THE EXACT ERROR IN ESTIMATING THE SPECTRAL DENSITY AT THE ORIGIN

BY SERENA NG AND PIERRE PERRON

Université de Montréal and C.R.D.E.

First version received November 1994

Abstract. This paper derives expressions for the exact bias and variance of a general class of spectral density estimators at the zero frequency, building on the work of Neave (The exact error in spectrum estimates. *Ann. Math. Statist.* 42 (1971), 961–75) who studied the case where the mean of the series is assumed known. These expressions are evaluated for 15 different windows and for a wide variety of stationary time series. The exact error of the estimators is found to depend on whether the sample mean has to be estimated, and some windows are noticeably inferior at certain values of the bandwidth. A response surface analysis reveals that the finite sample relationships between the bandwidth and the exact error are quite different from the ones suggested by asymptotic theory.

Keywords. Kernel; persistence measures; bandwidth; lag window; non-parametric inference.

1. INTRODUCTION

Recent developments on several fronts have made the spectral density at frequency zero a concept of increasing importance in economics. In studies of business cycles, the spectral density function is used to measure the degree of persistence of shocks on aggregate activities. In econometric applications, an estimate of the spectral density at frequency zero is often required in analyses of non-stationary time series and in Generalized Methods of Moments estimations that require a consistent estimate of the variance–covariance matrix.

An important element in the recent debate concerning whether or not an aggregate time series is better characterized by the presence of a unit root is the degree of persistence of the innovations on the level of the series. If a series contains a unit root, the innovations will have a lasting effect on its level and the shocks are said to be persistent. If the variable is stationary around a deterministic trend, the effects of the innovations will eventually vanish and the shocks are said to have zero persistence.

An area of research has developed in recent years to quantify the degree of persistence in aggregate data. Studies of this nature include Campbell and Mankiw (1987), Cochrane (1988), Watson (1986) and Clark (1987). Campbell and Mankiw's (1987) measure of persistence is motivated by the fact that for

0143-9782/96/04 379–408 JOURNAL OF TIME SERIES ANALYSIS Vol. 17, No. 4

an integrated process, say y_t, its first difference, denoted Δy_t, has a moving-average representation $\Delta y_t = \psi(L)\varepsilon_t$, $\varepsilon_t \sim$ i.i.d. $(0, \sigma_\varepsilon^2)$, where Δ is the first difference operator and L is the lag operator such that $Ly_t = y_{t-1}$. A natural measure of persistence is $\psi(1)$ since it is the cumulative effect of a unit shock on the level of y_t in the indefinite future. If y_t has a unit root, $\psi(1)$ will be non-zero, implying that a unit shock has a lasting effect on the level of y_t. However, if y_t is stationary, $\psi(1)$ will be zero, implying that a shock will only have transitory effects on y_t.

The role of the spectral density function in these analyses can be seen by noting that $f_{\Delta y}(0) = \sigma_\varepsilon^2 \psi(1)^2/2\pi$ is the non-normalized spectral density of Δy_t evaluated at the zero frequency. If Δy_t is an autoregressive moving-average (ARMA) process with autoregressive and moving-average polynomials $A(L)$ and $B(L)$ respectively, its non-normalized spectral density function at frequency zero is $f_{\Delta y}(0) = \{\sigma_\varepsilon^2 B(1)^2/A(1)^2\}/2\pi$, hence $\psi(1) = B(1)/A(1) = \{2\pi f_{\Delta y}(0)/\sigma_\varepsilon^2\}^{1/2}$. An estimate of $\psi(1)$ can therefore be recovered given an estimate of σ_ε^2 and the spectral density function evaluated at the origin.[1] Other measures of persistence such as the variance ratio statistic of Cochrane (1988), defined as $V(k) = \operatorname{var}(y_t - y_{t-k})/\{k \operatorname{var}(y_t - y_{t-1})\}$, can be shown to relate to $\psi(1)$ by noting that $V(k)$ can also be expressed as $1 + 2\sum_{j=1}^{k-1}(1 - j/k)\operatorname{corr}(\Delta y_t, \Delta y_{t-j})$, $\operatorname{corr}(\cdot)$ being the autocorrelation function. Hence $\lim_{k\to\infty} V(k) = 2\pi f_{\Delta y}(0)$.

The second need for an estimate of the spectral density at frequency zero is also related to the unit root literature. It is now known that classical inference does not apply when the regressors are integrated and new estimation and inference techniques have been developed for analyzing non-stationary time series. The need for a consistent estimate of the spectral density at the origin by these procedures is a rule rather than an exception.[2]

A third need for an estimator of the spectral density function at the origin arises in a somewhat different context. In econometric applications, especially in estimating the first-order conditions of rational expectations models, it is often necessary to correct for conditional heteroscedasticity and serial correlation in the residuals while ensuring that the resulting variance–covariance matrix is positive semi-definite. The widely used variance estimator proposed by Newey and West (1987) is a matrix version of the scaled spectral density at frequency zero using modified Bartlett weights to smooth the sample autocovariance function. In spite of the popularity of this estimator, little is known about the relative efficiency of the Bartlett weights as compared with other kernels which serve the same purpose. Our analysis sheds light on this issue.

Obtaining a consistent estimate of the spectral density function has been an area of much research (see Priestley (1981) for an overview). We shall analyze the properties of estimators at the zero frequency for two cases. The first is when the mean of the series is known. The second case is when the mean of the series is unknown, as in the analyses of Cochrane (1988) and Campbell and Mankiw (1987). Although the periodogram of a series and its mean-corrected variant are identical at non-zero frequencies, the equivalence breaks down at frequency zero (Brockwell and Davies, 1991, Section 10.4). It is therefore necessary to consider the case with and without an estimated mean.

2. NON-PARAMETRIC ESTIMATORS OF THE SPECTRAL DENSITY FUNCTION

Consider a real-valued, weakly stationary linear process X_t with finite mean $\mu = E(X_t)$ and autocovariance at lag v given by $R(v) = E\{(X_t - \mu)(X_{t+v} - \mu)\}$. It is assumed that $R(v)$ is continuous at $v = 0$ and $\int_{-\infty}^{\infty}|R(v)|\,dv < \infty$. Then the power spectrum of X_t exists over $-\pi \leqslant \omega \leqslant \pi$ and is defined as follows:

$$f(\omega) = \frac{1}{2\pi}\sum_{v=-\infty}^{\infty} R(v)\cos(v\omega) = \frac{1}{2\pi}\left\{R(0) + 2\sum_{v=1}^{\infty} R(v)\cos(v\omega)\right\} \tag{2.1}$$

noting that the relation $R(v) = \int_{-\pi}^{\pi} f(\omega)\cos(v\omega)\,d\omega$ holds. A formal treatment of the properties of $f(\omega)$ can be found in Priestley (1981).

Our focus is on a class of spectral density estimators defined for $\mu = 0$ by

$$f_T(\omega) = \frac{1}{2\pi}\sum_{v=-(T-1)}^{T-1} k_T^*\left(\frac{v}{M_T}\right)\hat{R}_T^*(v)\cos(v\omega) \tag{2.2}$$

where $\hat{R}_T^*(v) = T^{-1}\sum_{t=1}^{T-v}X_tX_{t+v}$ and $k_T^*(v/M_T)$ is a non-negative, bounded even function often referred to as a lag window or kernel. Many proposed windows assume that $k_T^*(v/M_T) = 0$ if $v > M_T$, in which case M_T acts as a truncation lag parameter. In general, M_T is the bandwidth which, together with the lag window, imposes different weights on different sample autocovariances. Estimators of the form (2.2) are consistent for $f(\omega)$ if $M_T/T \to 0$ and $M_T \to \infty$ as $T \to \infty$.

We are interested in estimating $\sigma_\varepsilon^2\psi(1)^2 = 2\pi f(0) = \sum_{v=-\infty}^{\infty}R(v)$ which, for notational simplicity, we shall denote as $h(0)$. Also denote θ as the ratio v/M_T. The estimator corresponding to (2.2) is given, for the case of a zero mean series, by

$$\hat{h}_T = \sum_{v=-(T-1)}^{T-1} k_T^*(\theta)\hat{R}_T^*(v). \tag{2.3}$$

In the case where the mean is unknown, the estimator is given by

$$\tilde{h}_T = \sum_{v=-(T-1)}^{T-1} k_T^*(\theta)\tilde{R}_T^*(v) \tag{2.4}$$

with $\tilde{R}_T^*(v) = T^{-1}\sum_{t=1}^{T-v}(X_t - \bar{X})(X_{t+v} - \bar{X})$ and $\bar{X} = T^{-1}\sum_{t=1}^{T}X_t$.

The asymptotic properties of this class of estimators were analyzed by Parzen (1957), for example, under the assumption that the mean of X_t is known. Let r be the characteristic exponent of $k_T^*(\theta)$, i.e. r is the largest integer such that $k_T^{*(r)} = \lim_{\theta\to 0}[\{1 - k_T^*(\theta)\}/|\theta|^r]$ exists, is finite and is non-zero. Let $h^r(0) = \sum_{-\infty}^{\infty}|s|^r|R(s)|$ be the Parzen 'generalized rth derivative' of the spectral density function evaluated at frequency zero. It coincides with the rth derivative of $h(0)$ only when r is even. If there exists a $q \geqslant r$ such that $h^q(0) < \infty$ and $T/(M_T)^r \to \infty$ as $T \to \infty$, the asymptotic bias can be approximated by (see Priestley, 1981)

$$\text{asymptotic bias } (\hat{h}_T) \approx (-M_T^r)^{-1} h^r(0) k_T^{*(r)}. \tag{2.5}$$

However, if $r > q$, then the relationship between bias and M_T is less precise and it can only be said that the asymptotic bias is $o(M^{-q})$ if $M_T/T \to 0$. The asymptotic variance of these kernel estimators has the property that

$$\lim_{T \to \infty} (T/M_T) \operatorname{var}(\hat{h}_T) = 2h^2(0) \int_{-1}^{1} k_T^{*2}(\theta)\, d\theta \tag{2.6}$$

if $M_T/T \to 0$ and $M_T \to \infty$ as $T \to \infty$.

The limiting bias and variance are kernel dependent since the value of r, $k_T^{*(r)}$ and k_T^{*2} are kernel specific. In large samples, it can readily be seen that the bias is decreasing, but the variance is increasing in the bandwidth. Note also that while the asymptotic bias depends on the curvature of the spectral density function evaluated at frequency zero, as given by $h^r(0)$, the variance depends on the value of $h(0)$ itself.

Our interest is in the bias and variance of $\hat{h}_T$ and $\tilde{h}_T$ in finite samples. A number of lag windows have been proposed to estimate the spectral density function. Useful references on this topic can be found in Priestley (1981), Hannan (1970) and Neave (1972). We select for analysis 15 lag windows that are more or less common in this literature. The complete list of windows is summarized in Table I. All the windows considered satisfy the conditions discussed in Andrews (1991), namely that $k_T^*(0) = 1$, $k_T^*(x) = k_T^*(-x)$, $\int_{-\infty}^{\infty} k_T^*(x)^2 < \infty$ and $k_T^*(\cdot)$ is continuous at 0 and all but a finite number of other points. With the exception of the quadratic window, $k_T^*(\theta)$ is set to 0 if $\theta > 1$. This restriction is not always imposed on the Daniell window, but was used in Neave (1972), whose definition we followed.

Of the 15 window generators considered, only six have non-negative Fourier transforms. These are the Bartlett (2, 3), Parzen (4), Bohman (7), Daniell (8) and Quadratic (14) windows. We include the Bartlett window specified as (10) because it was used in many studies which measure the degree of persistence in macroeconomic aggregates (e.g. Campbell and Mankiw, 1987). Also of interest is the quadratic spectral kernel found by Andrews (1991) to have some optimal properties, but it is the only kernel considered where M_T does not act as a truncation lag. The trapezoid window has recently been shown to have relatively small bias in Politis and Romano (1995), and is also included in the analysis.

Some of the windows listed above have been studied from both a theoretical and a numerical point of view. Neave (1972) found some optimality properties for the Parzen, normal and Bohman generators in that they have quite a narrow peak at $\theta = 0$. Neave (1971) analyzed the exact bias and variance of the spectral estimator $f_T(\omega)$ when the mean is known using the Parzen and the

TABLE I
LIST OF WINDOWS CONSIDERED

1 Truncated periodogram	$k^*(\theta) = 1$
2 Bartlett (a) with $\hat{R}_T^*(v)$	$k^*(\theta) = 1 - \theta$
3 Bartlett (b) with $\hat{R}_T(v)$	$k^*(\theta) = \dfrac{(1-\theta)T}{T-v}$
4 Parzen (a)	$k^*(\theta) = \begin{cases} 1 - 6\theta^2 + 6\theta^3 & 0 \leq \theta \leq \frac{1}{2} \\ 2(1-\theta)^3 & \frac{1}{2} \leq \theta \leq 1 \end{cases}$
5 Tukey–Hamming	$k^*(\theta) = 0.54 + 0.46\cos(\pi\theta)$
6 Tukey–Hanning	$k^*(\theta) = \frac{1}{2}\{1 + \cos(\pi\theta)\}$
7 Bohman	$k^*(\theta) = (1 - \theta)\cos(\pi\theta) + \sin(\pi\theta)/\pi$
8 Daniell	$k^*(\theta) = \dfrac{\sin(\pi\theta)}{\pi\theta}$
9 Parzen (b)	$k^*(\theta) = 1 - \theta^2$
10 Bartlett (c)	$k^*(\theta) = \dfrac{\{1 - v/(M_T+1)\}T}{T-v}$
11 Parzen (c)	$k^*(\theta) = \dfrac{1}{1+\theta^2}$
12 Tukey – Parzen	$k^*(\theta) = 0.436 + 0.564\cos(\pi\theta)$
13 Normal	$k^*(\theta) = \exp(-4.5\theta^2)$
14 Quadratic	$k^*(\theta) = \dfrac{25}{12\pi^2\theta^2}\left\{\dfrac{\sin(6\pi\theta/5)}{6\pi\theta/5} - \cos(6\pi\theta/5)\right\}$
15 Trapezoid	$k^*(\theta) = \begin{cases} 1 & 0 \leq \theta \leq \frac{1}{2} \\ 2(1-\theta) & \frac{1}{2} \leq \theta \leq 1 \end{cases}$

Tukey–Hamming windows for various shapes of the spectral density function. To the authors' knowledge, no extensive study has provided exact analytical results for estimators of the spectral density function in the important special case of zero frequency. The purpose of the following sections is to present such a study. To do this, we shall first derive the exact bias and variance of the estimators $\hat{h}_T$ and $\tilde{h}_T$.

3. THE EXACT BIAS AND VARIANCE OF $\hat{h}_T$, KNOWN MEAN

In this section, we derive the exact error of $\hat{h}_T$ as defined in (2.3) given an arbitrary lag window $k_T^*(v)$ and bandwidth M_T. Note that (2.3) is based on $\hat{R}_T^*$, the biased estimator of autocovariances. Of course, one could use the unbiased estimator $\hat{R}_T(v) = (T-v)^{-1}\sum_{t=1}^{T-v} X_t X_{t+v}$ instead of $\hat{R}_T^*(v)$. There is some evidence, however, that the biased estimator has better properties in terms of mean-squared error (MSE) (e.g. Priestley, 1981, p. 323). In most of the cases studied, the exact bias and variance are derived for $\hat{h}_T$ based on $\hat{R}_T^*(v)$. The exception is the Bartlett (b) and Bartlett (c) windows, which are based on $\hat{R}_T(v)$, so as to analyze the specification used by Campbell and Mankiw (1987) and Cochrane (1988).

For mathematical convenience, it is useful in the derivations to rescale the lag window so that $\hat{h}_T$ can be expressed in terms of $\hat{R}_T$ instead of $\hat{R}_T^*$, noting that both are even functions of v. By defining $k_T(v) = (1 - v/T)k_T^*(v)$, $k_T(0) = 0.5$, (2.3) can be rewritten as

$$\hat{h}_T = 2\sum_{v=0}^{T-1} k_T(v)\hat{R}_T(v). \tag{3.1}$$

Since $\hat{R}_T(v)$ is an unbiased estimate of $R(v)$, we have $E(\hat{h}_T) = 2\sum_{v=0}^{T-1} k_T(v)R(v)$ and the bias is given by

$$\text{bias}(\hat{h}_T) = E\{\hat{h}_T - h(0)\} = 2\sum_{v=1}^{T-1}\{k_T(v) - 1\}R(v) - 2\sum_{v=T}^{\infty} R(v). \tag{3.2}$$

The exact variance of $\hat{h}_T$ is given by

$$\begin{aligned}\text{var}(\hat{h}_T) &= E(\hat{h}_T^2) - \{E(\hat{h}_T)\}^2 \\ &= 4\sum_{u,\,v=0}^{T-1} k_T(u)k_T(v)[E\{\hat{R}_T(u)\cdot\hat{R}_T(v)\} - E\{\hat{R}_T(u)\}\cdot E\{\hat{R}_T(v)\}]\end{aligned}$$

whose close form was derived by Neave (1971) under the assumption that X_t is normal. Evaluating Neave's result at the zero frequency gives

$$\begin{aligned}\text{var}(\hat{h}_T) = 8\Bigg(&\sum_{v=0}^{T-1}\frac{k_T^2(v)}{(T-v)^2}\Bigg[\frac{T-v}{2}\{R^2(0)+R^2(v)\} \\ &+\sum_{x=1}^{T-v}(T-v-x)\{R^2(x)+R(x-v)R(x+v)\}\Bigg] \\ &+2\sum_{0\leqslant u\leqslant v\leqslant T-1}\frac{k_T(v)k_T(u)}{(T-v)(T-u)}\Bigg[\frac{T-u}{2}\sum_{x=0}^{u-v}\{R(x)R(x+v-u) \\ &+R(u-x)R(x+v)\} + \sum_{x=u-v+1}^{T-v}(T-v-x)\{R(x)R(x+v-u) \\ &+R(x-u)R(x+v)\}\Bigg]\Bigg).\end{aligned} \tag{3.3}$$

Expressions (3.2) and (3.3) enable us to compute numerically the exact bias and variance, and hence the MSE of the estimator $\hat{h}_T$ for various combinations of windows, bandwidths and underlying data generating processes for the X_t's.

4. THE EXACT BIAS AND VARIANCE OF $\tilde{h}_T$; UNKNOWN MEAN

The estimator given by (2.4) for the unknown mean case is expressed in terms of $\tilde{R}^*_T$. As in the previous section, it proves convenient to express $\tilde{h}_T$ in terms of the estimates $\tilde{R}_T(v) = (T-v)^{-1}\sum_{t=1}^{T-v}(X_t - \bar{X})(X_{t+v} - \bar{X})$. Letting $k_T(v) = (1 - v/T)k^*_T(v)$ gives

$$\tilde{h}_T = 2\sum_{v=0}^{T-1} k_T(v)\tilde{R}_T(v). \tag{4.1}$$

Since the sample mean is subtracted from each observation, $\tilde{R}_T(v)$ is not an exactly unbiased estimator of $R(v)$, the true autocovariance at lag v. Define the quantities

$$\operatorname{var}(\bar{X}) = E(\bar{X} - \mu)^2 = T^{-2}\sum_{t,\,s=1}^{T} R(t-s) = T^{-1}R(0) + 2\left\{T^{-1}\sum_{r=1}^{T}(1 - r/T)R(r)\right\} \tag{4.2}$$

and

$$\bar{R}(j) = T^{-1}\sum_{k=1}^{T} R(k-j). \tag{4.3}$$

It is then straightforward to show the following results:

$$E\{(X_j - \mu)(\bar{X} - \mu)\} = \bar{R}(j) \tag{4.4}$$

$$E\{(X_j - \bar{X})(X_i - \bar{X})\} = R(j-i) + \operatorname{var}(\bar{X}) - \bar{R}(j) - \bar{R}(i) \tag{4.5}$$

$$E\{\tilde{R}_T(v)\} = R(v) + \operatorname{var}(\bar{X}) - (T-v)^{-1}\sum_{t=1}^{T-v}\{\bar{R}(t+v) + \bar{R}(t)\}. \tag{4.6}$$

Using (4.6) we obtain the following expression for the bias of $\tilde{h}_T$:

$$\begin{aligned}\operatorname{bias}(\tilde{h}_T) &= E\{\tilde{h}_T - h(0)\}\\ &= \operatorname{bias}(\hat{h}_T) + 2\sum_{v=0}^{T-1} k_T(v)[\operatorname{var}(\bar{X}) - (T-v)^{-1}\sum_{t=1}^{T-v}\{\bar{R}(t) + \bar{R}(t+v)\}].\end{aligned} \tag{4.7}$$

It is straightforward to show, under the conditions assumed here, that the second term on the right-hand side of (4.7) is $O(M_T/T)$. Hence, the difference between the bias in the known and unknown mean cases vanishes as T increases. However, it also highlights the fact that larger differences can be expected if the truncation lag is large relative to the sample size.

The exact variance of $\tilde{h}_T$ can be derived following the method used by Neave (1971) for the case of $\hat{h}_T$. Using

$$E(\tilde{h}_T) = 2\sum_{v=0}^{T-1} k_T(v)E\{\tilde{R}_T(v)\}$$

$$E(\tilde{h}_T^2) = 4\sum_{u,\,v=0}^{T-1} k_T(v)k_T(u)E\{\tilde{R}_T(v)\tilde{R}_T(u)\}$$

and $\operatorname{var}(\tilde{h}_T) = E(\tilde{h}_T^2) - \{E(\tilde{h}_T)\}^2$, we have

$$\operatorname{var}(\tilde{h}_T) = 4\sum_{u,\,v=0}^{T-1} k_T(v)k_T(u)\operatorname{cov}\{\tilde{R}_T(v), \tilde{R}_T(u)\} \tag{4.8}$$

where

$$\operatorname{cov}\{\tilde{R}_T(v), \tilde{R}_T(u)\} = E\{\tilde{R}_T(v)\tilde{R}_T(u)\} - E\{\tilde{R}_T(v)\}E\{\tilde{R}_T(u)\}. \tag{4.9}$$

Using the expression for $\operatorname{cov}\{\tilde{R}_T(v), \tilde{R}_T(u)\}$ and the simplifications given in the appendix, it can be shown that

$$\begin{aligned}\operatorname{var}(\tilde{h}_T) = \operatorname{var}(\hat{h}_T) + 4\Bigg(&\sum_{u,\,v=0}^{T-1}(T-v)^{-1}(T-u)^{-1}\sum_{t=1}^{T-v}\sum_{s=1}^{T-u}k_T(u)k_T(v)\\
&\times [2\operatorname{var}(\bar{X})^2 + \operatorname{var}(\bar{X})\{R(t-s) + R(t+v-s-u)\\
&+ R(t-s-u) + R(t+v-s)\} - 2\operatorname{var}(\bar{X})\{\bar{R}(t+v) + \bar{R}(t)\\
&+ \bar{R}(s) + \bar{R}(s+u)\} + 2\{\bar{R}(t)\bar{R}(t+v) + \bar{R}(s)\bar{R}(s+u)\}\\
&+ \{\bar{R}(t+v) + \bar{R}(t)\}\{\bar{R}(s) + \bar{R}(s+u)\} - \bar{R}(t+v)\{R(t-s)\\
&+ R(t-s-u)\} + \bar{R}(t)\{R(t+v-s-u) + R(t+v-s)\}\\
&- \bar{R}(s+u)\{R(t-s) + R(t+v-s)\} + \bar{R}(s)\{R(t+v-s-u)\\
&+ R(t-s-u)\}]\Bigg).\end{aligned} \tag{4.10}$$

This expression for the exact variance is valid under normality of X_t, or at least, under the condition that the fourth cumulant of the distribution function is zero (see Priestley, 1981, p. 325). A comparison of (4.7) and (4.10) with (3.2) and (3.3) indicates that the bias and variance of $\tilde{h}_T(0)$ involve additional (not necessarily positive) terms that are absent from the corresponding expressions for the zero mean case. Thus, the exact error in estimating the spectral density at the origin depends on the treatment of the sample mean.

5. RESULTS

The expressions derived in Sections 3 and 4, while complex, can easily be evaluated numerically given (i) a value of the bandwidth M_T; (ii) a value for the number of observations, T; (iii) a lag window function $k_T(v)$; and (iv) the autocovariance function $R_T(v)$ of the underlying process. For each of the 15 windows described in Section 2, we consider three values for the sample size, $T = 50$, 100, 150, and for each sample size, ten different values for the bandwidth parameter M_T. These are chosen as follows:

(i) $T = 50$, $M_T = 2$, 4, 8, 12, 16, 20, 25, 30, 40, 49;
(ii) $T = 100$, $M_T = 2$, 4, 8, 14, 20, 30, 40, 50, 70, 99;
(iii) $T = 150$, $M_T = 2$, 4, 8, 14, 20, 30, 50, 70, 100, 149.

We increase M_T all the way up to $T - 1$ in order to analyze fully the effect of a large value of the bandwidth on the properties of the estimators. As will transpire from the results below, some estimators have peculiar properties at such extreme values of the bandwidth.

We analyze the behaviour of $\hat{h}_T$ and $\tilde{h}_T$ for 24 time series processes within the class of ARMA models. Those models, the associated autocovariance function and true value of the spectral density at frequency zero are given in Table II. For conciseness in the discussion below, we organize the 24 data generating processes into five groups:

(1) Nearly integrated models: AR1(0.9), AR2(1.30, −0.35), ARMA(0.9, 0.6) and ARMA (0.9, −0.5). The autoregressive root is close to the unit circle and hence $h(0)$ is large.

(2) Positively autocorrelated models: AR1(0.4), AR2(0.2, 0.4), ARMA(0.6, −0.5), and ARMA(0.9, −0.8).

(3) Negatively autocorrelated models: AR1(−0.4), AR1(−0.9).

(4) Positive moving-average models: MA1(0), MA1(0.5), MA1(1.0), MA2(0.5, 0.5), ARMA(−0.3, 0.6), ARMA(0.3, 0.6). The moving-average coefficients are positive and larger in absolute value than the autoregressive coefficients.

(5) Negative moving-average models: MA1(−0.5), MA1(−0.8), MA1(−1.0), MA2(−0.5, −0.5), ARMA(0.9, −1.0), ARMA(0.6, −1.0), ARMA(0.6, −0.8) and ARMA(−0.6, −1.0). The moving-average coefficients are negative and larger in absolute value than the autoregressive coefficients, and hence $h(0)$ is small.

Our experimental design is guided by the empirical properties of various macroeconomic time series of interest. The highly persistent processes in Group 1 are typical characterizations of the logarithm of many economic time series, and in many cases the data cannot reject an autoregressive unit root. As discussed earlier, the distinction between a trend stationary and a differenced stationary process is an important issue in empirical macroeconomics, and is the motivation for using the spectral density function at frequency zero of the first-differenced series to assess the degree of persistence of a shock. Groups 2–5 contain processes that are intended to capture different time series properties of such differenced series. Of particular concern is the case where a series is

TABLE II

DATA GENARATING PROCESSES USED IN THE EXPERIMENTS

	Autocovariance $R(v)$		$h(0)$	Parameters
noise	1	$v = 0$	1.0	–
	0	$v \neq 0$		
	$1 + b_1^2 + b_2^2$	$v = 0$	$(1 + b_1 + b_2)^2$	$b_1 = 0.5,\ b_2 = 0.0$
$+ b_1 e_{t-1} + b_2 e_{t-2}$	$b_1 + b_1 b_2$	$v = 1$		$b_1 = 1.0,\ b_2 = 0.0$
	b_2	$v = 2$		$b_1 = -0.5,\ b_2 = 0.0$
	0	$v > 2$		$b_1 = -0.8,\ b_2 = 0.0$
				$b_1 = -1.0,\ b_2 = 0.0$
				$b_1 = -0.5,\ b_2 = 0.5$
				$b_1 = -0.5,\ b_2 = -0.5$
$_{t-1} + e_t$	$\dfrac{a^{\lvert v\rvert}}{1 - a^2}$		$\dfrac{1}{(1 - a)^2}$	$a = 0.4$
				$a = 0.9$
				$a = -0.4$
				$a = -0.9$
	$\dfrac{1 - a_2}{(1 + a_2)(1 + a_1 - a_2)(1 - a_1 - a_2)}$	$v = 0$	$\dfrac{1}{(1 - a_1 - a_2)^2}$	$a_1 = 0.2,\ a_2 = 0.4$
$x_{t-1} + a_2 x_{t-2}$	$\dfrac{(1 - \mu_2^2)\mu_1^{v+1} - (1 - \mu_1^2)\mu_2^{v+1}}{(\mu_1 - \mu_2)(1 + \mu_1\mu_2)} R(0)$	$v > 0$		$a_1 = 1.3,\ a_2 = -0.35$
(1, 1)	$1 + b^2 + \dfrac{2ab}{1 - a^2}$	$v = 0$	$\dfrac{(1 + b)^2}{(1 + a)^2}$	$a = -0.3,\ b = 0.6$
$_{-1} + e_t + be_{t-1}$	$\dfrac{(1 + ab)(a + b)}{1 - a^2}$	$v = 1$		$a = 0.9,\ b = -1.0$
	$a^{v-1} R(1)$	$v \geqslant 2$		$a = 0.9,\ b = -0.8$
				$a = 0.9,\ b = -0.5$
				$a = 0.6,\ b = -1.0$
				$a = 0.6,\ b = -0.8$
				$a = 0.6,\ b = -0.5$
				$a = -0.6,\ b = -1.0$
				$a = 0.9,\ b = -1.0$
				$a = 0.9,\ b = 0.6$
				$a = 0.3,\ b = 0.6$

overdifferenced since the first-differenced series has a non-invertible moving-average component. We therefore include a variety of overdifferenced processes in Group 5 to assess the properties of the estimators in such cases. In choosing the sample sizes, we are motivated, in part, by the size of the samples typically encountered in macroeconomic analyses, and in part by computational issues as discussed in the appendix.

Our analysis results in the computation of 720 exact biases, variances and MSEs for each treatment of the sample mean. The complete set of results is available on request.

5.1. *Mean squared error*

The MSE is amongst the most commonly used criteria in kernel selection. It is of interest to identify the kernels with the smallest MSE. Tables III and IV list, for each sample size, the window and the associated bandwidth that generates the smallest MSE for each of the 24 time series models considered. We focus the discussion on the mean unknown case with $T = 100$ without loss of generality.[3] Although the overall results reveal that no single estimator is uniformly best in terms of MSE, it is clear that the Bartlett (a) and Bartlett (b) windows are always dominated by the other windows considered. The trapezoid window has the smallest MSE for processes in Group 1, i.e. those with an autoregressive root close to unity. This window also appears to be the best for processes in Group 4, where the positive correlation is of the moving-average type. For processes in Group 2, the picture is less clear, with the trapezoid, truncated and Tukey–Hanning windows performing better than other windows. For Group 3 with negative autocorrelation of the autoregressive type, the quadratic and Tukey–Hanning windows appear to perform well. For processes with a strong negative moving-average coefficient (Group 5), there is no clear winner except when the process is non-invertible and the mean is unknown, in which case the truncated and Bartlett (c) windows have the smallest MSE. This last result is not surprising since, by construction, these two windows always yield an estimate of 0 (the true value) when $M_T = T - 1$ and the mean is unknown.

The relationship between the MSE and the bandwidth is of utmost importance in practice. We select ten processes that are representative of the five data groups and graph this relationship for the Bartlett (a), Parzen (a), Bohman, quadratic and trapezoid windows in Figure 1, noting that the bandwidth increases as we move along the horizontal axis until $M_T = T - 1$. Hence, there are ten observations for each window. Since the relationship between the MSE and the bandwidth is similar for both treatments of the sample mean, we use the unknown mean case for illustration and focus on $T = 50$ without loss of generality. The following results are noteworthy.

(i) For processes with roots away from unity (Groups 2, 3 and 4), the optimal bandwidth is around four and all kernels perform equally well at this value (Figures 1(c), 1(d), 1(e), 1(f) and 1(g)).

TABLE III

WINDOW WITH LOWEST MSE AND THE ASSOCIATED BANDWITH; KNOWN MEAN

	$T = 50$		$T = 100$		$T = 150$	
Model	Window	M_T	Window	M_T	Window	M_T
AR1(0.9)	1	8	15	14	15	14
AR2(1.3, −0.35)	1	8	15	14	1	14
ARMA(0.9, 0.6)	1	8	15	14	15	14
ARMA(0.9, −0.5)	1	8	1	8	15	14
AR1(0.4)	15	2	1	2	1	2
AR2(0.2, 0.4)	1	4	1	4	15	8
ARMA(0.6, −0.5)	4	4	6	4	8	4
ARMA(0.9, −0.8)	14	8	1	8	1	8
AR1(− 0.4)	7, 13	4	5, 6, 12	4	5, 6, 12	4
AR1(−0.9)	4	8	4, 14	4	14	4
MA1(0)	4	2	4	2	4	2
MA1(0.5)	9	2	15	2	15	2
MA1(1)	9	2	15	2	15	2
MA2(0.5, 0.5)	12	2	1	2	1	2
ARMA(−0.3, 0.6)	12	2	8	2	8	2
ARMA(0.3, 0.6)	15	2	15	2	1	2
MA1(− 0.5)	4, 5, 6	8	4, 5, 6, 8	8	4, 5, 6, 8	8
MA1(−0.8)	[a]	25	[a]	50	[a]	100
MA1(−1)	15	25	4	99	4	100
ARMA(0.9, −1)	[a]	49	[a]	99	[a]	149
ARMA(0.6, −1)	15	49	9	99	15	149
ARMA(−0.6, −1)	[a]	49	[a]	70	[a]	100
ARMA(0.6, −0.8)	5	16	4	30	6	20
MA2(−0.5, −0.5)	15	49	1	99	3	149

Note: [a]Window 2 has a significantly larger MSE but all other windows identical MSEs up to the second decimal place.

(ii) For nearly integrated processes (Group 1), a bandwidth of at least one-fifth the sample size seems desirable (Figures 1(a) and 1(b)). There is more variation in the optimal bandwidth across windows, with the Bohman and Parzen (a) windows requiring more lags than the Bartlett and the quadratic spectral windows.

(iii) For models with large negative moving-average coefficients (Group 5, Figures 1(h), 1(i) and 1(j)), the MSE declines with M_T. This negative relationship approaches monotonocity as the moving-average coefficient approaches -1. Thus, unlike models of Groups 2, 3 and 4, when the cost of a larger than optimal M_T tends to be higher than a smaller than optimal M_T, the opposite is true for models of Group 5.

Figure 1 also indicates that the performance of the estimator using any of the five kernels is similar at their optimal bandwidth. It is at sub-optimal bandwidths that differences across kernels become significant. In particular, the cost of using an unnecessarily large bandwidth is particularly high for the trapezoid window. On the other hand, when M_T is too small, the MSE

TABLE IV

WINDOW WITH LOWEST MSE AND THE ASSOCIATED BANDWIDTH; UNKNOWN MEAN

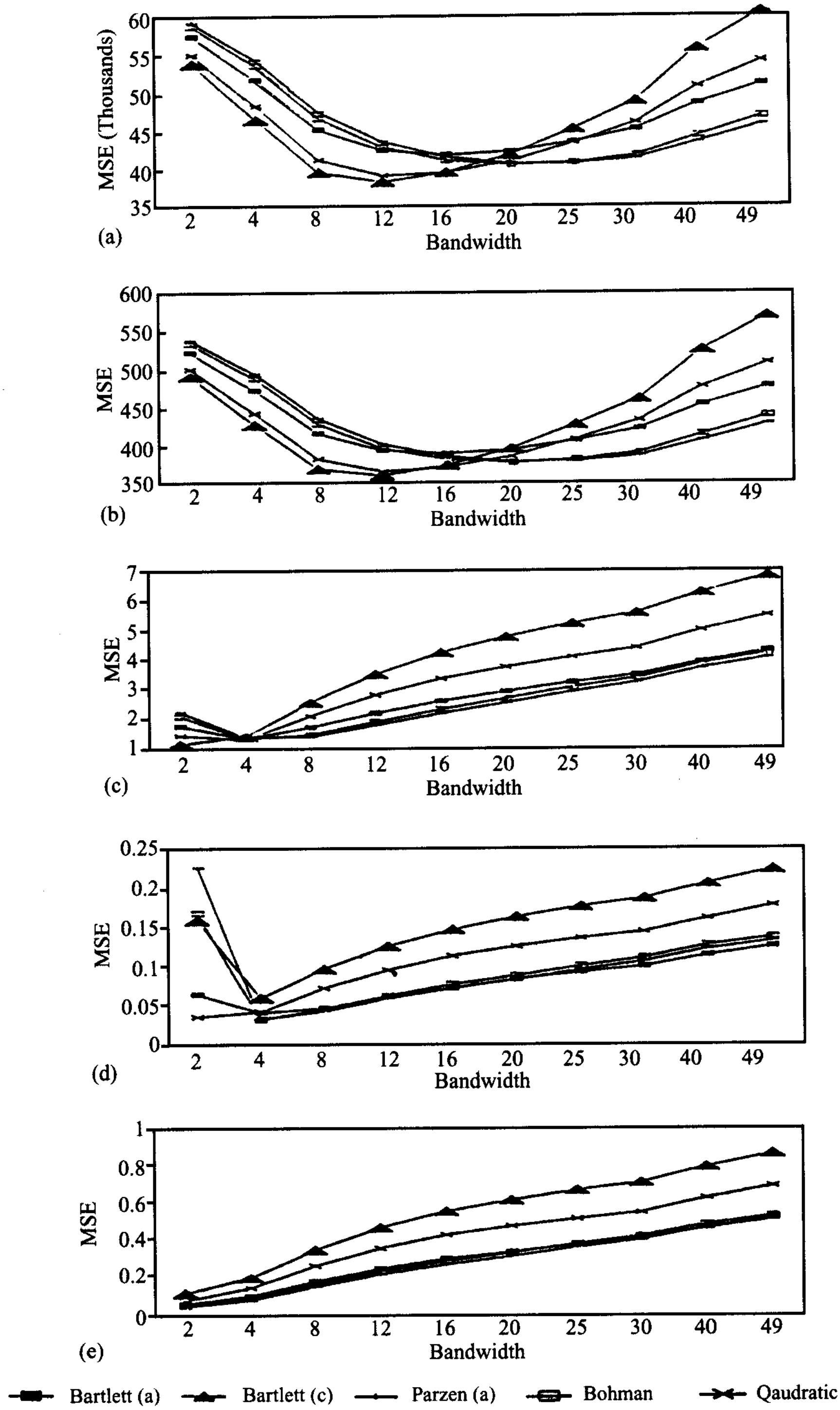
	T = 50		T = 100		T = 150	
Model	Window	M_T	Window	M_T	Window	M_T
AR1(0.9)	1	8	15	14	1	14
AR2(1.3, −0.35)	14	16	1	14	1	14
ARMA(0.9, 0.6)	1	8	15	8	1	14
ARMA(0.9, −0.5)	1	8	15	14	1	14
AR1(0.4)	15	2	1	2	1	2
AR2(0.2, 0.4)	1	4	1	4	15	8
ARMA(0.6, −0.5)	14	2	6	4	8	4
ARMA(0.9, −0.8)	14	8	1	8	1	8
AR1(− 0.4)	4, 13	4	6	4	6, 12	4
AR1(−0.9)	4	8	14	4	14	4
MA1(0)	4	2	4	2	4	2
MA1(0.5)	15	2	15	2	15	2
MA1(1)	15	2	15	2	15	2
MA2(0.5, 0.5)	1	2	1	2	1	2
ARMA(−0.3, 0.6)	12	2	8	2	8	2
ARMA(0.3, 0.6)	15	2	15	2	1	2
MA1(− 0.5)	4, 6	12	5	8	4, 5, 6	8
MA1(−0.8)	11	49	9	99	2	125
MA1(−1)	1, 10[a]	49	1, 10[a]	99	1, 10[a]	149
ARMA(0.9, −1)	1, 10[a]	49	1, 10[a]	99	1, 10[a]	149
ARMA(0.6, −1)	1, 10[a]	49	1, 10[a]	99	1, 10[a]	149
ARMA(−0.6, −1)	1, 10[a]	49	1, 10[a]	99	1, 10[a]	149
ARMA(0.6, −0.8)	2	40	2	30	4, 5, 13	30
MA2(−0.5, −0.5)	1, 10[a]	49	1, 10[a]	99	1, 10[a]	149

Note: [a]Windows 1 and 10 are 0 by construction when $M_T = T - 1$. Hence, when $h(0) = 0$, which occurs when the moving-average root is 1, these windows yield the smallest MSE by definition.

associated with the Parzen (a) window is large relative to the MSE of other kernels at the same bandwidth. Thus, the desirability of a window at a certain bandwidth does not imply its desirability at other bandwidths.

5.2. *Bias, variance and the sample mean*

Although for a given data generating process and kernel the MSE of $\hat{h}_T$ and $\tilde{h}_T$ appears quantitatively similar, the trade-off between bias and variance depends on whether the mean has to be estimated. To begin, we observe that windows 1 (truncated) and 10 (Bartlett (c)) always yield an estimated spectral density of zero with $M_T = T - 1$ when the sample mean is unknown. Accordingly, the bias recorded for these two windows at $M_T = T - 1$ is the negative of the true spectral density at frequency zero with a corresponding variance of zero. While less extreme, the variance of other estimators is also attracted towards zero as M_T increases, a feature that is unique to the case when the sample mean is unknown.


MSE (Thousands)
60
55
50
45
40
35
2 4 8 12 16 20 25 30 40 49
(a)
Bandwidth
600
550
500
450
400
350
MSE
2 4 8 12 16 20 25 30 40 49
(b)
Bandwidth
7
6
5
4
3
2
1
MSE
2 4 8 12 16 20 25 30 40 49
(c)
Bandwidth
0.25
0.2
0.15
0.1
0.05
0
MSE
2 4 8 12 16 20 25 30 40 49
(d)
Bandwidth
1
0.8
0.6
0.4
0.2
0
MSE
2 4 8 12 16 20 25 30 40 49
(e)
Bandwidth
Bartlett (a)
Bartlett (c)
Parzen (a)
Bohman
Qaudratic

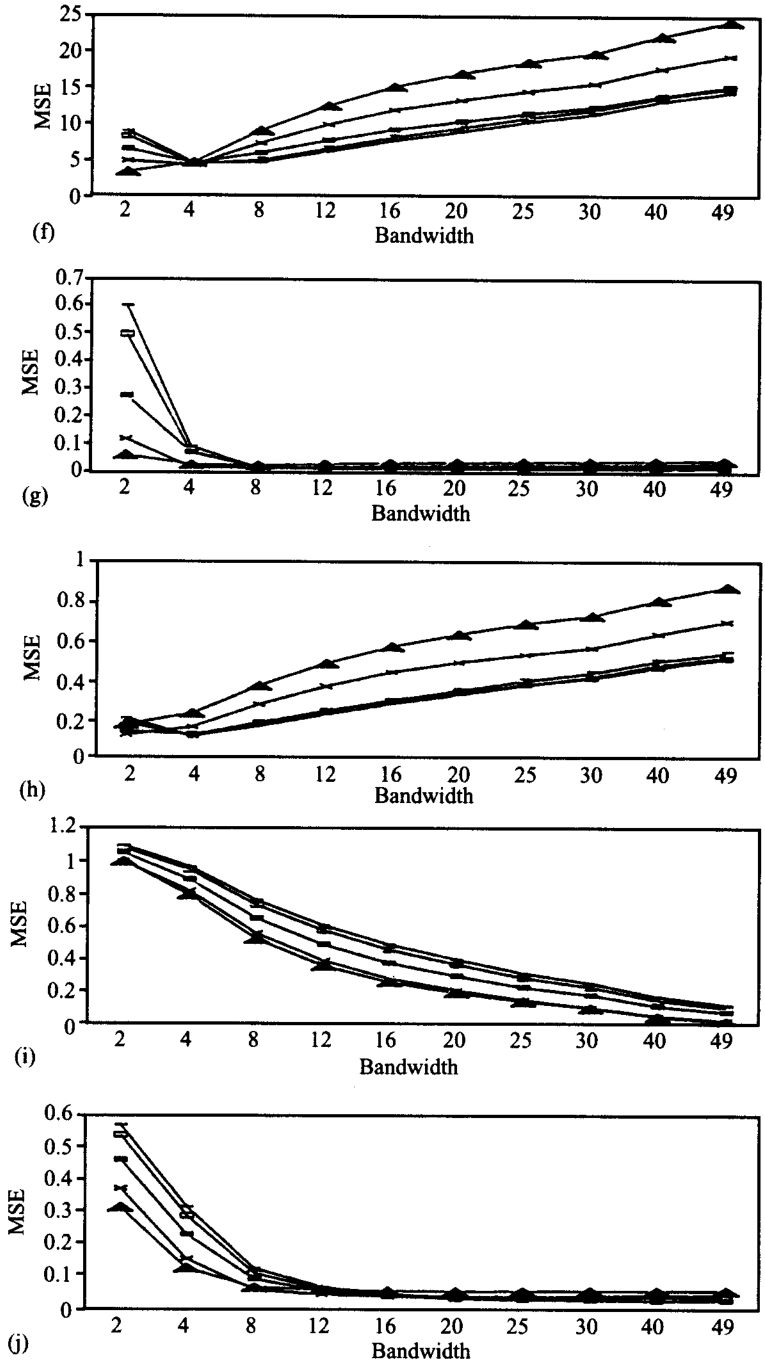


FIGURE 1. (a) ARMA(0.9, 0.6); (b) ARMA(0.9, −0.5); (c) AR(0.4); (d) AR(−0.4); (e) white noise; (f) ARMA(0.3, 0.6); (g) MA(−0.5); (h) ARMA(−0.5, −0.5); (i) ARMA(0.9, −1.0); (j) ARMA(0.6, −0.8).

The more substantive difference is that while bias is usually decreasing and variance increasing in the bandwidth when the mean is known, this is not the case when the sample mean has to be estimated. The difference in bias between the two cases can be explained using the relation (4.7). As discussed earlier, the difference in the two biases is $O(M_T/T)$. Hence for small values of the truncation lag (relative to the total sample), one can expect the biases to be similar. However, for large truncation lags, the bias in the unknown mean case is expected to diverge from that of the known mean case. This is consistent with our finding that the exact bias for stationary and invertible processes is monotonically decreasing in the bandwidth in the known mean case but that it eventually increases in the unknown mean case.

Both bias and variance tend to be non-linear functions of the bandwidth when the sample mean is unknown. However, while minimum bias always occurs at a unique bandwidth, a small variance can be attained at either a very small or a very large bandwidth. To highlight the differences in the exact error arising from the treatment of the sample mean, we select one time series from each of the five groups and depict the relationship between the bandwidth and the bias/variance at $T = 50$ for representative kernels. The main findings in Figures 2–7 are as follows:

(i) For models with positive serial correlation, such as AR(0.9), AR(0.4) and MA(0.5), we note from Figures 2, 3 and 5 that (a) $h(0)$ is underestimated as the bias is negative whether or not the mean has to be estimated; (b) the bias and variance are not monotonic in the bandwidth when the sample mean is unknown; and (c) the more persistent the process, the larger is the bandwidth which minimizes the bias when the mean is unknown.

(ii) For models with negative serial correlation such as AR1(-0.4) and MA1(-0.5), we note from Figures 4 and 6 that (a) bias is positive when the mean is known and declines as the bandwidth increases; and (b) bias is positive at small bandwidths but is negative as M_T increases when the mean is unknown, the absolute bias does not decline monotonically in M_T.

(iii) For models with negative moving-average components such as MA1(-0.5) we note from Figure 6 that variance is not monotonic in M_T even when the mean is known. The variance *declines* with M_T when the moving-average coefficient is close to -1. This is so for both treatments of the sample mean, as seen from Figure 7.

Looking at the results across the five groups makes it clear that the sign and magnitude of the bias is determined by whether the autoregressive coefficients are larger than the moving-average coefficients, and by whether the dominant root is positive or negative. The positive bias in estimating the spectral density of models with negative serial correlation is best understood by considering the case when the moving-average coefficient is -1. Since $h(0) = 0$, estimators constructed to yield positive definite estimates must necessarily overestimate $h(0)$. A continuity argument can be used to explain why the estimators have the tendency to overestimate $h(0)$ when it is close to zero.

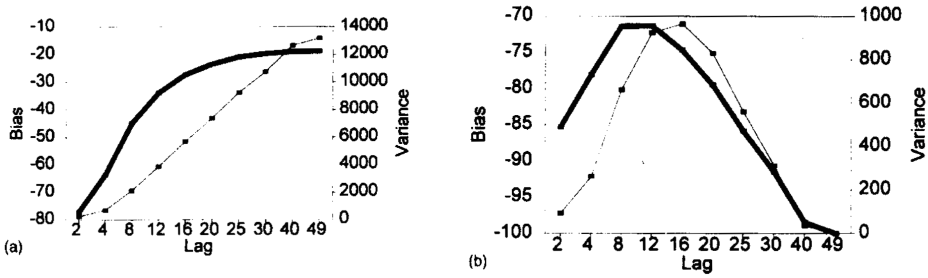


FIGURE 2. AR1(0.9), truncated: (a) known mean; (b) unknown mean.

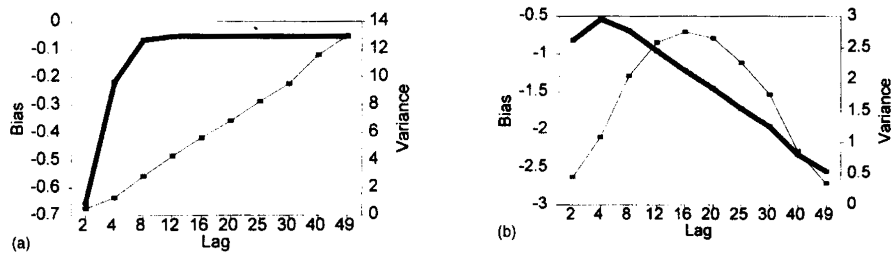


FIGURE 3. AR1(0.4), trapezoid: (a) known mean; (b) unknown mean.

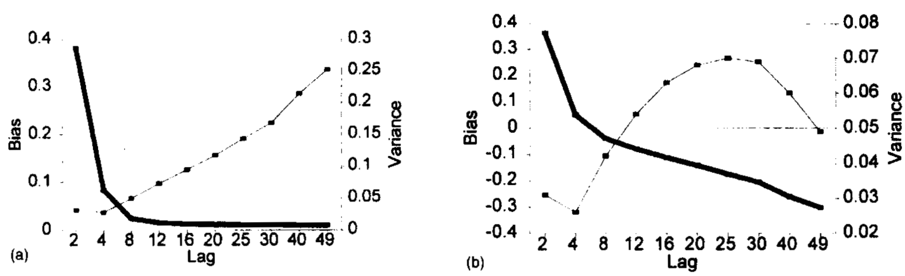


FIGURE 4. AR1(−0.4), normal: (a) known mean; (b) unknown mean.

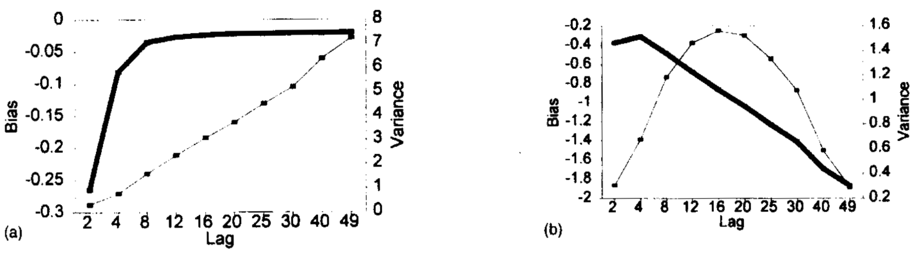


FIGURE 5. MA1(0.5), Parzen (b): (a) known mean; (b) unknown mean.

TABLE V

BIAS, VARIANCE, MSE AND RANKINGS FOR SELECTED PROCESSES, KNOWN MEAN, $T = 50$

	AR1(0.9)		AR1(0.4)		AR1(−0.4)		MA1(0.5)		MA1(−0.5)	
s										
	−18.8500	(1)	−0.0530	(1)	0.0100	(1)	−0.0200	(1)	0.0200	(1)
	−30.9740	(14)	−0.1040	(15)	0.0200	(15)	−0.0400	(15)	0.0400	(15)
	−19.2230	(3)	−0.0540	(4)	0.0100	(1)	−0.0200	(1)	0.0200	(1)
	−31.4010	(15)	−0.0660	(14)	0.0100	(1)	−0.0220	(12)	0.0220	(12)
	−25.5290	(9)	−0.0580	(9)	0.0100	(1)	−0.0210	(7)	0.0210	(7)
	−26.1090	(10)	−0.0590	(10)	0.0100	(1)	−0.0210	(7)	0.0210	(7)
	−29.9610	(13)	−0.0640	(12)	0.0100	(1)	−0.0220	(12)	0.0220	(12)
	−23.9940	(8)	−0.0570	(8)	0.0100	(1)	−0.0210	(7)	0.0210	(7)
	−22.2780	(6)	−0.0550	(5)	0.0100	(1)	−0.0200	(1)	0.0200	(1)
	−18.8500	(1)	−0.0530	(1)	0.0100	(1)	−0.0200	(1)	0.0200	(1)
	−21.7770	(5)	−0.0550	(5)	0.0100	(1)	−0.0200	(1)	0.0200	(1)
	−27.0390	(11)	−0.0600	(11)	0.0100	(1)	−0.0210	(7)	0.0210	(7)
	−29.8680	(12)	−0.0640	(12)	0.0100	(1)	−0.0220	(12)	0.0220	(12)
	−22.7430	(7)	−0.0560	(7)	0.0100	(1)	−0.0210	(7)	0.0210	(7)
	−19.4100	(4)	−0.0530	(1)	0.0100	(1)	−0.0200	(1)	0.0200	(1)
iance										
	227.8050	(15)	1.0480	(15)	0.0930	(15)	0.8470	(15)	0.0710	(15)
	37.5920	(5)	0.2360	(5)	0.0250	(1)	0.2250	(5)	0.0160	(1)
	38.3570	(7)	0.2400	(7)	0.0250	(1)	0.2290	(7)	0.0170	(5)
	21.2650	(1)	0.1440	(1)	0.0270	(5)	0.1420	(1)	0.0160	(1)
	46.8900	(8)	0.2750	(8)	0.0250	(1)	0.2530	(8)	0.0160	(1)
	37.5920	(5)	0.2360	(5)	0.0250	(1)	0.2250	(5)	0.0170	(5)
	25.2670	(3)	0.1660	(2)	0.0280	(7)	0.1630	(2)	0.0170	(5)
	48.4680	(9)	0.2980	(9)	0.0290	(9)	0.2800	(9)	0.0180	(10)
	58.5430	(10)	0.3550	(11)	0.0390	(13)	0.3310	(11)	0.0230	(12)
	86.1580	(13)	0.4430	(12)	0.0300	(11)	0.3770	(12)	0.0160	(1)
	119.1360	(14)	0.5880	(14)	0.0370	(12)	0.4900	(14)	0.0230	(12)
	24.8580	(2)	0.1830	(4)	0.0290	(9)	0.1890	(4)	0.0200	(11)
	26.3260	(4)	0.1710	(3)	0.0280	(7)	0.1660	(3)	0.0170	(5)
	58.5530	(11)	0.3520	(10)	0.0270	(5)	0.3250	(10)	0.0170	(5)
	84.1170	(12)	0.5000	(13)	0.0650	(14)	0.4590	(13)	0.0340	(14)

E

4152.2152	(1)	1.1309	(13)	0.1058	(15)	0.8474	(15)	0.0714	(15)
4439.2040	(15)	1.0806	(10)	0.0491	(11)	0.4851	(10)	0.0334	(12)
4426.5401	(14)	1.0757	(9)	0.0479	(10)	0.4790	(7)	0.0329	(11)
4325.8574	(10)	1.1070	(11)	0.0362	(4)	0.4606	(5)	0.0232	(2)
4327.4131	(11)	1.0422	(5)	0.0379	(8)	0.4748	(6)	0.0229	(1)
4284.5441	(6)	1.0156	(3)	0.0361	(3)	0.4851	(10)	0.0232	(2)
4316.9310	(8)	1.0641	(6)	0.0352	(2)	0.4796	(8)	0.0237	(5)
4289.3652	(7)	1.0668	(8)	0.0364	(6)	0.4214	(2)	0.0240	(7)
4235.7863	(4)	1.1418	(15)	0.0394	(9)	0.4012	(1)	0.0282	(9)
4405.3282	(13)	1.1236	(12)	0.0496	(12)	0.4879	(12)	0.0323	(10)
4322.2379	(9)	1.0210	(4)	0.0505	(13)	0.5367	(14)	0.0341	(13)
4262.0848	(5)	0.9978	(2)	0.0363	(5)	0.5173	(13)	0.0238	(6)
4335.7692	(12)	1.0655	(7)	0.0351	(1)	0.4846	(9)	0.0235	(4)
4194.7095	(2)	1.1354	(14)	0.0372	(7)	0.4319	(3)	0.0258	(8)
4225.8268	(3)	0.9277	(1)	0.0667	(14)	0.4594	(4)	0.0344	(14)

ankings are given in parentheses.

TABLE VI

Bias, Variance, MSE and Rankings for Selected Processes, Unknown Mean, $T = 50$

	AR1(0.9)		AR1(0.4)		AR1(−0.4)		MA1(0.5)		MA1(−0.5)	
s										
	−71.3420	(1)	−0.5570	(2)	−0.0620	(15)	−0.2390	(2)	−0.0050	(3)
	−76.6120	(15)	−0.7860	(14)	−0.0080	(3)	−0.4400	(15)	0.0090	(7)
	−75.4680	(14)	−0.7720	(13)	−0.0210	(8)	−0.4300	(14)	−0.0190	(13)
	−74.7160	(11)	−0.6990	(12)	−0.0280	(9)	−0.3600	(9)	0.0140	(11)
	−74.4790	(9)	−0.6830	(8)	0.0000	(1)	−0.3440	(8)	0.0120	(10)
	−74.0430	(8)	−0.6630	(6)	0.0060	(2)	−0.3390	(7)	0.0190	(13)
	−74.6200	(10)	−0.6880	(9)	−0.0370	(12)	−0.3680	(10)	0.0050	(3)
	−73.7770	(7)	−0.6790	(7)	−0.0170	(5)	−0.3220	(4)	0.0000	(1)
	−73.1890	(4)	−0.6560	(5)	−0.0350	(10)	−0.3100	(3)	−0.0150	(12)
	−75.4000	(13)	−0.7910	(15)	−0.0380	(13)	−0.4250	(13)	0.0110	(8)
	−73.4050	(6)	−0.6550	(4)	−0.0360	(11)	−0.3700	(11)	−0.0260	(15)
	−73.3440	(5)	−0.6310	(3)	0.0140	(4)	−0.3300	(6)	−0.0110	(8)
	−74.9340	(12)	−0.6950	(11)	−0.0400	(14)	−0.3740	(12)	0.0030	(2)
	−72.7650	(3)	−0.6880	(9)	−0.0200	(7)	−0.3230	(5)	−0.0060	(6)
	−71.6710	(2)	−0.5370	(1)	−0.0180	(6)	−0.1530	(1)	0.0050	(3)
riance										
	0.0210	(1)	0.0000	(1)	0.0000	(1)	0.0000	(1)	0.0000	(1)
	15.7270	(8)	0.2120	(9)	0.0240	(9)	0.2090	(9)	0.0080	(8)
	2.8320	(3)	0.0140	(3)	0.0020	(3)	0.0120	(3)	0.0010	(3)
	8.9430	(4)	0.1300	(4)	0.0250	(11)	0.1330	(5)	0.0150	(13)
	19.5070	(10)	0.2450	(11)	0.0230	(8)	0.2330	(11)	0.0100	(10)
	15.7270	(8)	0.2120	(9)	0.0240	(9)	0.2090	(9)	0.0120	(11)
	10.6080	(6)	0.1500	(5)	0.0260	(12)	0.1520	(6)	0.0150	(13)
	20.2390	(11)	0.2670	(12)	0.0260	(12)	0.2600	(13)	0.0080	(8)
	24.4160	(12)	0.3180	(14)	0.0180	(6)	0.2890	(14)	0.0060	(5)
	0.0210	(1)	0.0000	(1)	0.0000	(1)	0.0000	(1)	0.0000	(1)
	48.7550	(15)	0.1690	(8)	0.0070	(4)	0.1170	(4)	0.0020	(4)
	10.5300	(5)	0.1670	(7)	0.0280	(15)	0.1780	(8)	0.0150	(13)
	11.0390	(7)	0.1540	(6)	0.0260	(12)	0.1550	(7)	0.0140	(12)
	24.4250	(13)	0.3150	(13)	0.0190	(7)	0.3000	(15)	0.0060	(5)
	35.0090	(14)	0.3490	(15)	0.0160	(5)	0.2490	(12)	0.0060	(5)

SE

5761.3884	(1)	1.2112	(3)	0.0920	(15)	0.8061	(15)	0.0625	(15)
6332.7720	(15)	1.3187	(13)	0.0376	(11)	0.5678	(12)	0.0181	(7)
6235.5370	(13)	1.3102	(12)	0.0359	(9)	0.5611	(11)	0.0185	(8)
6145.2037	(11)	1.3463	(14)	0.0294	(3)	0.5202	(5)	0.0162	(1)
6102.2181	(9)	1.2421	(7)	0.0320	(6)	0.5579	(9)	0.0171	(5)
6075.0971	(8)	1.2202	(4)	0.0310	(4)	0.5678	(12)	0.0164	(2)
6129.3460	(10)	1.2974	(10)	0.0287	(2)	0.5326	(6)	0.0170	(3)
6057.4607	(7)	1.2373	(5)	0.0320	(6)	0.4875	(3)	0.0180	(6)
5985.9987	(4)	1.2823	(8)	0.0390	(12)	0.4474	(2)	0.0212	(12)
6253.7630	(14)	1.3671	(15)	0.0372	(10)	0.5590	(10)	0.0191	(11)
6040.5450	(6)	1.2379	(6)	0.0414	(13)	0.5789	(14)	0.0270	(13)
5995.4363	(5)	1.1992	(2)	0.0322	(8)	0.5559	(8)	0.0188	(9)
6155.0963	(12)	1.3026	(11)	0.0285	(1)	0.5396	(7)	0.0170	(3)
5936.9075	(3)	1.2863	(9)	0.0316	(5)	0.4918	(4)	0.0190	(10)
5815.0472	(2)	1.1129	(1)	0.0573	(14)	0.4464	(1)	0.0294	(14)

ankings are given in parentheses.

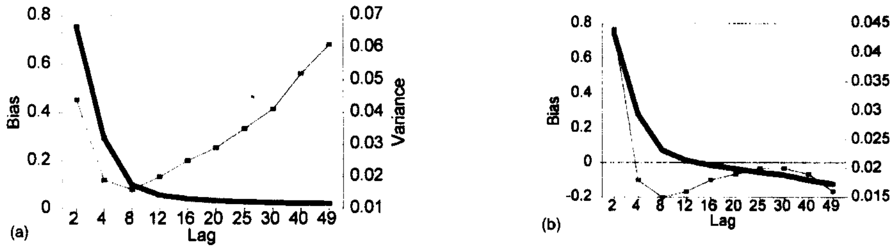


FIGURE 6. MA1(−0.5), Parzen (a): (a) known mean; (b) unknown mean.

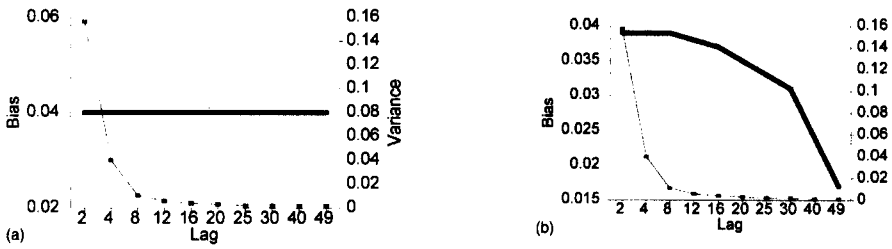


FIGURE 7. MA1(−1.0), trapezoid: (a) known mean; (b) unknown mean.

The finding that both the bias and the variance are lower with negative than with positive serial correlation is in accord with theory. This follows from the fact that $h(0) = B(1)/A(1)$. The closer is the moving-average coefficient to -1, the smaller is $B(1)$. The more negative is the autoregressive coefficient, the larger is $A(1)$. Negative serial correlation has the effect of reducing the absolute value of $h(0)$, and hence the absolute exact error induced by the estimators. Thus, for a moving-average component of the same size, a positive autoregressive component magnifies both the bias and the variance, while a negative autoregressive component reduces these quantities. Accordingly, the MSE of positively autocorrelated series is generally much larger than that of negatively autocorrelated series. This can be seen by comparing AR1(0.4), Figure 1(c) or 3, with AR1(−0.4), Figure 1(d) or 4. Similarly, a negative moving-average component reduces the MSE of the estimators. This is highlighted by the dramatically different scaling in the MSE of the two nearly integrated models depicted in Figures 1(a) and 1(b) for ARMA(0.9, 0.6) and ARMA(0.9, −0.5).

In Tables V and VI, we report the bias, variance and MSE and the rankings associated with these quantities for time series processes selected from each of the five groups. As expected, windows that tend to yield a large bias also tend to yield a low variance. Since the bias and variance are minimized at different bandwidths, the rankings of the MSE need not coincide with those for bias and variance. The rankings also depend on whether the mean has to be estimated. Although some windows might have the same ranking with both treatments of the sample mean, their implications in terms of bias and variance can be rather different. In general, the exact bias for Groups 1, 2 and 4 is larger, and the variance smaller, when the mean has to be estimated. The MSE is usually lower when the mean is known with positively correlated processes, but is lower when the mean is unknown with negatively correlated processes. However, when the truncated and trapezoid windows perform well in a mean-squared sense, it is unambiguously because of their small biases.

6. A RESPONSE SURFACE ANALYSIS

The objective of this section is to provide a concise framework for evaluating how the bias and variance vary with parameters such as sample size, bandwidth and coefficients underlying the time series models in finite samples. This is accomplished using a response surface analysis such as discussed in Hendry (1984). The analysis involves estimating, for each window, one equation for bias and one for variance.

Since the characteristic exponent r for most of the windows considered here is even,[4] the bias can be approximated in large samples by (2.5) under some conditions. Using this asymptotic result as basis, we then incorporate factors found in the previous section to influence the exact bias into the regression. These are functions of the value of the spectral density at the origin, the bandwidth, the sample size and how close is the sum of the autoregressive parameters to unity. Let a_i and b_i be the autoregressive and moving-average coefficients corresponding to an ARMA(p, q) process. After a specification search, we settle on the following equation for bias:

$$\text{bias}\,(h_T) = \alpha_0 \frac{h''(0)}{M_T^2} + \alpha_1 \frac{(1-\sum a_i)^{-2}}{M_T^3} + \alpha_2 \frac{h(0)^3}{T^3} + \alpha_3 \frac{h(0)^2/M_T}{T} + \alpha_4 \frac{h(0)}{T}. \tag{6.1}$$

The specification for variance is based on the asymptotic consideration in (2.6) that the variance of the estimator is proportional to $(M_T/T)h(0)^2$. As in the bias equation, we also allow other variables to enter the equation for variance. The best model (in terms of explanatory power) is found to be one which allows $h(0)$ to have a non-linear effect on the variance. The equation for variance is

$$\text{var}\,(\hat{h}_T) = \beta_0 \frac{h(0)^2 M_T}{T} + \beta_1 \frac{1}{T^2} + \beta_2 \frac{M_T}{T} + \beta_3 \frac{h(0)}{T} + \beta_4 \frac{h(0)^2}{T}. \tag{6.2}$$

The estimates $\hat{\alpha}_i$ and $\hat{\beta}_i$, $i = 0, \ldots, 4$, are obtained by ordinary least-squares regressions of (6.1) and (6.2). In cases when the mean of the series is known, the estimations are based on the 720 observations derived from the various combinations of sample sizes, bandwidths and data generating processes.

However, the estimated spectral density at frequency zero and its variance are attracted towards zero as M_T increases when the sample mean is unknown. To prevent such idiosyncrasies from dominating the response surface analysis, we only use observations with $M_T \leqslant T/2$ in the regressions. This leaves 504 observations.

Tables VII–X summarize the regression results. The $\bar{R}^2$s for the bias equations are higher when the mean is unknown, but the explanatory power of

TABLE VII
RESPONSE SURFACE ANALYSIS, SUMMARY OF THE BIAS EQUATIONS; KNOWN MEAN

Window	$h''(0)/M^2$	$(1-\sum a_i)^{-2}/M^3$	$h(0)/T^3$ (10^4)	$\{h(0)^2/M^2\}/T$	$h(0)/T$	R^2
1 Truncated periodogram	0.005	2.45	2.61	0.15	−26.06	0.88
2 Bartlett (a)	0.005	2.60	6.53	0.18	−51.93	0.90
3 Bartlett (b)	0.005	2.88	6.04	0.16†	−46.86	0.86
4 Parzen (a)	0.005	2.71	7.05	0.20	−56.14	0.89
5 Tukey–Hamming	0.005	2.77	5.45	0.19	−45.54	0.88
6 Tukey–Hanning	0.005	2.80	5.70	0.19	−47.23	0.88
7 Bohman	0.005	2.74	6.69	0.20	−53.82	0.88
8 Daniell	0.005	2.80	5.02	0.18	−42.60	0.87
9 Parzen (b)	0.005	2.80	4.45	0.18	−38.73	0.87
10 Bartlett (c)	0.005	2.73	5.87	0.15	−45.43	0.86
11 Parzen (c)	0.005	2.62	3.91	0.17	−34.96	0.88
12 Tukey–Parzen	0.005	2.85	6.09	0.19	−49.94	0.87
13 Normal	0.005	2.73	6.63	0.20	−53.42	0.88
14 Quadratic	0.005	2.79	4.91	0.18	−40.60	0.87
15 Trapezoid	0.005	2.82	3.64	0.17	−33.31	0.86

Note: All entries are statistically significant at the 5% level except those indicated by a dagger.

TABLE VIII
RESPONSE SURFACE ANALYSIS, SUMMARY OF THE BIAS EQUATIONS; UNKNOWN MEAN

Window	$h''(0)/M^2$	$(1-\sum a_i)^{-2}/M^3$	$h(0)/T^3$ (10^4)	$\{h(0)^2/M^2\}/T$	$h(0)/T$	R^2
1 Truncated periodogram	0.002	0.54	1.00	0.17	−78.68	0.97
2 Bartlett (a)	0.003	1.09	1.17	0.20	−87.88	0.98
3 Bartlett (b)	0.003	1.11	1.16	0.20	−88.03	0.98
4 Parzen (a)	0.003	1.25	1.19	0.21	−89.03	0.97
5 Tukey–Hamming	0.003	1.16	1.09	0.21	−83.82	0.98
6 Tukey–Hanning	0.003	1.22	1.10	0.21	−84.27	0.98
7 Bohman	0.003	1.25	1.16	0.21	−87.73	0.98
8 Daniell	0.003	1.14	1.06	0.21	−82.40	0.98
9 Parzen (b)	0.003	1.08	1.04	0.21	−80.90	0.98
10 Bartlett (c)	0.002	1.00	1.14	0.19	−86.02	0.98
11 Parzen (c)	0.002	0.84	1.04	0.19	−80.86	0.98
12 Tukey–Parzen	0.003	1.30	1.11	0.22	−84.98	0.97
13 Normal	0.003	1.23	1.16	0.21	−87.69	0.98
14 Quadratic	0.003	1.11	1.07	0.21	−81.66	0.98
15 Trapezoid	0.003	1.03	0.99	0.20	−78.09	0.98

Note: All entries are statistically significant at the 5% level.

TABLE IX

RESPONSE SURFACE ANALYSIS, SUMMARY OF THE VARIANCE EQUATIONS; KNOWN MEAN

Window	$h^2(0)M/T$	$1/T^2$ (10^6)	M/T^2 (10^4)	$h(0)/T$ (10^4)	$h(0)^2/T$ (10^5)	$h(0)/T^2$	R^2
1 Truncated periodogram	1.71	1.57	−7.91	1.13	−7.35	−5.44	0.96
2 Bartlett (a)	0.82	0.70	−3.67	0.37	−7.04	−1.67	0.98
3 Bartlett (b)	1.50	0.98	−5.43	0.30	−15.86	−1.08	0.97
4 Parzen (a)	0.68	0.67	−3.55	0.29	−7.33	−1.25	0.97
5 Tukey–Hamming	0.98	0.84	−4.42	0.44	−8.54	−1.98	0.98
6 Tukey–Hanning	0.93	0.82	−4.31	0.41	−8.67	−1.82	0.98
7 Bohman	0.74	0.71	−3.73	0.31	−7.68	−1.38	0.98
8 Daniell	1.08	0.91	−4.77	0.50	−9.04	−2.26	0.98
9 Parzen (b)	1.23	1.01	−5.29	0.59	−9.40	−2.73	0.98
10 Bartlett (c)	1.52	0.90	−5.14	0.30	−15.49	−1.01	0.98
11 Parzen (c)	1.36	1.14	−5.83	0.73	−8.05	−3.46	0.98
12 Tukey–Parzen	0.87	0.79	−4.19	0.37	−8.88	−1.61	0.98
13 Normal	0.75	0.70	−3.72	0.32	−7.59	−1.40	0.98
14 Quadratic	1.16	0.95	−4.99	0.55	−9.00	−2.54	0.98
15 Trapezoid	1.47	1.22	−6.33	0.74	−10.74	−3.41	0.98

TABLE X

RESPONSE SURFACE ANALYSIS, SUMMARY OF THE VARIANCE EQUATIONS; UNKNOWN MEAN

Window	$h^2(0)M/T$	$1/T^2$ (10^6)	M/T^2 (10^4)	$h(0)/T$ (10^4)	$h(0)^2/T$ (10^5)	$h(0)/T^2$	R^2
1 Truncated periodogram	0.39	0.58	−5.49	0.46	−1.68†	−2.34	0.83
2 Bartlett (a)	0.27	0.36	−3.53	0.16	−2.14	−0.79	0.91
3 Bartlett (b)	0.32	0.40	−3.95	0.18	−2.51	−0.86	0.91
4 Parzen (a)	0.27	0.37	−3.68	0.12	−2.41	−0.57	0.90
5 Tukey–Hamming	0.33	0.44	−4.33	0.19	−2.63	−0.94	0.91
6 Tukey–Hanning	0.34	0.44	−4.41	0.18	−2.78	−0.86	0.91
7 Bohman	0.29	0.39	−3.88	0.13	−2.52	−0.64	0.90
8 Daniell	0.36	0.46	−4.60	0.22	−2.77	−1.07	0.91
9 Parzen (b)	0.38	0.49	−4.83	0.26	−2.77	−1.28	0.90
10 Bartlett (c)	0.32	0.40	−3.98	0.19	−2.37	−0.92	0.92
11 Parzen (c)	0.35	0.47	−4.53	0.31	−2.02	−1.57	0.87
12 Tukey–Parzen	0.35	0.46	−4.62	0.16	−3.05	−0.75	0.91
13 Normal	0.28	0.38	−3.81	0.13	−2.45	−0.65	0.91
14 Quadratic	0.38	0.47	−4.63	0.24	−2.75	−1.21	0.90
15 Trapezoid	0.45	0.58	−5.68	0.32	−3.16	−1.58	0.90

Note: All entries are statistically significant at the 5% level except those indicated by a dagger.

the variance equations is higher when the mean is known. Based on standard errors corrected for heteroscedasticity,[5] most estimates are statistically significant at the 5% level. We therefore indicate coefficients that are insignificant instead.

The results in Tables VII and VIII indicate that although the quantitative effect depends on the treatment of the mean, all regressors have the same qualitative effect on the bias using any of the windows whether or not the mean has to be estimated. The estimated coefficients on $(1-\sum a_i)^{-2}/M^3$ confirm that bias increases significantly as the sum of the autoregressive coefficients approaches one. The autoregressive parameters have an additional effect on bias via $h(0)$, since $h(0)=B(1)^2/A(1)^2$. By contrast, the moving-average parameters affect bias only to the extent that they affect $h(0)$. It is to be noted that the main reason why (6.1) performs well with both treatments of the mean, even though the difference in the two biases is $O(M_T/T)$, is that the sample for the mean unknown case is restricted to values of $M_T < T/2$. Hence, we implicitly restrict the sample to those cases where the biases are least different in the known and unknown mean cases.

For the variance equations, the estimates reported in Tables IX and X suggest a non-linear relationship between the variance of the estimators on the one hand and the bandwidth and $h(0)$ on the other. Interestingly, the sum of the autoregressive coefficients found significant in the bias equations are not significant in the variance equations. The autoregressive and moving-average coefficients therefore have no statistically significant effect on variance beyond that through $h(0)$.

The coefficient on $h''(0)/M^2$ in the bias equations is always highly significant suggesting that the bias is larger the steeper is the spectral density function. The approximation for the asymptotic bias in (2.5) assumes conditions such as $M_T/T \rightarrow 0$ for large T, and on values of r and q. These conditions are violated in some of the cases considered, and could account for the discrepancies between the estimated coefficients and the asymptotic values reported in Priestley (1981).

The coefficient $\hat{\beta}_0$ in (6.2) should correspond to the asymptotic proportionality factor $2\int_0^1 k^*(\theta)^2\,d\theta$. Unlike the asymptotic approximation for bias, which depends on such parameters as r and q as stated in (2.5), the asymptotic approximation of variance permits a more meaningful comparison with the estimates $\hat{\beta}_0$. In general, the estimates are smaller than the asymptotic values tabulated in Priestley (1981).[6] To see how sensitive these results are to the sample size, we also computed the exact bias and variance of the estimators for $T=250$. Including these observations and dropping those corresponding to $T=50$ tends to increase $\hat{\beta}_0$ by 10%. Conversely, restricting the sample size to observations derived from $25 \leqslant T \leqslant 100$ reduces $\hat{\beta}_0$ to values much smaller than the tabulated asymptotic values. This suggests that asymptotic approximations should be used with caution when working with small sample sizes.

Analogous asymptotic values for α_0 and β_0 are not available for the mean unknown case. While $\hat{\alpha}_0$ and $\hat{\beta}_0$ are always significant and positive, they tend to be smaller than the estimates for the known mean case. This is particularly so for $\hat{\beta}_0$ of the variance equations. Care should also be taken in applying results valid when the mean is known to the case when the sample mean has to be estimated.

7. CONCLUSION

Our analysis reveals a marked difference in the finite sample bias and variance between estimators which have to estimate the sample mean from those which do not. While all windows produce exact errors of comparable size at optimal bandwidths, there is substantial variation in the performance of the estimators at suboptimal bandwidths. For most processes included in our simulations, small values of the bandwidth are adequate, but substantially larger values are necessary when the process is nearly integrated or when it has large negative moving-average errors. Asymptotic relationships which characterize the exact error are found to be poor guides for the sample sizes considered here. This has practical implications for plug-in rules which use the asymptotic values to guide the selection of the bandwidth.

APPENDIX: DERIVATION AND COMPUTATION OF VAR ($\tilde{h}_T$)

To derive the variance for $\tilde{h}_T$, consider the expectation of the cross products of the autocovariances:

$$E\{\tilde{R}_T(v)\tilde{R}_T(u)\} = (T-v)^{-1}(T-u)^{-1}\sum_{t=1}^{T-v}\sum_{s=1}^{T-u} E\{(X_t-\bar{X})(X_{t+v}-\bar{X})(X_s-\bar{X})(X_{s+u}-\bar{X})\}. \quad \text{(A1)}$$

It has been shown in Isserlis (1918) that for normally distributed variates A, B, C and D having zero means, $E(ABCD) = E(AB)E(CD) + E(AC)E(BD) + E(AD)E(BC)$. Substituting $(X_t - \bar{X})$, $(X_{t+v} - \bar{X})$, $(X_s - \bar{X})$ and $(X_{s+u} - \bar{X})$ for A, B, C and D, we have

$$\begin{aligned} E\{\tilde{R}_T(v)\tilde{R}_T(u)\} &= (T-v)^{-1}(T-u)^{-1}\sum_{t=1}^{T-v}\sum_{s=1}^{T-u}[E\{(X_t-\bar{X})(X_{t+v}-\bar{X})\}E\{(X_s-\bar{X})(X_{s+u}-\bar{X})\} \\ &\quad + E\{(X_t-\bar{X})(X_s-\bar{X})\}E\{(X_{t+v}-\bar{X})(X_{s+u}-\bar{X})\} \\ &\quad + E\{(X_t-\bar{X})(X_{s+u}-\bar{X})\}E\{(X_{t+v}-\bar{X})(X_s-\bar{X})\}]. \end{aligned} \quad \text{(A2)}$$

In view of (4.5), (4.6), (4.9) and (A2) we have

$$\begin{aligned} \operatorname{cov}\{\tilde{R}_T(v), \tilde{R}_T(u)\} &= (T-v)^{-1}(T-u)^{-1}\sum_{t=1}^{T-v}\sum_{s=1}^{T-u}[\{R(t-s)+\operatorname{var}(\bar{X})-\bar{R}(t)-\bar{R}(s)\} \\ &\quad \cdot\{R(t+v-s-u)+\operatorname{var}(\bar{X})-\bar{R}(t+v)-\bar{R}(s+u)\}+\{R(t-s-u) \\ &\quad +\operatorname{var}(\bar{X})-\bar{R}(t)-\bar{R}(s+u)\}\{R(t+v-s)+\operatorname{var}(\bar{X})-\bar{R}(t+v)-\bar{R}(s)\}] \\ &= (T-v)^{-1}(T-u)^{-1}\sum_{t=1}^{T-v}\sum_{s=1}^{T-u}[R(t-s)R(t+v-s-u) \\ &\quad + R(t-s-u)R(t+v-s)+2\operatorname{var}(\bar{X})^2+\operatorname{var}(\bar{X})\{R(t-s) \end{aligned}$$

$$+ R(t+v-s-u) + R(t-s-u) + R(t+v-s)\} - 2\operatorname{var}(\bar{X})\{\bar{R}(t+v)$$
$$+ \bar{R}(t) + \bar{R}(s) + \bar{R}(s+u)\} + 2\{\bar{R}(t)\bar{R}(t+v) + \bar{R}(s)\bar{R}(s+u)\}$$
$$+ \{\bar{R}(t+v) + \bar{R}(t)\}\{\bar{R}(s) + \bar{R}(s+u)\} - \bar{R}(t+v)\{R(t-s) + R(t-s-u)\}$$
$$- \bar{R}(t)\{R(t+v-s-u) + R(t+v-s)\} - \bar{R}(s+u)\{R(t-s) + R(t+v-s)\}$$
$$- \bar{R}(s)\{R(t+v-s-u) + R(t-s-u)\}]. \tag{A3}$$

It follows from (4.8), (A3) and the definition of $\operatorname{var}(\hat{h}_T)$ that

$$\begin{aligned}\operatorname{var}(\tilde{h}_T) = 4\sum_{u,\,v=0}^{T-1} (T-v)^{-1}(T-u)^{-1}\Bigg(\sum_{t=1}^{T-v}\sum_{s=1}^{T-u} & k_T(u)k_T(v) \times [R(t-s)R(t+v-s-u) \\ & + R(t-s-u)R(t+v-s) + 2\operatorname{var}(\bar{X})^2 + \operatorname{var}(\bar{X})\{R(t-s) \\ & + R(t+v-s-u) + R(t-s-u) + R(t+v-s)\} \\ & - 2\operatorname{var}(\bar{X})\{\bar{R}(t+v) + \bar{R}(t) + \bar{R}(s) + \bar{R}(s+u)\} + 2\{\bar{R}(t)\bar{R}(t+v) \\ & + \bar{R}(s)\bar{R}(s+u)\} + \{\bar{R}(t+v) + \bar{R}(t)\}\{\bar{R}(s) + \bar{R}(s+u)\} \\ & - \bar{R}(t+v)\{R(t-s) + R(t-s-u)\} + \bar{R}(t)\{R(t+v-s-u) \\ & + R(t+v-s)\} - \bar{R}(s+u)\{R(t-s) + R(t+v-s)\} \\ & + \bar{R}(s)\{R(t+v-s-u) + R(t-s-u)\}]\Bigg).\end{aligned} \tag{A4}$$

Equation (4.10) follows upon rearranging terms.

The expression (4.10) involves repeated calculations of many terms. To simplify the computations, we define

$$C(i,j) = \sum_{k=1}^{j} R(k-i). \tag{A5}$$

It is straightforward to show that

$$\sum_{s=1}^{T-\mu} R(t-s) = \begin{cases} T\bar{R}(t+u) - C(t+u,\,u) & \text{if } t+u \leqslant T \\ T\bar{R}(t) - C(t+u-T,\,u) & \text{otherwise} \end{cases} \tag{A6}$$

$$\sum_{s=1}^{T-\mu} R(t+v-s) = \begin{cases} T\bar{R}(t+u+v) - C(t+\mu+v,\,\mu) & \text{if } t+\mu+v \leqslant T \\ T\bar{R}(t+v) - C(t+\mu+v-T,\,\mu) & \text{otherwise} \end{cases} \tag{A7}$$

$$\sum_{s=1}^{T-\mu} R(t-s-u) = T\bar{R}(t) - C(t,\,u) \tag{A8}$$

$$\sum_{s=1}^{T-\mu} R(t+v-u-s) = T\bar{R}(t+v) - C(t+v,\,u). \tag{A9}$$

Interchanging s and μ by t and v gives expressions for $\sum_{t=1}^{T-\mu} R(t+v-s-u)$, $\sum_{t=1}^{T-\mu} R(t+v-s)$, $\sum_{t=1}^{T-\mu} R(t-s)$ and $\sum_{t=1}^{T-\mu} R(t-s-u)$. Using (A6)–(A9) reduces (4.10) from a double summation to two single summations.

All the calculations are performed using Turbo C Version 2 with supplementary numerical routines from Press *et al.* (1988). The numerical computations were very time consuming due to the multiple loops involved. For example, with $T = 100$, the task took over a day to be executed on a 486 66 Mhz. This precludes sample sizes greater than 250.

NOTES

1. A consistent estimate of σ_ε^2 can be obtained in the time domain using the sum of squared residuals from long autoregression estimated by ordinary least squares. One can also use a non-parametric method as suggested by Hannan and Nicholls (1977) to obtain strongly consistent and asymptotically normal estimates. Alternatively, one can taper the series in the frequency domain, a technique which Pukkila and Nyquist (1985) found to be particularly effective in reducing bias and variance when the series is near integrated.
2. These statistics are valid under quite general regularity conditions on the moments of the innovations. See Phillips (1987) and Phillips and Perron (1988).
3. The picture is basically similar at other sample sizes, except that with $T = 150$ the truncated window now performs better than the trapezoid window for processes in Group 1, and the Daniell window performs reasonably well for processes in Group 2. The ranking is broadly similar when the mean is known.
4. The exceptions are the Bartlett, truncated, quadratic and trapezoid windows.
5. The regressions are estimated using the 'ROBUST' option in RATS Version 4.20.
6. These are 4/3 for Bartlett (a), 1.08 for Parzen (a), 3/2 for Tukey–Hanning and 1.59 for Tukey–Hamming.

ACKNOWLEDGEMENTS

The authors acknowledge grants from the Social Science and Humanities Research Council of Canada (SSHRC) and the Fonds de la Formation de Chercheurs et l'Aide à la Recherche du Québec (FCAR). The second author would also like to thank the National Science and Engineering Research Council of Canada (NSERC) for financial support. Useful comments from an anonymous referee are also gratefully acknowledged.

REFERENCES

ANDREWS, D. W. K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–58.

BROCKWELL, P. J. and DAVIES, R. A. (1991) *Time Series: Theory and Methods*, 2nd Edn. New York: Springer.

CAMPBELL, J. Y. and MANKIW, N. G. (1987) Are output fluctuations transitory? *Q. J. Econ.* 102, 857–80.

CLARK, P. K. (1987) The cyclical component of U.S. economic activity. *Q. J. Econ.* 102, 798–814.

COCHRANE, J. H. (1988) How big is the random walk in GNP? *J. Polit. Econ.* 96, 893–920.

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

— and NICHOLLS, D. F. (1977) The estimation of the prediction error variance. *Am. Stat. Assoc.* 72, 834–40.

HENDRY, D. F. (1984) Monte Carlo experimentation in econometrics. In *Handbook of Econometrics*, Vol. II (eds Z. Griliches and M. D. Intriligator). Amsterdam: North-Holland, pp. 939–74.

ISSERLIS, L. (1918) On a formula for the product moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika* 12, 134–39.

NEAVE, H. R. (1971) The exact error in spectrum estimates. *Ann. Math. Statist.* 42, 961–75.

— (1972) A comparison of lag window generators. *Am. Statist. Assoc.* 67, 152–58.

NEWEY, W. K. and WEST, K. D. (1987) A simple, positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–8.

PARZEN, E. (1957) On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Statist.* 28, 329–48.

PHILLIPS, P. C. B. (1987) Time series regression with unit roots. *Econometrica* 55, 277–302.

— and PERRON, P. (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–46.

POLITIS, D. N. and ROMANO, J. P. (1995) Bias-corrected nonparametric spectral estimation. *J. Time Ser. Anal.* 16, 67–103.

PRESS, W., FLANNERY, B., TEUKOLSKY, S. and VETTERLING, W. (1988) *Numerical Recipes in C.* Cambridge: Cambridge University Press.

PRIESTLEY, M. B. (1981) *Spectral Analysis and Time Series*, 2 Vols. New York: Academic Press.

PUKKILA, T. and NYQUIST, H. (1985) On the frequency domain estimation of the innovation variance of a stationary univariate time series. *Biometrika* 72, 317–23.

WATSON, M. W. (1986) Univariate detrending methods with stochastic trends. *J. Monetary Econ.* 18, 49–76.