.013	0.415	0.532	0.678	0.294	0.088	0.023	0.010	0.016	0.708	0.074	0.013
0.0	0.8	400	0.242	0.032	0.005	0.003	0.166	0.117	0.642	0.038	0.021	0.002	0.002	0.004	0.657	0.005	0.003
0.0	0.8	1600	0.231	0.017	0.002	0.001	0.089	0.033	0.035	0.011	0.006	0.001	0.001	0.001	0.644	0.002	0.001
Mean MSE			0.119	0.087	0.054	0.026	0.100	0.109	0.194	0.062	0.040	0.018	0.016	0.026	0.287	0.054	0.026
0.5	0.4	100	0.123	0.222	0.203	0.074	0.157	0.223	0.250	0.159	0.096	0.109	0.083	0.081	0.303	0.203	0.073
0.5	0.4	400	0.078	0.044	0.044	0.057	0.047	0.068	0.207	0.025	0.036	0.071	0.037	0.035	0.271	0.044	0.054
0.5	0.4	1600	0.068	0.013	0.014	0.044	0.021	0.020	0.016	0.009	0.013	0.027	0.014	0.015	0.263	0.014	0.041
0.5	0.8	100	0.410	0.200	0.134	0.039	0.573	0.747	0.853	0.412	0.118	0.023	0.015	0.015	0.959	0.135	0.043
0.5	0.8	400	0.356	0.056	0.019	0.014	0.243	0.272	0.628	0.057	0.030	0.009	0.006	0.007	0.921	0.019	0.015
0.5	0.8	1600	0.337	0.033	0.005	0.005	0.131	0.132	0.063	0.016	0.009	0.004	0.003	0.003	0.908	0.005	0.005
Mean MSE			0.229	0.095	0.070	0.039	0.195	0.244	0.336	0.113	0.051	0.040	0.026	0.026	0.604	0.070	0.039
0.8	0.4	100	0.174	0.241	0.291	0.083	0.230	0.319	0.362	0.218	0.098	0.198	0.133	0.098	0.443	0.292	0.087
0.8	0.4	400	0.126	0.059	0.067	0.034	0.077	0.206	0.158	0.036	0.037	0.099	0.055	0.045	0.414	0.067	0.035
0.8	0.4	1600	0.110	0.021	0.021	0.018	0.034	0.188	0.030	0.013	0.013	0.040	0.024	0.019	0.409	0.021	0.018
0.8	0.8	100	0.611	0.450	0.190	0.214	0.718	0.842	0.861	0.701	0.352	0.196	0.162	0.112	0.963	0.189	0.203
0.8	0.8	400	0.560	0.226	0.084	0.175	0.459	0.517	0.719	0.294	0.169	0.048	0.048	0.033	0.937	0.084	0.165
0.8	0.8	1600	0.544	0.180	0.026	0.141	0.344	0.198	0.345	0.152	0.065	0.048	0.042	0.017	0.929	0.026	0.141
Mean MSE			0.354	0.196	0.113	0.111	0.310	0.378	0.412	0.236	0.122	0.105	0.077	0.054	0.683	0.113	0.108
Mean MSE, All			0.217	0.121	0.075	0.054	0.187	0.224	0.297	0.126	0.067	0.049	0.036	0.034	0.491	0.075	0.053

(a) 10,000 iterations each experiment, based on VMA(5) and GARCH(1,1) processes with *t*-distributed errors and parameters obtained from Dow Jones returns data;
(b) daily return rates aggregated over nonoverlapping periods of 5 and 20 days;
(c) based on parametric VAR(*p*) estimation, with *p* selected using AIC and SBC;
(d) Schwert (1989) lag selection rule $k = \lceil 4(T/100)^{1/4} \rceil$;
(e) Newey-West covariance matrix with standard weight vector of ones and no prewhitening.
(f) equation 3.14 with $a = 0$, ζ = 2 and Newey-West based selection of k;
(g) equation 3.14 with $a = 0$, ζ = 4 and Newey-West based selection of k;
(h) equation 3.14 with $a = 0$, ζ = 12 and Newey-West based selection of k;
(i) equations 3.13 and 3.14 with $-10 \le a \le 10$, ζ = 2 and Newey-West based selection of a and k;
(j) equations 3.13 and 3.14 with $-10 \le a \le 10$, ζ = 4 and Newey-West based selection of a and k;
(k) equations 3.13 and 3.14 with $-10 \le a \le 10$, ζ = 12 and Newey-West based selection of a and k;
(l) equation 3.15 with VAR(1) prewhitening and Andrews-Monahan based selection of k;
(m) equation 3.15 with VAR-AIC prewhitening and Andrews-Monahan based selection of k;
(n) equation 3.15 with VAR-SBC prewhitening and Andrews-Monahan based selection of k.