AN AUTOREGRESSIVE SPECTRAL DENSITY ESTIMATOR AT FREQUENCY ZERO FOR NONSTATIONARITY TESTS

PIERRE PERRON
Boston University
and
C.R.D.E.

SERENA NG
Boston College

Many unit root and cointegration tests require an estimate of the spectral density function at frequency zero of some process. Commonly used are kernel estimators based on weighted sums of autocovariances constructed using estimated residuals from an AR(1) regression. However, it is known that with substantially correlated errors, the OLS estimate of the AR(1) parameter is severely biased. In this paper, we first show that this least-squares bias induces a significant increase in the bias and mean-squared error (MSE) of kernel-based estimators. We then consider a variant of the autoregressive spectral density estimator that does not share these shortcomings because it bypasses the use of the estimate from the AR(1) regression. Simulations and local asymptotic analyses show its bias and MSE to be much smaller than those of a kernel-based estimator when there is strong negative serial correlation. We also include a discussion about the appropriate choice of the truncation lag.

1. INTRODUCTION

The statistical analysis of models with nonstationary variables has received considerable attention in the last decade, as seen from the many theoretical results that have been developed and the numerous applications that have been reported. It is by now common practice to report the outcome of some unit root test on each variable, perform tests for the presence of cointegration, and, using one of the many asymptotically optimal procedures, estimate the cointegrating vectors. For a review, see Campbell and Perron (1991), Stock (1994), and Watson (1994).

Consider for example the following simple relation between a scalar y_t and an m-dimensional vector x_t with all variables being $I(1)$:

$$y_t = \beta' x_t + v_t. \tag{1.1}$$

Of special interest is to test the null hypothesis of no cointegration. This is often done applying a unit root test to the estimated residuals, $\hat{v}_t = y_t - \hat{\beta}' x_t$, where $\hat{\beta}$ is the ordinary least squares (OLS) estimate of β. Note that (1.1) contains the univariate unit root problem as a special case with $\beta = 0$ and $\hat{v}_t = y_t$. This model is quite general because substantial heterogeneity and autocorrelation are permitted in the errors $\{v_t\}$ and the first differences of the data, Δx_t. Usually some kind of "mixing conditions" are imposed (see, e.g., Phillips and Perron, 1988), such that one can apply a functional central limit theorem to the partial sums of the errors.

The basis of many unit root tests is the following first-order autoregression:

$$\hat{v}_t = \alpha \hat{v}_{t-1} + u_t,$$

with the least-squares estimate denoted $\hat{\alpha}$ and $\hat{u}_t = \hat{v}_t - \hat{\alpha} \hat{v}_{t-1}$. It is by now well known that, under the null hypothesis of no cointegration (or the null hypothesis of a unit root when $\beta = 0$), the least-squares estimator, $\hat{\alpha}$, converges to 1 at the fast rate of T. However, the limiting distributions of $T(\hat{\alpha} - 1)$ and of its associated t-statistic depend on nuisance parameters arising from serial correlation in the errors $\{\Delta v_t\}$. A popular approach to remove this dependence of the asymptotic distribution on the nuisance parameters has been to apply some kind of transformation to the basic least-squares estimates. For the univariate case, early examples of transformed unit root tests are those of Phillips (1987) and Phillips and Perron (1988). In the multivariate case, transformed statistics were proposed by Phillips and Ouliaris (1990). To apply these transformations, consistent estimates of $\sigma^2_{\Delta v} = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E(\Delta v_t^2)$ and $\sigma^2 = \lim_{T \to \infty} T^{-1} E(S_T^2)$, with $S_T = \sum_{t=1}^{T} \Delta v_t$, are needed. These are usually based on the estimated residuals $\hat{u}_t$ because $\hat{\alpha}$ converges to 1 at the fast rate of T. For example, to estimate $\sigma^2_{\Delta v}$ one uses $s_u^2 = T^{-1} \sum_{t=1}^{T} \hat{u}_t^2$, and to estimate σ^2 the most popular method has been to use a kernel-based estimator of the form

$$s_{WA}^2 = T^{-1} \sum_{t=1}^{T} \hat{u}_t^2 + 2T^{-1} \sum_{j=1}^{(T-1)} k(j, M_T) \sum_{t=k+1}^{T} \hat{u}_t \hat{u}_{t-k}. \tag{1.2}$$

Here $k(j, M_T)$ is some kernel that weighs the sample autocovariances and M_T is a bandwidth that acts as a truncation lag parameter when $k(j, M_T) = 0$ for $|j| > M_T$. Although many variants of unit root and cointegration tests have been proposed, almost all use in some way such estimators to eliminate the effect of nuisance parameters on the asymptotic distribution.

It is important to note that the preceding estimators of $\sigma^2_{\Delta v}$ and σ^2 both depend on the properties of $\hat{\alpha}$ via the use of the estimated residuals $\hat{u}_t$. However, it has also been shown that the least-squares estimate of α is severely biased in samples of typical sizes (and remains so even in quite large samples) when there is sub-

stantial correlation in the errors. This feature has been extensively documented in Perron (1996) for the univariate case and is recently analyzed in Ng and Perron (1997a) for the multivariate case. This can explain the substantial size distortions of the tests in the presence of important serial correlation.

Given these biases in the least-squares estimates, one would like to construct cointegration and unit root tests that are affected as little as possible by the dependence on $\hat{\alpha}$. An obvious possibility is to use the residuals under the null hypothesis, i.e., $\Delta\hat{v}_t$ (or Δy_t in the univariate unit root problem). However, if both $\sigma_{\Delta v}^2$ and σ^2 are estimated using the residuals under the null hypothesis, it was shown by Phillips and Ouliaris (1990) that the tests become inconsistent. This result was more or less perceived as implying an impossibility to altogether avoid the use of the least-squares estimates $\hat{\alpha}$ in constructing estimates of the nuisance parameters.

The theme of this paper is that, on the contrary, it is possible to construct estimates of the nuisance parameters that are consistent under the null hypothesis and ensure consistent tests while avoiding any dependence on $\hat{\alpha}$. The idea is to use the residuals under the null hypothesis, $\Delta\hat{v}_t$, to construct s_u^2 and to use a particular formulation of the autoregressive spectral density estimator to estimate the spectral density at frequency zero of Δv_t. Such an estimator was first proposed by Stock (1990) and is defined by $s_{AR}^2 = s_{ek}^2/(1 - \hat{b}(1))^2$, where $\hat{b}(1) = \sum_{j=1}^k \hat{b}_j$ and $s_{ek}^2 = T^{-1}\sum_{t=k+1}^T \hat{e}_{tk}^2$ with $\hat{b}_j$ and $\hat{e}_{tk}$ obtained from the following autoregression:

$$\Delta\hat{v}_t = b_0\hat{v}_{t-1} + \sum_{j=1}^k b_j \Delta\hat{v}_{t-j} + e_{tk}.$$

It is the aim of this paper to analyze the properties of such an autoregressive spectral density estimator. Without much loss of generality we concentrate on the univariate case where $\beta = 0$ with $\hat{v}_t = y_t$, and the problem of interest is that of testing for a unit root. The focus is on the properties of the estimator when there is substantial correlation in the error process. We first show that such an estimator of the spectral density at frequency zero has much smaller biases and mean-squared errors (MSE) compared to a kernel-based estimator of the form (1.2) that is constructed using the least-squares residuals. These features are analyzed using both simulations and local asymptotic analyses where the errors are modeled as AR(1) or MA(1) processes with parameters approaching the boundaries -1 or $+1$ as the sample size increases. The qualitative results obtained extend immediately to the multivariate framework.

The plan of the paper is as follows. Section 2 motivates the analysis in terms of the Z_α test for the presence of a unit root. Section 3 discusses the data-generating processes used for the simulations and presents the results. Section 4 presents a framework to analyze the local asymptotic properties of the spectral density estimators. Section 5 summarizes the implications of the different estimators of the spectral density at frequency zero for the unit root tests. In particular, we discuss

how the use of the autoregressive spectral density estimator allows unit root tests that show little size distortions even in the presence of substantial serial correlation in the errors. Section 6 offers concluding comments. A technical contribution of this paper is to derive the limit of the autoregressive spectral density estimator in several local asymptotic frameworks. These proofs are contained in a mathematical appendix.

2. MOTIVATION

We motivate our analysis with the problem of testing for a unit root. We consider a series $\{y_t\}_{t=0}^T$ generated by

$$y_t = \alpha y_{t-1} + u_t, \tag{2.1}$$

with $\alpha = 1$ under the null hypothesis. The errors $\{u_t\}$ are assumed to be a linear process of the form $u_t = \sum_{i=0}^\infty b_i e_{t-i}$ with e_t being independent and identically distributed, $e_t \sim$ i.i.d. $(0, \sigma_e^2)$. Let $B(z) = 1 + \sum_{i=1}^\infty b_i z^i = 1/(1 + \sum_{i=1}^\infty a_i z^i) = 1/A(z)$. We further assume that $B(z)$ is nonzero on the unit circle, that $A(1) \neq 0$, and that $k^{1/2} \sum_{i=1}^\infty |a_{k+i}| \to 0$ for some increasing sequence k (note that the latter condition is automatically satisfied if u_t is a stationary and invertible autoregressive moving average (ARMA) process). Also, $y_0 = 0$, for simplicity.

We shall focus on the Z_α test developed in Phillips (1987) and extended in Phillips and Perron (1988). The test is defined as

$$Z_\alpha = T(\hat{\alpha} - 1) - (s^2 - s_u^2) \bigg/ \left(2T^{-2} \sum_{t=1}^T y_{t-1}^2 \right), \tag{2.2}$$

where $\hat{\alpha}$ is the OLS estimate of the autoregressive parameter in (2.1), $s_u^2 = T^{-1} \sum_{t=1}^T \hat{u}_t^2$, $\hat{u}_t = y_t - \hat{\alpha} y_{t-1}$, and s^2 is a consistent estimator of σ^2. The analysis can easily be extended to the case where additional deterministic components are included in the regression (2.1). The form of Z_α remains the same if y_{t-1} is replaced by its demeaned or detrended counterpart. See Ng and Perron (1997b) for a discussion of issues pertaining to detrending and the estimation of the spectral density function.

A consistent estimator of σ^2 often used is the nonparametric estimator, s_{WA}^2, defined in (1.2). Simulation results of Schwert (1989), DeJong, Nankervis, Savin, and Whiteman (1992), and Phillips and Perron (1988), among others, have shown that Z_α based upon s_{WA}^2 suffers from severe size distortions, especially when there is substantial negative correlation in the residuals u_t (see Haug, 1993, concerning cointegration tests). There is also evidence that the choice of the kernel and the methods to choose the truncation lag do not affect much the finite-sample properties of the test (see Kim and Schmidt, 1990).

The bad size properties of the test can be explained by the fact that $\hat{\alpha}$ is severely biased. In Perron (1996), it was shown that the finite-sample distribution of the normalized least-squares estimator, $T(\hat{\alpha} - 1)$, is very badly approximated by its

limiting distribution when there is substantial serial correlation in the residuals and the adequacy of the asymptotic approximation deteriorates when additional deterministic components are included in the regression. Indeed, the properties of $\hat{\alpha}$ effect the properties of Z_α not only in a direct way via $T(\hat{\alpha} - 1)$, but also in indirect ways via the estimates of the nuisance parameters s_{WA}^2 and s_u^2 because they are constructed using the least-squares residuals $\hat{u}_t$ and hence depend on the biased estimate $\hat{\alpha}$.

The size problem found in Z_α is not unique to the use of the kernel estimator. As an alternative to s_{WA}^2, consider the standard autoregressive spectral density estimator following the work of Berk (1974). It is defined by $s_B^2 = s_{ek}^2/(1 - \hat{b}(1))^2$, where $\hat{b}(1) = \sum_{j=1}^{k} \hat{b}_j$ and $s_{ek}^2 = T^{-1} \sum_{t=k+1}^{T} \hat{e}_{tk}^2$ with $\hat{b}_j$ and $\hat{e}_{tk}$ obtained from the following autoregression:

$$\hat{u}_t = \sum_{j=1}^{k} b_j \hat{u}_{t-j} + e_{tk}. \tag{2.3}$$

The estimator can be seen as a parametric autoregressive approximation of σ^2. Because $T(\hat{\alpha} - 1) = O_p(1)$, one can use the results of Berk (1974) to show that s_B^2 is consistent provided $k^3/T \to 0$ and $k \to \infty$ as $T \to \infty$. Simulations showed that unit root tests continue to have severely distorted sizes when based upon s_B^2. The problem is that the autoregressive spectral density estimator still depends on the estimated residuals, $\hat{u}_t$. Hence, a starting point to modify the Z_α test is to try to get rid of the dependence of the test statistic on $\hat{\alpha}$.

As a first step in eliminating this dependence, let us analyze the case where, instead of using $s_u^2 = T^{-1} \sum_{t=1}^{T} \hat{u}_t^2$ as an estimate of σ_u^2, we use the residuals under the null hypothesis, i.e., $s_{\Delta y}^2 = T^{-1} \sum_{t=1}^{T} \Delta y_t^2$. This estimator is obviously also consistent under the null hypothesis of a unit root. A little algebra shows that, in this case, Z_α can be written as

$$MZ_\alpha = (T^{-1} y_T^2 - s^2) \Big/ \left(2T^{-2} \sum_{t=1}^{T} y_{t-1}^2 \right), \tag{2.4}$$

which is exactly the modified unit root test proposed by Stock (1990) and further analyzed by Perron and Ng (1996). It is called a modified Z_α test because it can also be written as (see Ng and Perron, 1976b)

$$MZ_\alpha = Z_\alpha + (T/2)(\hat{\alpha} - 1)^2. \tag{2.5}$$

Because $\hat{\alpha}$ converges to 1 at rate T, the correction factor is asymptotically negligible and Z_α and MZ_α are asymptotically equivalent. However, when large negative serial correlation is present in the residuals and hence $\hat{\alpha}$ is severely biased, the correction factor $(T/2)(\hat{\alpha} - 1)^2$ can be important even in quite large samples.

The representation (2.4) is interesting in several aspects. First, it shows that using the residuals under the null hypothesis to construct a consistent estimate of σ_u^2 eliminates the dependence of the unit root test on $\hat{\alpha}$ were it not for the fact that

the spectral density estimate s^2 remains constructed using the least-squares residuals. Second, writing $MZ_\alpha = (y_T^2 - Ts^2)/(2T^{-1} \sum_{t=1}^{T} y_{t-1}^2)$, it is easy to see that a necessary condition for the test to be consistent against stationary alternatives is that Ts^2 diverges under such alternatives. This is important because it shows that we cannot construct s_{WA}^2 using the residuals under the null hypothesis if σ_u^2 is estimated using the same residuals, i.e., Δy_t. This is because $Ts_{WA}^2 = O_p(1)$ under stationary alternatives when constructed with Δy_t as shown by Phillips and Ouliaris (1990).

The challenge therefore is to construct an estimator s^2 that is consistent for σ^2 under the null hypothesis, with Ts^2 diverging under stationary alternatives, and is such that the estimator does not depend on $\hat{\alpha}$. This is achieved using a modified autoregressive spectral density estimator based on the first differences of the data. Such an estimator, which we denote by s_{AR}^2, is defined in the present context as

$$s_{AR}^2 = s_{ek}^2/(1 - \hat{b}(1))^2, \tag{2.6}$$

where $s_{ek}^2 = T^{-1} \sum_{t=k+1}^{T} \hat{e}_{tk}^2$, $\hat{b}(1) = \sum_{j=1}^{k} \hat{b}_j$, with $\hat{b}_j$ and $\{\hat{e}_{tk}\}$ obtained from the following autoregression estimated by OLS:

$$\Delta y_t = b_0 y_{t-1} + \sum_{j=1}^{k} b_j \Delta y_{t-j} + e_{tk}. \tag{2.7}$$

Under the conditions stated on the errors $\{u_t\}$, consistency of the parameter estimates in the preceding regression under the null hypothesis that y_t has a unit root follows from the results of Berk (1974), Said and Dickey (1984), and Ng and Perron (1995) provided the truncation lag is such that $k \to \infty$ and $k^3/T \to 0$ as $T \to \infty$. Consistency of s_{AR}^2 for σ^2 follows.

The preceding autoregressive spectral density estimator differs from (2.3) in two ways. First, it uses Δy_t instead of $\hat{u}_t$, and second, the lagged level y_{t-1} is included as a regressor. The introduction of the lagged level is of no importance under the null hypothesis of a unit root because $\hat{b}_0 \to 0$ at rate T. In other words, s_B^2 and s_{AR}^2 are asymptotically equivalent under the null hypothesis. The introduction of the lagged level, however, ensures the consistency of unit root tests by making the estimate bounded below by zero under stationary alternatives. This is an important property because the requirement that Ts^2 diverges is now satisfied. An overlooked advantage of s_{AR}^2 over s_B^2 is that the former based upon (2.7) is immune to potentially severe biases in $\hat{\alpha}$ caused by the presence of substantial correlation in the errors.

Given that s_{AR}^2 based upon (2.7) does not depend on $\hat{\alpha}$, it is likely to provide a better estimate of σ^2 than s_{WA}^2 in finite samples. The next sections of this paper establish that this is indeed the case.

3. THE EXPERIMENTAL DESIGN

In this section, we discuss the experimental design used to evaluate the relative properties of s_{WA}^2 and s_{AR}^2. We keep the design very simple to better highlight the

types of correlation that induce problems of inference. To that effect, we consider errors as being generated by simple AR(1) or MA(1) models. Hence, the data-generating processes are of the form

$$y_t = y_{t-1} + u_t, \tag{3.1}$$

where the initial condition is set to $y_0 = 0$ and the errors $\{u_t\}$ are generated by either of the following:

$$MA(1): u_t = e_t + \theta e_{t-1},$$
$$AR(1): u_t = \rho u_{t-1} + e_t, \tag{3.2}$$

with $u_0 = e_0 = 0$ and $e_t \sim$ i.i.d. $(0, \sigma_e^2)$. Note that, in this case, the true value of σ^2 is $\sigma_e^2(1 + \theta)^2$ and $\sigma_e^2/(1 - \rho)^2$ for moving average and autoregressive models, respectively. We present both simulation experiments and theoretical analyses based on these specifications.

In the simulations, we consider the case where the data are assumed to have an unknown mean. Correspondingly, s_{WA}^2 is constructed using residuals $\hat{u}_t$ obtained from the regression (2.1) with a constant included. Also, the regression used to construct the autoregressive spectral density estimator is

$$\Delta y_t = c + b_0 y_{t-1} + \sum_{j=1}^{k} b_j \Delta y_{t-j} + e_{tk}. \tag{3.3}$$

The aim of the simulation experiments is to quantify the bias and MSE of s_{WA}^2 and s_{AR}^2 for a range of values of θ and ρ. The emphasis of our discussion is on cases where there is substantial serial correlation in the errors. The innovations $\{e_t\}$ are generated as i.i.d. $N(0,1)$ random variables using the GASDEV function in Press, Teukolsky, Vetterling, and Flannery (1992). In all cases, 2,000 replications are used. Three sample sizes are considered, $T = 100$, 200, and 500.

3.1. Results for s_{WA}^2 with Estimated Residuals

We report results for the kernel-based estimator constructed as in (1.2) using the Parzen window. This is a kernel that operates with a truncation point. Although other kernels are possible, the choice of this kernel is with little loss of generality given that our focus is on processes for u_t with roots close to the boundary of unity. The Parzen kernel was found to produce estimates with relatively good finite-sample properties in Ng and Perron (1996). It is among the best windows that provide non-negative estimates by construction and for which the bandwidth acts as a truncation lag parameter. We also tried other windows such as the quadratic spectral advocated by Andrews (1991). The results are qualitatively similar.

Several methods to choose the truncation lag were considered. We analyzed the properties of the estimator using fixed truncation lags ranging from 1 to some maximal order $M_T(\max)$ that increases with the sample size. We set $M_T(\max) =$

6, 10, and 14 for $T = 100$, 200, and 500, respectively. For data-dependent selection rules, our base case is the asymptotically optimal data-dependent method suggested by Andrews (1991) using an AR(1) approximation. Another experiment that we tried was to calculate the optimal bandwidth using an ARMA(1,1) approximation with the true values of the parameters. This led to estimates with even worse properties, especially in the negative MA(1) case. The reason is that for such a process the optimal bandwidth is relatively large and for reasons explained subsequently, the properties of s_{WA}^2 deteriorate as the bandwidth increases because the estimated residuals are not good approximations to the true residuals. We also considered experiments using the prewhitening device suggested by Andrews and Monahan (1992). This produced significant improvements only for AR(1) errors with positive coefficients. However, there were neither significant improvements nor marked deteriorations in cases with large negative AR(1) or MA(1) coefficients. To conserve space, these results will not be reported (but are available on request). Readers will be reminded of the advantages of prewhitening where appropriate.

The results are presented in Tables 1 and 2 for the bias and MSE, respectively. Consider first the base case with i.i.d. errors. For a given sample size the bias decreases, as expected, as the truncation lag increases. For a fixed truncation lag, it also decreases rapidly as the sample size increases. The MSE eventually increases with the truncation lag, but it is relatively small in all cases and decreases rapidly as T increases, especially using an automatic bandwidth selection procedure. For models with positive moving average coefficients, the bias and MSE are large for small values of the truncation lag, but both decrease substantially as the truncation lag increases. However, for a given truncation lag, the error decreases less rapidly as the sample size increases than in the i.i.d. case.

Consider now the case with a large negative moving average coefficient. Here the bias and MSE initially decrease as the truncation lag increases but start increasing at larger lags. More importantly, for $\theta = -.8$ the bias and MSE barely decrease as T increases even when an automatic bandwidth selection procedure is used. The MSE is, in all cases, several orders of magnitude greater than in the i.i.d. case. Indeed, the bias and MSE at $\theta = -.8$ are very large in relative terms because the true value is $\sigma^2 = .04$. Although the bias and MSE diminish somewhat more quickly in cases of large negative autoregressive errors, the estimator still gives imprecise estimates of σ^2 given that the true values are small when ρ is negative. When the autoregressive coefficient is positive, both the bias and MSE decrease as the truncation lag increases but they are very large compared to other cases and again decrease only very slowly as T increases.

We also present, in the last column of Tables 1 and 2, the minimal value of the bias and MSE for each case over all possible integer valued bandwidths. While these figures correspond to the best case possible and cannot, in general, be attained in practice, they provide a useful benchmark for comparison with the bias and MSE of s_{AR}^2 later.

TABLE 1. Exact bias of s_{WA}^2 using the Parzen window[a]

		σ^2	$M_T = 1$	2	3	4	5	$M_T(\text{max})$	Auto	Min
$T = 100$, $M_T(\text{max}) = 6$	i.i.d.	1.00	−0.054	−0.051	−0.046	−0.040	−0.034	−0.029	−0.037	−0.014
MA = 0.80		3.24	−1.688	−1.321	−0.877	−0.650	−0.536	−0.476	−0.428	−0.426
	0.50	2.25	−1.053	−0.820	−0.536	−0.388	−0.311	−0.269	−0.220	−0.227
	0.20	1.44	−0.450	−0.358	−0.246	−0.188	−0.157	−0.140	−0.134	−0.125
	−0.20	0.64	0.341	0.262	0.170	0.130	0.116	0.115	0.137	0.115
	−0.50	0.25	0.857	0.705	0.531	0.465	0.455	0.470	0.457	0.455
	−0.80	0.04	1.071	1.003	0.942	0.957	1.014	1.088	0.939	0.942
AR = 0.80		25.00	−22.655	−21.765	−20.480	−19.368	−18.373	−17.504	−12.244	−12.286
	0.50	4.00	−2.746	−2.448	−2.042	−1.741	−1.512	−1.340	−0.903	−0.887
	0.20	1.56	−0.562	−0.465	−0.342	−0.267	−0.220	−0.192	−0.193	−0.157
	−0.20	0.69	0.296	0.213	0.121	0.088	0.082	0.086	0.093	0.082
	−0.50	0.44	0.791	0.544	0.294	0.251	0.273	0.309	0.260	0.251
	−0.80	0.31	1.925	1.372	0.913	1.031	1.257	1.484	1.148	0.913
$T = 200$, $M_T(\text{max}) = 10$	i.i.d.	1.00	−0.028	−0.026	−0.021	−0.017	−0.013	0.003	−0.015	0.000
MA = 0.80		3.24	−1.640	−1.256	−0.789	−0.545	−0.415	−0.240	−0.229	−0.233
	0.50	2.25	−1.032	−0.792	−0.500	−0.346	−0.264	−0.155	−0.149	−0.151
	0.20	1.44	−0.426	−0.328	−0.209	−0.145	−0.110	−0.057	−0.077	−0.052
	−0.20	0.64	0.367	0.279	0.174	0.122	0.100	0.098	0.113	0.088
	−0.50	0.25	0.926	0.735	0.509	0.408	0.372	0.447	0.368	0.366
	−0.80	0.04	1.207	1.084	0.958	0.947	0.999	1.480	0.953	0.947

AR = 0.80	25.00	−22.434	−21.431	−19.975	−18.696	−17.531	−13.284	−7.956	−8.011
0.50	4.00	−2.709	−2.395	−1.967	−1.647	−1.401	−0.807	−0.669	−0.612
0.20	1.56	−0.542	−0.441	−0.312	−0.231	−0.178	−0.082	−0.128	−0.067
−0.20	0.69	0.320	0.228	0.124	0.083	0.071	0.086	0.077	0.069
−0.50	0.44	0.825	0.542	0.248	0.178	0.178	0.281	0.184	0.178
−0.80	0.31	2.172	1.395	0.688	0.700	0.849	1.579	0.947	0.688
$T = 500$, $M_T(\max) = 14$ i.i.d.	1.00	−0.013	−0.011	−0.009	−0.007	−0.005	0.011	−0.007	0.001
MA = 0.80	3.24	−1.618	−1.225	−0.745	−0.490	−0.352	−0.121	−0.117	−0.118
0.50	2.25	−1.014	−0.769	−0.471	−0.312	−0.226	−0.080	−0.082	−0.073
0.20	1.44	−0.413	−0.314	−0.193	−0.127	−0.091	−0.025	−0.048	−0.013
−0.20	0.64	0.388	0.292	0.176	0.117	0.087	0.070	0.081	0.059
−0.50	0.25	0.967	0.743	0.475	0.342	0.279	0.308	0.245	0.239
−0.80	0.04	1.365	1.151	0.911	0.830	0.836	1.579	0.866	0.830
AR = 0.80	25.00	−22.305	−21.237	−19.681	−18.306	−17.043	−9.893	−4.900	−4.755
0.50	4.00	−2.683	−2.356	−1.908	−1.571	−1.309	−0.446	−0.386	−0.273
0.20	1.56	−0.533	−0.431	−0.300	−0.219	−0.166	−0.048	−0.094	−0.036
−0.20	0.69	0.341	0.239	0.123	0.073	0.053	0.055	0.049	0.041
−0.50	0.44	0.867	0.553	0.220	0.126	0.109	0.181	0.109	0.108
−0.80	0.31	2.327	1.378	0.472	0.383	0.451	1.116	0.583	0.383

[a]The column labeled Min gives the smallest value of the bias over all possible integer-valued bandwidths. Sometimes, the bias obtained using the automatic bandwidth selection procedure with an AR(1) approximation is slightly smaller (column Auto). This can occur because the automatic procedure selects a bandwidth that is not necessarily integer-valued.

TABLE 2. MSE of s^2_{WA} using the Parzen window[a]

		σ^2	$M_T = 1$	2	3	4	5	$M_T(\max)$	Auto	Min
$T = 100, M_T(\max) = 6$	i.i.d.	1.00	0.022	0.024	0.032	0.040	0.048	0.056	0.040	0.022
MA = 0.80		3.24	2.926	1.884	1.023	0.788	0.760	0.802	1.244	0.760
0.50		2.25	1.149	0.741	0.411	0.326	0.321	0.342	0.485	0.321
0.20		1.44	0.225	0.161	0.116	0.111	0.121	0.135	0.139	0.111
−0.20		0.64	0.137	0.086	0.047	0.038	0.037	0.040	0.041	0.037
−0.50		0.25	0.769	0.518	0.299	0.236	0.234	0.258	0.232	0.234
−0.80		0.04	1.194	1.038	0.911	0.945	1.067	1.236	0.908	0.911
AR = 0.80		25.00	5.13E2	4.74E2	4.21E2	3.78E2	3.43E2	3.14E2	2.17E2	2.02E2
0.50		4.00	7.599	6.104	4.396	3.390	2.792	2.454	2.695	2.203
0.20		1.56	0.340	0.252	0.176	0.153	0.152	0.160	0.177	0.152
−0.20		0.69	0.111	0.064	0.033	0.030	0.034	0.039	0.032	0.030
−0.50		0.44	0.677	0.320	0.101	0.084	0.105	0.136	0.093	0.084
−0.80		0.31	4.094	2.072	1.040	1.429	2.146	2.991	1.706	1.040
$T = 200, M_T(\max) = 10$	i.i.d.	1.00	0.010	0.011	0.015	0.019	0.023	0.045	0.019	0.010
MA = 0.80		3.24	2.730	1.651	0.755	0.486	0.416	0.558	0.712	0.413
0.50		2.25	1.086	0.664	0.315	0.212	0.188	0.264	0.283	0.188
0.20		1.44	0.193	0.125	0.072	0.060	0.062	0.102	0.077	0.060
−0.20		0.64	0.146	0.087	0.039	0.025	0.021	0.031	0.025	0.021
−0.50		0.25	0.876	0.550	0.267	0.177	0.153	0.250	0.151	0.153
−0.80		0.04	1.494	1.193	0.932	0.920	1.032	2.302	0.934	0.920

AR = 0.80	25.00	5.03E2	4.59E2	4.00E2	3.52E2	3.11E2	1.90E2	1.37E2	1.24E2
0.50	4.00	7.367	5.796	3.988	2.900	2.225	1.302	1.433	1.295
0.20	1.56	0.305	0.212	0.127	0.096	0.086	0.115	0.102	0.086
-0.20	0.69	0.114	0.061	0.025	0.018	0.018	0.032	0.018	0.018
-0.50	0.44	0.707	0.306	0.068	0.041	0.045	0.119	0.048	0.041
-0.80	0.31	4.992	2.052	0.566	0.663	0.996	3.487	1.214	0.556
$T = 500$, $M_T(\text{max}) = 14$ i.i.d.	1.00	0.004	0.004	0.006	0.008	0.010	0.026	0.007	0.004
MA = 0.80	3.24	2.634	1.530	0.610	0.320	0.228	0.321	0.362	0.198
0.50	2.25	1.038	0.608	0.250	0.138	0.103	0.147	0.138	0.094
0.20	1.44	0.175	0.105	0.048	0.031	0.027	0.055	0.034	0.027
-0.20	0.64	0.155	0.089	0.035	0.018	0.013	0.018	0.013	0.011
-0.50	0.25	0.943	0.557	0.228	0.121	0.083	0.130	0.068	0.067
-0.80	0.04	1.888	1.334	0.839	0.713	0.740	2.776	0.801	0.713
AR = 0.80	25.00	4.97E2	4.51E2	3.88E2	3.36E2	2.92E2	1.08E2	69.908	62.849
0.50	4.00	7.208	5.574	3.691	2.547	1.823	0.596	0.644	0.596
0.20	1.56	0.288	0.192	0.101	0.064	0.048	0.066	0.046	0.042
-0.20	0.69	0.121	0.061	0.019	0.010	0.008	0.018	0.009	0.008
-0.50	0.44	0.762	0.310	0.051	0.019	0.016	0.055	0.018	0.016
-0.80	0.31	5.544	1.941	0.250	0.199	0.287	1.864	0.496	0.199

[a]The column labeled Min gives the smallest value of the MSE over all possible integer-valued bandwidths. Sometimes, the bias obtained using the automatic bandwidth selection procedure with an AR(1) approximation is slightly smaller (column Auto). This can occur because the automatic procedure selects a bandwidth that is not necessarily integer valued.

3.2. Results for the Autoregressive Spectral Density Estimator, s_{AR}^2

This subsection discusses results pertaining to the behavior of s_{AR}^2 constructed using the augmented autoregression (3.3). For the construction of the autoregressive spectral density estimator, s_{AR}^2, the only nuisance parameter to determine is the order of the autoregression k. We first considered deterministic rules whereby k is a fixed value in the integer interval between 1 and $k$max. We used $k$max = 4, 8, and 14 for $T = 100$, 200, and 500, respectively. We also considered data-dependent rules whereby k is chosen according to statistical criteria. This includes (i) a general to specific recursive procedure using a 5 and a 10% t-test for the significance of the last lag (given the upper bound $k$max) and (ii) rules based on the AIC and the Schwartz information criteria. The t-test tends to select orders of truncation that are higher than information-based rules, with the order of truncation increasing with the significance level of the test. Thus, of the data-dependent rules considered, the Schwartz criterion produces the tightest model and the 10% t-test is the most liberal.

The results for the bias and MSE are presented in Tables 3 and 4, respectively. A notable property of s_{AR}^2 is that its bias and MSE decrease rapidly as the sample size increases. However, for $T = 100$ or smaller, the bias and variance of s_{AR}^2 can be large, especially at an overly liberal value of k.

For small sample sizes, for example, at $T = 100$, the results are sensitive to the choice of $k$max. For larger values of $k$max, the estimates exhibit occasional outliers that increase the bias and MSE substantially. This occurs because when k is large relative to the total sample the biases of the least-squares estimates $\hat{b}_i$ are such that $\sum_{i=1}^{k} \hat{b}_i$ is occasionally close to 1, causing a singularity in the denominator of s_{AR}^2. This problem is less severe when no constant is included in the autoregression but more so when a time trend is included. When the sample size is larger, say $T = 150$ or greater, this sensitivity to the choice of $k$max disappears.

Consider first the base case with i.i.d. errors. For low values of the truncation lag, the bias and MSE are small but increase substantially as the truncation lag increases. With positive moving average errors, the bias eventually decreases and the MSE increases as the truncation lag increases for a fixed T. However, the errors decrease noticeably as T increases. While the variations in performance across selection procedures reduce as the sample size increases, the tight Schwartz criterion tends to produce the smallest MSE with positive moving average models.

It is of interest to note that both the bias and the MSE display a clear oscillating pattern as the truncation lag varies from odd to even values. More precisely, bias and MSE are substantially smaller at even than at odd lags. This can be seen from the fact that the autoregressive representation of a moving average model has coefficients $(-\theta)^i$ at the ith lag. An even k always ensures that the calculation of $\sum_{i=1}^{k} \hat{b}_i$ is balanced, in the sense that the number of odd and even terms always match.

For negative moving average errors, several features are noteworthy. First, the bias and MSE of s_{AR}^2 both decrease as the truncation lag increases for a fixed T

when $\theta \leq -0.5$. Accordingly, a more liberal data-dependent method (e.g., the recursive t-test) produces in this case smaller MSE than one that tends to select a tight structure (e.g., Schwartz's criterion). Thus, while the MSE obtained using a recursive t-test is higher than that obtained using the Schwartz criterion when θ is positive, the reverse is true when θ is negative.

We now turn to cases of AR(1) errors. It is useful to note at the outset that because the true autoregressive order of u_t is one in all data-generating processes considered, any overparameterization of the autoregression will lead to increases in the MSE. Accordingly, it is easy to understand why a tight selection procedure such as that based on the Schwartz criterion might produce estimates that have the lowest MSE. This is indeed the case with $|\rho| < 0.8$, where we observed that bias and MSE increase as the truncation lag increases given a fixed T and they both decrease rapidly as T increases. When ρ is close to one, the bias and MSE of the estimator are large but diminish as T increases. When ρ is close to -1, bias and MSE increase somewhat as k increases with a given T but fall rapidly as T increases.

In view of the oscillating magnitude of the bias and MSE for the case of MA(1) errors with positive coefficient, the errors associated with data-dependent rules could be further reduced if the rules are specified to choose over a range of even-valued truncation lags. Simulations to that effect are presented in the last four columns of Tables 3 and 4. As can be seen, the bias and MSE are substantially reduced in the positive moving average case. Indeed, with a search restricted to even lags, the MSE with k selected using recursive t-tests on the significance of the last lag is decreased in all cases, even with AR(1) errors. For the data-dependent methods using the AIC and the Schwartz criteria, the MSE is reduced with MA(1) errors but slightly increased with AR(1) errors. The latter can be explained by the fact that the true order ($k = 1$) is outside the permissible range for k, being 2 to $k$max. In practical settings, one should expect a pure numerical advantage in using an even number of autoregressive lags in empirical work.

3.3. Comparison of s_{WA}^2 and s_{AR}^2

For the i.i.d. case, a tight selection criterion, such as the Schwartz's criterion, permits s_{AR}^2 to have as low a bias as s_{WA}^2 with the MSE remaining slightly higher. For positive moving average errors, the MSE of s_{AR}^2 at the optimal truncation lag for a given T is higher than the MSE of s_{WA}^2 similarly evaluated at an optimal truncation lag, especially at $T = 100$, though the difference is only marginal at larger sample sizes.

Significant differences between s_{WA}^2 and s_{AR}^2 surface when θ is negative, in particular if it is close to -1. The bias and MSE of s_{AR}^2 are dramatically smaller and decrease much more rapidly than s_{WA}^2 as T increases. For example, with $T = 200$ and $\theta = -0.8$, the smallest MSE for s_{WA}^2 (with a bandwidth of 4) is .920

TABLE 3. Bias of the autoregressive spectral density estimator, s_{AR}^2

		σ^2	$k = 1$	2	3	4	5
$T = 100$, $k$max $= 4$	i.i.d.	1.00	0.023	0.073	0.158	0.232	—
	MA $= 0.80$	3.24	1.654	−0.652	1.406	0.010	—
	0.50	2.25	0.786	−0.109	0.615	0.504	—
	0.20	1.44	0.136	0.083	0.240	0.360	—
	−0.20	0.64	0.092	0.067	0.114	0.167	—
	−0.50	0.25	0.332	0.168	0.128	0.114	—
	−0.80	0.04	0.673	0.435	0.334	0.261	—
	AR $= 0.80$	25.00	−3.328	−1.972	−0.001	3.638	—
	0.50	4.00	−0.044	0.195	0.631	0.964	—
	0.20	1.56	0.036	0.108	0.237	0.306	—
	−0.20	0.69	0.027	0.052	0.113	0.170	—
	−0.50	0.44	0.026	0.047	0.093	0.130	—
	−0.80	0.31	0.019	0.034	0.060	0.086	—
$T = 200$, $k$max $= 8$	i.i.d.	1.00	0.014	0.038	0.072	0.094	0.132
	MA $= 0.80$	3.24	1.597	−0.692	1.035	−0.186	1.027
	0.50	2.25	0.703	−0.211	0.329	0.131	0.355
	0.20	1.44	0.121	0.038	0.113	0.144	0.201
	−0.20	0.64	0.076	0.035	0.049	0.061	0.084
	−0.50	0.25	0.310	0.141	0.083	0.061	0.057
	−0.80	0.04	0.612	0.345	0.222	0.154	0.115
	AR $= 0.80$	25.00	−1.704	−1.050	−0.142	0.574	1.440
	0.50	4.00	−0.054	0.060	0.187	0.268	0.400
	0.20	1.56	0.017	0.057	0.125	0.155	0.213
	−0.20	0.69	0.014	0.031	0.059	0.071	0.097
	−0.50	0.44	0.009	0.020	0.035	0.047	0.061
	−0.80	0.31	0.007	0.013	0.024	0.032	0.044
$T = 500$, $k$max $= 14$	i.i.d.	1.00	0.004	0.014	0.030	0.035	0.047
	MA $= 0.80$	3.24	1.545	−0.726	0.794	−0.348	0.572
	0.50	2.25	0.670	−0.252	0.209	−0.010	0.131
	0.20	1.44	0.102	0.008	0.046	0.052	0.070
	−0.20	0.64	0.067	0.019	0.018	0.021	0.030
	−0.50	0.25	0.293	0.124	0.063	0.038	0.027
	−0.80	0.04	0.564	0.291	0.176	0.116	0.081
	AR $= 0.80$	25.00	−0.638	−0.328	0.038	0.243	0.473
	0.50	4.00	−0.004	0.037	0.090	0.116	0.168
	0.20	1.56	−0.003	0.011	0.039	0.052	0.071
	−0.20	0.64	0.003	0.008	0.016	0.023	0.032
	−0.50	0.25	0.003	0.005	0.011	0.016	0.019
	−0.80	0.04	0.002	0.005	0.010	0.014	0.017

	All lags				Even lags			
kmax	t-10	t-5	AIC	SW	t-10	t-5	AIC	SW
—	0.115	0.084	0.058	0.032	0.110	0.097	0.098	0.073
—	0.422	0.524	0.454	0.561	−0.680	−0.716	−0.109	−0.331
—	0.448	0.476	0.419	0.488	−0.081	−0.101	0.102	−0.052
—	0.232	0.194	0.156	0.135	0.156	0.120	0.158	0.096
—	0.122	0.114	0.093	0.088	0.098	0.088	0.088	0.068
—	0.175	0.211	0.200	0.265	0.156	0.160	0.153	0.166
—	0.351	0.406	0.412	0.534	0.362	0.384	0.357	0.413
—	−0.289	−1.016	−2.197	−3.106	−0.215	−0.454	−1.217	−1.846
—	0.382	0.184	0.114	−0.035	0.337	0.254	0.309	0.221
—	0.146	0.120	0.097	0.053	0.148	0.135	0.157	0.118
—	0.100	0.075	0.061	0.040	0.082	0.068	0.086	0.055
—	0.065	0.047	0.035	0.028	0.063	0.057	0.057	0.049
—	0.048	0.037	0.031	0.020	0.045	0.040	0.042	0.034
0.248	0.129	0.090	0.037	0.018	0.091	0.067	0.051	0.038
0.669	0.716	0.724	0.643	0.558	−0.403	−0.555	0.213	−0.173
0.661	0.374	0.292	0.181	0.189	−0.068	−0.138	0.015	−0.166
0.402	0.253	0.196	0.106	0.112	0.136	0.087	0.067	0.043
0.162	0.112	0.103	0.067	0.073	0.074	0.062	0.045	0.035
0.077	0.096	0.119	0.117	0.195	0.115	0.124	0.111	0.136
0.066	0.105	0.140	0.165	0.297	0.162	0.204	0.162	0.270
4.830	2.008	0.561	−0.969	−1.638	0.613	−0.072	−0.232	−0.976
0.949	0.442	0.250	0.026	−0.047	0.300	0.200	0.107	0.059
0.409	0.225	0.154	0.048	0.028	0.165	0.119	0.086	0.059
0.191	0.120	0.086	0.032	0.016	0.088	0.066	0.052	0.032
0.110	0.056	0.041	0.017	0.010	0.046	0.034	0.026	0.020
0.079	0.040	0.027	0.010	0.007	0.033	0.027	0.018	0.014
0.184	0.115	0.082	0.011	0.005	0.074	0.046	0.019	0.014
0.438	0.393	0.411	0.317	0.198	−0.185	−0.284	0.070	−0.211
0.409	0.260	0.200	0.080	−0.021	−0.034	−0.139	−0.047	−0.201
0.266	0.185	0.149	0.055	0.092	0.095	0.052	0.015	0.003
0.109	0.081	0.071	0.044	0.060	0.047	0.036	0.021	0.019
0.046	0.045	0.057	0.059	0.105	0.073	0.090	0.066	0.111
0.015	0.025	0.036	0.048	0.099	0.044	0.063	0.047	0.098
3.087	1.710	0.838	−0.236	−0.586	0.707	0.171	−0.164	−0.333
0.791	0.478	0.313	0.024	−0.002	0.310	0.213	0.054	0.037
0.278	0.153	0.092	0.011	−0.004	0.079	0.046	0.023	0.012
0.116	0.070	0.044	0.006	0.003	0.044	0.026	0.011	0.008
0.080	0.050	0.038	0.006	0.003	0.030	0.024	0.007	0.005
0.056	0.041	0.027	0.005	0.002	0.029	0.017	0.007	0.005

TABLE 4. MSE of the autoregressive spectral density estimator, s_{AR}^2

		σ^2	$k = 1$	2	3	4	5
$T = 100$, $kmax = 4$	i.i.d.	1.00	0.070	0.145	0.295	0.545	—
	MA = 0.80	3.24	5.514	1.584	8.556	3.649	—
	0.50	2.25	1.544	0.714	2.373	3.273	—
	0.20	1.44	0.242	0.339	0.776	1.475	—
	−0.20	0.64	0.037	0.056	0.113	0.334	—
	−0.50	0.25	0.125	0.041	0.037	0.045	—
	−0.80	0.04	0.483	0.229	0.197	0.179	—
	AR = 0.80	25.00	2.29E2	3.43E2	6.89E2	2.23E3	—
	0.50	4.00	2.638	4.236	8.210	13.592	—
	0.20	1.56	0.242	0.422	0.778	1.168	—
	−0.20	0.69	0.029	0.067	0.166	0.308	—
	−0.50	0.44	0.009	0.023	0.056	0.104	—
	−0.80	0.31	0.003	0.011	0.024	0.047	—
$T = 200$, $kmax = 8$	i.i.d.	1.00	0.032	0.059	0.096	0.141	0.196
	MA = 0.80	3.24	3.767	0.932	3.057	1.336	4.944
	0.50	2.25	0.919	0.329	0.804	0.785	1.369
	0.20	1.44	0.111	0.140	0.240	0.332	0.464
	−0.20	0.64	0.018	0.021	0.036	0.053	0.076
	−0.50	0.25	0.102	0.025	0.012	0.011	0.014
	−0.80	0.04	0.387	0.129	0.057	0.029	0.018
	AR = 0.80	25.00	1.27E2	1.51E2	1.90E2	2.34E2	2.80E2
	0.50	4.00	1.203	1.672	2.273	3.111	4.048
	0.20	1.56	0.104	0.166	0.273	0.372	0.510
	−0.20	0.69	0.012	0.026	0.045	0.068	0.101
	−0.50	0.44	0.004	0.009	0.016	0.025	0.035
	−0.80	0.31	0.001	0.004	0.007	0.011	0.018
$T = 500$, $kmax = 14$	i.i.d.	1.00	0.011	0.021	0.030	0.040	0.053
	MA = 0.80	3.24	2.849	0.693	1.251	0.529	1.206
	0.50	2.25	0.612	0.168	0.266	0.223	0.324
	0.20	1.44	0.045	0.047	0.072	0.094	0.115
	−0.20	0.64	0.010	0.008	0.012	0.017	0.022
	−0.50	0.25	0.088	0.017	0.006	0.004	0.003
	−0.80	0.04	0.322	0.087	0.032	0.014	0.007
	AR = 0.80	25.00	49.436	55.797	68.198	77.351	90.378
	0.50	4.00	0.468	0.616	0.784	0.916	1.090
	0.20	1.56	0.036	0.057	0.084	0.116	0.141
	−0.20	0.69	0.005	0.009	0.015	0.020	0.027
	−0.50	0.44	0.001	0.003	0.005	0.008	0.010
	−0.80	0.31	0.000	0.001	0.002	0.003	0.004

	All lags				Even lags			
*k*max	*t*-10	*t*-5	AIC	*SW*	*t*-10	*t*-5	AIC	*SW*
—	0.296	0.226	0.169	0.090	0.259	0.225	0.224	0.145
—	5.858	6.720	5.630	6.229	1.915	1.818	3.351	2.844
—	2.706	2.290	2.254	1.546	1.518	1.100	2.055	1.352
—	0.987	0.712	0.582	0.309	0.827	0.591	0.895	0.467
—	0.249	0.222	0.179	0.039	0.235	0.217	0.204	0.057
—	0.061	0.075	0.067	0.094	0.043	0.042	0.040	0.041
—	0.207	0.245	0.245	0.350	0.193	0.199	0.186	0.218
—	9.80E2	8.32E2	3.57E2	2.49E2	9.37E2	8.96E2	4.66E2	3.55E2
—	7.662	5.104	4.253	2.781	6.514	4.905	5.331	4.464
—	0.752	0.633	0.548	0.343	0.664	0.588	0.694	0.460
—	0.195	0.142	0.100	0.061	0.148	0.110	0.177	0.074
—	0.051	0.038	0.024	0.012	0.046	0.041	0.041	0.028
—	0.022	0.015	0.012	0.004	0.019	0.017	0.017	0.011
0.532	0.274	0.200	0.064	0.036	0.201	0.152	0.099	0.059
6.354	6.118	5.915	4.957	3.888	1.733	1.289	3.854	2.072
5.765	2.289	1.791	1.072	0.767	1.108	0.758	2.260	0.562
1.431	0.856	0.618	0.183	0.121	0.651	0.476	0.219	0.150
0.219	0.125	0.090	0.032	0.019	0.094	0.072	0.035	0.021
0.038	0.031	0.035	0.027	0.054	0.029	0.028	0.021	0.024
0.012	0.023	0.041	0.052	0.131	0.048	0.066	0.047	0.095
5.76E2	3.96E2	2.79E2	1.67E2	1.36E2	2.78E2	2.32E2	2.15E2	1.57E2
10.969	5.629	3.617	1.616	1.263	4.172	3.196	1.975	1.674
1.643	0.779	0.587	0.180	0.112	0.591	0.459	0.271	0.169
0.396	0.186	0.126	0.036	0.014	0.118	0.074	0.066	0.027
0.097	0.051	0.034	0.009	0.004	0.038	0.025	0.016	0.009
0.044	0.024	0.017	0.004	0.002	0.018	0.014	0.007	0.004
0.280	0.186	0.136	0.019	0.012	0.124	0.082	0.025	0.021
2.662	2.325	2.225	1.652	1.070	1.137	0.963	1.293	0.662
1.482	0.949	0.757	0.302	0.280	0.608	0.442	0.274	0.191
0.555	0.370	0.248	0.054	0.047	0.228	0.134	0.064	0.047
0.112	0.075	0.048	0.012	0.010	0.050	0.030	0.011	0.008
0.017	0.012	0.012	0.008	0.016	0.015	0.015	0.009	0.016
0.001	0.001	0.002	0.004	0.014	0.004	0.007	0.004	0.013
2.82E2	1.78E2	1.44E2	66.407	50.885	1.34E2	1.12E2	63.150	55.917
6.056	2.844	2.020	0.596	0.476	2.059	1.495	0.725	0.616
0.714	0.421	0.295	0.063	0.038	0.261	0.170	0.083	0.057
0.155	0.109	0.061	0.009	0.005	0.084	0.037	0.013	0.010
0.057	0.035	0.024	0.003	0.002	0.021	0.016	0.005	0.003
0.026	0.019	0.014	0.001	0.001	0.014	0.009	0.002	0.001

whereas the smallest MSE for s_{AR}^2 (with $k = 8$) is .012. The comparisons are even more dramatic with $T = 500$.

For AR(1) errors, with $|\rho| < 0.8$, a tight selection procedure such as those based on information criteria permits the MSE's to be as small as those obtained with s_{WA}^2 evaluated with an optimal selection procedure for the bandwidth. The statistic s_{AR}^2 is inferior to s_{WA}^2 only in the case of positive AR(1) errors when prewhitening is applied.

When ρ is close to the boundary of one, u_t is nearly nonstationary with a power spectrum that becomes unbounded when ρ approaches 1. It is then not surprising that neither s_{AR}^2 nor s_{WA}^2 produces satisfactory estimates of σ^2. Whereas the bias remains high compared to other cases, it is substantially smaller with s_{AR}^2 than with s_{WA}^2. Also, unlike the bias of s_{WA}^2, the bias of s_{AR}^2 decreases noticeably as T increases.

For ρ close to -1, the bias and MSE of s_{AR}^2 are dramatically smaller than those of s_{WA}^2. Indeed, even though a tight selection procedure produces the smallest MSE for s_{AR}^2, any data-dependent rule would have resulted in a dramatically smaller bias and MSE than s_{WA}^2 constructed using an optimal bandwidth. For example, with $T = 100$ and $\rho = -0.8$, the MSE of s_{AR}^2 using a data-dependent method to select k varies from .022 (t-10) to .004 (SW) whereas the smallest MSE of s_{WA}^2 (at $M_T = 3$) is 1.04. The differences remain as important for larger sample size.

Our simulation results therefore lead to the following general observations. In well-behaved cases ($\theta > -.5$ and $|\rho| < .8$), both s_{AR}^2 and s_{WA}^2 produce good estimates of σ^2, but s_{WA}^2 is somewhat better in a mean-squared sense if evaluated at the optimal bandwidth. In such cases, the choice of the optimal truncation lag appears to be more important than the choice between the two estimators. In the AR(1) case with a large value of ρ, both estimators have rather poor properties. However, s_{AR}^2 has noticeably smaller biases and slightly smaller MSE. In cases of large negative serial correlation (moving average or autoregressive), the properties of s_{AR}^2 are dramatically superior to those of s_{WA}^2 irrespective of the method to choose k. Therefore the cases in which the choice of the estimator for the spectral density at frequency zero matters are when $\theta \rightarrow -1$, and $|\rho| \rightarrow \pm 1$. In all three cases, s_{WA}^2 appears inferior.

4. THEORETICAL RESULTS

To analyze the behavior of the estimates from a theoretical perspective, we adopt the approach of Nabeya and Perron (1994), treating the moving average or autoregressive coefficients as local to the relevant boundaries. We consider a slight extension of the models specified by (3.1) and (3.2) with $\{y_t\}$ generated by the following nearly integrated model:

$$y_t = (1 + c/T)y_{t-1} + u_t. \tag{4.1}$$

The series has an autoregressive root local to unity with noncentrality parameter c. Under the null hypothesis of a unit root, $c = 0$. The advantage of this

generalization is that it allows deriving the local asymptotic power of unit root tests. Our results are used to that effect in Perron and Ng (1996). There are three relevant cases. The first is when the moving average coefficient is local to -1, in which case the process is described by

$$u_t = e_t + \theta_T e_{t-1},$$
$$\theta_T = -1 + \delta/\sqrt{T}. \tag{4.2}$$

Throughout, $\{e_t\}$ is assumed to be i.i.d. $(0, \sigma_e^2)$. This specifies that the moving average coefficient approaches -1 at rate $\sqrt{T}$. As T increases, the errors have a noninvertible moving average representation and $\{y_t\}$ is white noise. Hence, this model was labeled as a "nearly-integrated nearly white noise" process. The second case is when the autoregressive coefficient is local to $+1$, and the process is described by

$$u_t = \rho_T u_{t-1} + e_t,$$
$$\rho_T = 1 + \phi/T. \tag{4.3}$$

This specifies that the autoregressive coefficient approaches $+1$ at rate T. As T increases, the errors have a unit root, and $\{y_t\}$ has accordingly two unit roots. Hence, this model was labeled as a "nearly twice integrated" process. The third case is when the autoregressive coefficient is local to -1, and the process is described by

$$u_t = \rho_T u_{t-1} + e_t,$$
$$\rho_T = -(1 + \phi/T). \tag{4.4}$$

This specifies that the autoregressive coefficient approaches -1 at rate T. As T increases, the errors have a negative unit root and $y_t = y_{t-2} + e_t$, a process with a unit root at period 2. Hence, this model was labeled as a "nearly seasonally integrated" process.

All these specifications were found to be useful in providing good approximations to the finite-sample distribution of the least-squares estimator in an autoregression of order one. Our aim in characterizing the limits of s_{WA}^2 and s_{AR}^2 in these local frameworks is similarly to obtain better approximations and additional insights about their behavior when there is substantial serial correlation in the errors. We also summarize relevant results about the implied behavior of the unit root tests.

4.1. Local Asymptotic Properties of s_{WA}^2

In this section, we consider the limit of s_{WA}^2 for the case where M_T acts as a truncation lag. The results are stated in the following lemma.

LEMMA 4.1. *Let $\{y_t\}$ be generated by (4.1) and let s_{WA}^2 be constructed as in (1.2).*

(a) *Suppose that $\{u_t\}$ is generated by (4.2), then $M_T^{-1} s_{WA}^2$ is $O_p(1)$.*
(b) *Suppose that $\{u_t\}$ is generated by (4.3), then $(M_T T)^{-1} s_{WA}^2$ is $O_p(1)$.*
(c) *Suppose that $\{u_t\}$ is generated by (4.4), then $(M_T T)^{-1} s_{WA}^2$ is $O_p(1)$.*

This lemma is proved in Perron and Ng (1996) en route to explaining the properties of unit root tests that adopt a kernel estimate for σ^2. These asymptotic limits of s_{WA}^2 are, however, interesting in their own right. In all cases considered s_{WA}^2 is not only an inconsistent estimator of σ^2 but diverges as T increases (because M_T is required to increase as T increases). The rate of divergence is more severe in the autoregressive cases compared to the negative moving average case. These theoretical results are in accord with the simulations reported earlier, namely, that biases and MSE are large and do not decrease much as the sample size increases when the autoregressive coefficient is close to ± 1 or the moving average coefficient is close to -1.

The results of the preceding lemma hold irrespective of the choice of the kernel and are the reason for our earlier claim that the choice of the Parzen kernel in the simulations is without loss of generality. The choice of the kernel affects the $O_p(1)$ factors in the lemma but not the rate of divergence of the estimators. The unimportance of the choice of the kernel in these situations is corroborated by the empirical findings of Kim and Schmidt (1990). Whereas the choice of the kernel is of secondary importance for the issue considered here, the value of the truncation point M_T is of special importance because it dictates the rate of divergence of s_{WA}^2.[1]

4.2. Local Asymptotic Properties of the Autoregressive Spectral Density Estimator, s_{AR}^2

The regression used to construct the autoregressive spectral density estimator is evidently the same regression used to construct the unit root test of Said and Dickey (1984). However, the noise function of the three cases of interest each has, in the limit, a root on the unit circle. Hence, we cannot appeal to results in Said and Dickey (1984) to derive the limit of s_{AR}^2 in the local asymptotic frameworks. To that effect, we provide, in the Appendix, detailed proofs of the results stated in this section.

Consider first the case pertaining to a large negative moving average coefficient. Because $\sigma^2 = \sigma_e^2 (1 + \theta_T)^2$, we have that the limiting value is 0 because $\theta_T \to -1$ as $T \to \infty$. The next theorem shows s_{AR}^2 to be consistent in this case.

THEOREM 4.1. *Let $\{y_t\}$ be generated by (4.1) and (4.2). Let s_{AR}^2 be obtained by applying OLS to (2.7). Then $s_{AR}^2 \to 0$ provided $k \to \infty$ and $k/T \to 0$ as $T \to \infty$.*

Because y_t is a white noise process in the limit, Δy_t is overdifferenced. In spite of this, Theorem 4.1 shows that the augmented autoregression can still be used to construct a consistent estimate of σ^2. Because in the limit σ^2 is 0, all that is required is that $\hat{b}(1) \to \infty$ as $T \to \infty$, a result that follows if $k \to \infty$ and $k/T \to 0$ as $T \to \infty$. The consistency of s_{AR}^2 in this case is to be contrasted with the limit of

s_{WA}^2, which diverges. This explains why in the simulations, the bias and MSE are much smaller with s_{AR}^2 than with s_{WA}^2.

Consider now the case pertaining to a large positive autoregressive coefficient. Note that because $\sigma^2 = \sigma_e^2/(1 - \rho_T)^2$, we have that the limiting value satisfies $T^{-2}\sigma^2 \to \sigma_e^2/\phi^2$ as $T \to \infty$ given that $\rho_T = (1 + \phi/T)$.

THEOREM 4.2. *Let $\{y_t\}$ be generated by* (4.1) *and* (4.3). *Let s_{AR}^2 be obtained by applying* OLS *to* (2.7) *with $k \to \infty$ and $k = o(T^{1/3})$. Let $T(\hat{b}(1) - b(1)) \to \eta$, where η is a random variable defined in the Appendix (equation* (A.7)*). Then $T^{-2}s_{AR}^2 \to \sigma_e^2/(c + \phi + \eta)^2$.*

Note that in this case s_{AR}^2 is not consistent for the true value even under the null hypothesis of a unit root ($c = 0$). Furthermore, it converges to a random variable in the limit. Whereas u_t is a unit root process with noncentrality parameter ϕ, in an augmented autoregression in Δy_t it is a unit root process with noncentrality parameter $c + \phi + \eta$. Thus, the augmented autoregression will not, in general, identify u_t as a unit root process even when $\phi = 0$. This accounts for the relatively poor performance of s_{AR}^2 in models with positive residual autocorrelation. Recall, however, that $s_{WA}^2 = O_p(M_T T)$ so that it understates σ^2 because $M_T/T \to 0$ as $T \to \infty$. But, unlike s_{WA}^2, s_{AR}^2 is of the same order as the true value σ^2. Hence, our theoretical result indicates that we can still expect s_{AR}^2 to be a better estimator than s_{WA}^2, even though both are likely to have poor properties because neither is consistent.

Consider now the case pertaining to a large negative autoregressive coefficient. Note that because $\sigma^2 = \sigma_e^2/(1 - \rho_T)^2$, we have that the limiting value satisfies $\sigma^2 \to \sigma_e^2/4$ as $T \to \infty$ because $\rho_T = -(1 + \phi/T)$. The next theorem shows s_{AR}^2 to be consistent for σ^2 in this case.

THEOREM 4.3. *Let $\{y_t\}$ be generated by* (4.1) *and* (4.4). *Let s_{AR}^2 be obtained by applying* OLS *to* (2.7) *with $k \to \infty$ and $k = o(T^{1/3})$ as $T \to \infty$. Then $s_{AR}^2 \to \sigma_e^2/4$.*

In the standard framework, Δy_t is a stationary process when y_t is integrated of order one. For the data-generating process in question, y_t has a seasonal unit root of period two, and hence Δy_t remains nonstationary. Heuristically, consistency of s_{AR}^2 follows from the fact that all the variables in the augmented autoregression are $I(1)$. Although the number of regressors increases with the sample size, we show in the Appendix that consistency of the parameter estimates continues to hold as in a regression with a fixed number of $I(1)$ regressors. The consistency of s_{AR}^2 in this case is again to be contrasted with the limit of s_{WA}^2, which diverged.

5. IMPLICATIONS FOR UNIT ROOT TESTS

We now consider the implications of the local limits of s_{WA}^2 and s_{AR}^2 for unit root tests using the same local asymptotic frameworks. The spectral density estimator is, of course, not the only quantity that affects the properties of unit root tests. The

sample moments of other quantities also matter. The following two lemmas summarize the relevant asymptotic results derived in Perron and Ng (1996).

LEMMA 5.2. *Let* $\{y_t\}$ *be generated by* (4.1) *and let* s_{WA}^2 *be constructed as in* (1.2).

   (a) *Suppose that* $\{u_t\}$ *is generated by* (4.2), *then* $(M_T T)^{-1} Z_\alpha$ *and* $(M_T T)^{-1} MZ_\alpha$ *are* $O_p(1)$.
   (b) *Suppose that* $\{u_t\}$ *is generated by* (4.3), *then* Z_α *and* MZ_α *are* $O_p(1)$.
   (c) *Suppose that* $\{u_t\}$ *is generated by* (4.4), *then* $(M_T T)^{-1} Z_\alpha$ *and* $(M_T T)^{-1} MZ_\alpha$ *are* $O_p(1)$.

In all cases, the divergence of Z_α *and* MZ_α *is to* $-\infty$.

The implications for the unit root tests depend on the particular cases considered. With negative serial correlation, Z_α and MZ_α diverge to $-\infty$ at rate $(M_T T)$. If a statistic has a limiting distribution that diverges to $-\infty$ and critical values from a bounded distribution are used in hypothesis testing, the consequence will be large size distortions. This is essentially why size distortions are reported for Z_α. Even though such results are widely reported for the negative moving average case, the problem is important in the negative autoregressive case also. In such cases, the selection of M_T in unit root tests entails considerations beyond the usual bias-variance trade-off of s_{WA}^2 as analyzed in Andrews (1991), because increasing the truncation lag can aggravate size distortions in the tests. In the case of autoregressive errors with positive coefficients, Z_α and MZ_α remain bounded as T increases even though s_{WA}^2 diverges. Hence, smaller size distortions are expected. These results are consistent with the simulations reported in the preceding section.

We now consider the limit behavior of the same unit root tests when s_{AR}^2 is used as the spectral density estimator at frequency zero.

LEMMA 5.3. *Let* $\{y_t\}$ *be generated by* (4.1) *and let* s_{AR}^2 *be obtained by applying* OLS *to* (2.7) *as an estimator of* σ^2.

   (a)[2] *Suppose that* $\{u_t\}$ *is generated by* (4.2), *then* $Z_\alpha = O_p(T)$ *but* $MZ_\alpha = O_p(1)$.
   (b) *Suppose that* $\{u_t\}$ *is generated by* (4.3), *then* Z_α, MZ_α *are both* $O_p(1)$.
   (c) *Suppose that* $\{u_t\}$ *is generated by* (4.4), *then* $Z_\alpha = O_p(T)$ *but* $MZ_\alpha = O_p(1)$.

For the two cases of negative serial correlation ((a) and (c)), the implications are first that Z_α remains with large size distortions even if s_{AR}^2 is used instead of s_{WA}^2. This is because the bias in the least-squares estimator still affects Z_α directly via $\hat{\alpha}$ and indirectly via the least-squares residuals (when constructing s_u^2). However, the statistic MZ_α is now bounded in probability in the local asymptotic framework where the MA(1) or the AR(1) coefficient converges to -1 as $T \to \infty$.

The foregoing analysis suggests that, if we construct MZ_α using s_{AR}^2, we will essentially have a unit root test that does not have any dependence on $\hat{\alpha}$. Whereas the limiting distributions are different from those obtained using the standard asymptotic framework, we also found the standard asymptotic distribution to be a reasonable approximation to the finite-sample distribution of MZ_α. For this

reason, use of the standard asymptotic critical values yielded unit root tests with good size properties for all the parameters considered in the simulations. Details are contained in Perron and Ng (1996). The consequence is dramatic improvements in size properties over unit root tests that do have a dependence on $\hat{\alpha}$ (e.g., Z_α) in the problematic parameter space. To give an idea of the magnitude of the size improvement, consider the MA(1) case with $\theta = -0.8$ and $T = 100$. The size of MZ_α using s_{AR}^2 is .09, whereas the size of Z_α using s_{WA}^2 is .98, when the nominal size of the test is .05. Such contrasts in size remain in larger samples.

The preceding lemmas also indicate that to have unit root tests with good properties, simply replacing s_{WA}^2 by s_{AR}^2 will not be sufficient; we need to remove total dependence of the test statistic on $\hat{\alpha}$. As discussed in Perron and Ng (1996), there exist several other tests that also do not have a dependence on $\hat{\alpha}$; for example, a modified Sargan–Bhargava test suggested by Stock (1990) or a modification of the Z_t test developed in Phillips (1987). The autoregressive spectral density estimator discussed here can therefore be used in a rather broad range of applications.

6. CONCLUSIONS

This paper has considered estimating the nuisance parameter σ^2 in the context of unit root or cointegration tests. We have shown that a particular formulation of the autoregressive spectral density estimator can provide estimates far superior to the traditional kernel-based estimator constructed using least-squares residuals. The gains are important in cases of strong negative correlation, and there are little losses in accuracy in the other cases. When used in conjunction with tests that do not depend on $\hat{\alpha}$, it allows unit root or cointegration tests to have substantially improved size in the presence of strong serial correlation in the residuals. Also, this marked reduction in size distortions does not come at the expense of a reduction in power. The estimator is very easy to construct and requires basically only a standard autoregression estimated by OLS. For these reasons, we believe that this estimator is of substantial interest for applications.

An issue that remains unsolved is an optimal method to select the order k of the autoregression. The relative merits of data-dependent methods for selecting k are discussed in Ng and Perron (1995) in the context of testing for a unit root from an augmented regression such as (2.7). Whereas we advocated the use of a general to specific recursive procedure on the ground that it produces unit root tests with better finite-sample size and power, it does not follow that this procedure is better in the context of producing estimates s_{AR}^2 that have the smallest MSE. As seen from the results here, too large a $k$max can induce excessive variability in the estimates when the sample size is small. As well, a liberal selection rule is preferred with negative moving average errors, but a conservative rule is preferred with positive moving average errors, and with (finite order) autoregressive errors in general. This being said, it is not clear that the MSE of the spectral density estimators is the appropriate criterion for selecting k, because our ultimate objective is to test for the presence of a unit root and not to obtain an estimate of σ^2 that is as precise as possible. Clearly, the optimal lag length should depend on the

underlying data-generating process. Hence, an important avenue for future research is to devise optimal data-dependent rules for s_{AR}^2 that produce unit root tests with good size both when the root of the error process is away from the unit circle and when it is close to it. These issues are discussed in Ng and Perron (1997b).

Finally, it is important to note that the estimators considered here are clearly aimed at providing estimates of the nuisance parameters in the context of testing for unit roots or cointegration. In this case, the class of possible estimators is constrained by the requirement that the estimates be bounded (or at least converge to zero at a rate slower than T) under stationary alternatives. This is needed to ensure consistency of the tests. If one is interested solely in an estimate of (2π times) the spectral density function at frequency zero of some series, say Δy_t, then better estimates are available. Because one is no longer constrained to use the least-squares residuals to construct s_{WA}^2, the first-differences Δy_t can be used. Also, in the construction of the autoregressive spectral density estimator s_{AR}^2, one need not include the lagged level y_{t-1} in the autoregression (2.7). These alternative constructions not only ensure consistency of the estimators under stationary alternatives but also more efficient estimates when the level of the series contains a unit root.

NOTE

1. Lee and Phillips (1994) suggested an ARMA prewhitened long-run variance estimator that has better properties than standard kernel estimators and can reduce size distortions in Z_α.

2. Theorem 4.3 and the results in Perron and Ng (1996) are actually not sufficient to show that $MZ_\alpha = Op(1)$. What is required is that $Ts_{AR}^2 = Op(1)$. This holds if $k \to \infty$, with $k = cT^{1/2}$ for some constant c. In Ng and Perron (1997b), we showed that model selection procedures based on the AIC and a modified version of it, which works better in finite samples, imply a selected k that satisfies this requirement. Methods based on the BIC do not, however, satisfy the requirement.

REFERENCES

Andrews, D.W.K. (1991) Heteroskedastic and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–854.

Andrews, D.W.K. & J. Monahan (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953–966.

Berk, K.N. (1974) Consistent autoregressive spectral estimates. *Annals of Statistics* 2, 489–502.

Campbell, J.Y. & P. Perron (1991) Pitfalls and opportunities: What macroeconomists should know about unit roots. In O.J. Blanchard & S. Fisher (eds.), *NBER Macroeconomics Annual*, vol. 6, 141–201. Cambridge: MIT Press.

Chang, M.C. (1989) *Testing for Overdifferencing*. Ph.D. Dissertation, North Carolina State University.

Chang, M.C. & D.A. Dickey (1994) Recognizing overdifferenced time series. *Journal of Time Series Analysis* 15, 1–18.

DeJong, D.N., J.C. Nankervis, N.E. Savin, & C.H. Whiteman (1992) The power problems of unit root tests in time series with autoregressive errors. *Journal of Econometrics* 92, 323–343.

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

Golub, G.H. & C.F. van Loan (1984) *Matrix Computations.* Baltimore: John Hopkins University Press.

Haug, A.A. (1993) Residual based tests for cointegration: A Monte Carlo study of size distortions. *Economics Letters* 41, 345–351.

Kim, K. & P. Schmidt (1990) Some evidence on the accuracy of Phillips-Perron tests using alternative estimates of nuisance parameters. *Economics Letters* 34, 345–350.

Lee, C.C. & P.C.B. Phillips (1994) An Arma Prewhitened Long-Run Variance Estimator. Manuscript, Yale University.

Nabeya, S. & P. Perron(1994) Local asymptotic distributions related to the AR(1) model with dependent errors. *Journal of Econometrics* 62, 229–264.

Ng, S. & P. Perron (1995) Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association* 90, 268–281.

Ng, S. & P. Perron (1996) The exact error in estimating the spectral density at the origin. *Journal of Time Series Analysis*, 17, 379–408.

Ng, S. & P. Perron (1997a) Estimation and inference in nearly unbalanced nearly cointegrated systems. *Journal of Econometrics* 79, 53–81.

Ng, S. & P. Perron (1997b) Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power. Working paper 369, Boston College.

Perron, P. (1996) The adequacy of asymptotic approximations in the near-integrated autoregressive model with dependent errors. *Journal of Econometrics* 70, 317–350.

Perron, P. & S. Ng (1996) Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *Review of Economic Studies* 63, 435–465.

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

Phillips, P.C.B. & S. Ouliaris (1990) Asymptotic properties of residual based tests for cointegration. *Econometrica* 58, 165–193.

Phillips, P.C.B. & P. Perron (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–346.

Press, W.H., S. Teukolsky, W. Vetterling, & B. Flannery (1992) *Numerical Recipes in C*, 2nd ed. Cambridge: Cambridge University Press.

Said, S.E. & D.A. Dickey (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599–607.

Schwert, G.W. (1989) Tests for unit roots, a Monte Carlo investigation. *Journal of Business and Economic Statistics* 7, 147–160.

Stock, J.H. (1990) A Class of Tests for Integration and Cointegration. Manuscript, Harvard University.

Stock, J.H. (1994) Unit roots and structural breaks. In R.F. Engle & D. McFadden (eds.), *Handbook of Econometrics*, vol. IV, ch. 46. Amsterdam: Elsevier.

Watson, M.W. (1994) Vector autoregressions and cointegration. In R.F. Engle & D. McFadden (eds.), *Handbook of Econometrics*, vol. IV, ch. 47. Amsterdam: Elsevier.

APPENDIX: MATHEMATICAL RESULTS

The following regression equation estimated by OLS is considered throughout this appendix:

$$\Delta y_t = b_0 y_{t-1} + \sum_{i=1}^{k} b_i \Delta y_{t-i} + e_{tk}. \tag{A.1}$$

We denote the OLS estimates by $\{\hat{b}_i\}_{i=0}^{k}$ and the estimated variance of the residuals as $s_{ek}^2 = T^{-1} \sum_{t=k+1}^{T} \hat{e}_{tk}^2$. As a matter of notation, we let $\Rightarrow$ denote weak convergence in distribution and $\rightarrow$ convergence in probability. Also, $W(r)$ is the unit Wiener process defined on $C[0,1]$, and $J_c(r) = \int_0^r \exp((r-s)c)\,dW(s)$ is the Ornstein–Uhlenbeck process with drift param-

eter c. The norm of a matrix B is defined by $\|B\| = \sup\{\|Bx\| : \|x\| < 1\}$, where $\|x\| = (x'x)^{1/2}$ is the standard euclidian norm of a column vector x. We also let C denote an arbitrary constant that is not necessarily the same throughout.

Proof of Theorem 4.1. We need to show $\sum_{i=1}^{k} b_i \to \infty$, that $s_{ek}^2 = T^{-1}\sum_{t=k+1}^{T} \hat{e}_{tk}^2$ is bounded as $T \to \infty$, and hence that $s_{AR}^2 = s_{ek}^2/(1 - \sum_{i=1}^{k} \hat{b}_i)^2 \to 0$ with $k/T \to 0$ and $k \to \infty$ as $T \to \infty$. It is useful first to note the following representation derived in Nabeya and Perron (1994) for $\{y_t\}$ generated by (4.1) and (4.2). Define $X_t = (1 + c/T)X_{t-1} + e_t$, $a_T = (1 - \delta/\sqrt{T})(1 - c/T)$, $b_T = 1 - (1 - c/T)(1 - \delta/\sqrt{T})$ with $a_T \to 1$ and $T^{1/2}b_T \to \delta$ as $T \to \infty$, then

$$y_t = a_T e_t + b_T X_t + o_p(T^{-1/2}), \tag{A.2}$$

where the presence of the $o_p(T^{-1/2})$ term is due to the fact that we specify $\alpha = 1 + c/T$ instead of $\alpha = \exp(c/T)$ as in Nabeya and Perron (1994).

Some of the arguments that follow are similar to those in Chang (1989) and Chang and Dickey (1994). We define the following vectors of dimension $(k + 1)$:

$$U_t' = (y_{t-1}, Z_t'), \quad \text{with } Z_t' = (\Delta y_{t-1}, \dots, \Delta y_{t-k}),$$

$$\hat{V}_T = T^{-1}\sum_{t=k+1}^{T} U_t \Delta y_t,$$

$$V = \sigma_e^2(-1, -1, 0, \dots, 0)',$$

and the following $(k + 1)$ by $(k + 1)$ matrices:

$$\hat{R}_T = T^{-1}\sum_{t=k+1}^{T} U_t U_t' = T^{-1}\begin{bmatrix} \sum_{t=k+1}^{T} y_{t-1}^2 & \sum_{t=k+1}^{T} y_{t-1}Z_t' \\ \sum_{t=k+1}^{T} y_{t-1}Z_t & \sum_{t=k+1}^{T} Z_t Z_t' \end{bmatrix},$$

$$\tilde{R}_T = \sigma_e^2\begin{bmatrix} \sigma_e^{-2}T^{-1}\sum_{t=k+1}^{T}(a_T^2 e_{t-1}^2 + b_T^2 X_{t-1}^2) & 1 & & & \\ 1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix},$$

$$R = \sigma_e^2\begin{bmatrix} 1 + \delta^2\int_0^1 J_c(r)^2\, dr & 1 & & & \\ 1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}.$$

Using this notation, we have $\hat{\beta} = \hat{R}_T^{-1}\hat{V}_T$, where $\hat{\beta} = (\hat{b}_0, \hat{b}_1, \dots, \hat{b}_k)$. We first state a few results that are useful for subsequent derivations.

LEMMA A.1. *Let $X_t = (1 + c/T)X_{t-1} + e_t$, then* (a) $|E(X_t e_s)| \leq C$; (b) $|E(X_t X_s)| \leq TC$ *for $t, s, \leq T$;* (c) $E(\sum_{t=1}^T X_t e_t)^2 = O(T^2)$; (d) $\sum_{t=1}^T X_{t-1} e_t = O_p(T)$; *and* (e) $T^{-2} \sum_{t=1}^T X_t^2 \Rightarrow \sigma_e^2 \int_0^1 J_c(r)^2 \, dr$.

Parts (a) through (d) are straightforward generalizations of results in Fuller (1976); part (e) is a standard result for near-integrated processes. The following lemma collects some useful results derived in Nabeya and Perron (1994).

LEMMA A.2. *Let $\{y_t\}$ be generated according to (4.1) and (4.2) and let $e_\infty = \lim_{T \to \infty} e_T/\sigma_e$. Then as $T \to \infty$,* (a) $T^{-1} \sum_{t=1}^T y_{t-1}^2 \Rightarrow \sigma_e^2 + \sigma_e^2 \delta^2 \int_0^1 J_c(r)^2 \, dr$; (b) $T^{-1} \sum_{t=1}^T y_{t-1} u_t \Rightarrow -\sigma_e^2$; (c) $y_T \Rightarrow \sigma_e e_\infty + \sigma_e \delta J_c(1)$; *and* (d) $T^{-1} \sum_{t=1}^T u_t^2 \Rightarrow 2\sigma_e^2$.

Using Lemma A.2, it is straightforward to derive the convergence results stated in the following lemma.

LEMMA A.3. *Let $\{y_t\}$ be generated by (4.1) and (4.2). Then for $i, j = 1, \ldots, k$,*

(a)

$$T^{-1} \sum_{t=k+1}^T y_{t-1} u_{t-i} \Rightarrow \begin{cases} \sigma_e^2 & \text{if } i = 1, \\ 0 & \text{otherwise;} \end{cases}$$

(b)

$$T^{-1} \sum_{t=k+1}^T y_{t-j} y_{t-i} \Rightarrow \begin{cases} \sigma_e^2 \left(1 + \delta^2 \int_0^1 J_c(r)^2 \, dr \right) & \text{if } i = j, \\ \sigma_e^2 \delta^2 \int_0^1 J_c(r)^2 \, dr & \text{if } i \neq j; \end{cases}$$

(c)

$$T^{-1} \sum_{t=k+1}^T u_{t-i} u_{t-j} \Rightarrow \begin{cases} 2\sigma_e^2 & \text{if } i = j, \\ -\sigma_e^2 & \text{if } |i - j| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

The following bounds can also be derived using Lemma A.1.

LEMMA A.4. *Let $\{y_t\}$ be generated according to (4.1) and (4.2):* (a) $|E[y_t y_s]| \leq C$ *if $s = t$ or $s = t + 1$, and $|E[y_t y_s]| \leq CT^{-1/2}$ otherwise;* (b) $E[u_t u_s] = 0$ *for $|t - s| > 1$ and $|E[u_t u_s]| \leq C$ otherwise;* (c) $E[\sum_{t=k+1}^T y_t u_{t-j}]^2 \leq CT^2$ *if $j = -1$ and $E[\sum_{t=k+1}^T y_t u_{t-j}]^2 \leq CT$ if $j \neq -1$;* (d) $E[\sum_{t=k+1}^T u_t u_{t-j}]^2 \leq CT$ *for $|j| > 1$;* (e) $E[\sum_{t=k+1}^T y_t y_{t-j}]^2 \leq CT^2$ *for any j.*

We now consider the limiting behavior of the moment matrix $\hat{R}_T$.

LEMMA A.5. (a) $\|\hat{R}_T - R\| \to 0$; (b) $\|\hat{R}_T - \tilde{R}_T\| = O_p(k/T^{1/2})$.

Proof. To prove (a), we show that each element of $\hat{R}_T$ converges in distribution to the corresponding element of R. Consider first the $(1,1)$ element of $\hat{R}_T$. Using Lemma A.2, we have

$$T^{-1} \sum_{t=k+1}^{T} y_{t-1}^2 \Rightarrow \sigma_e^2 \left(1 + \delta^2 \int_0^1 J_c(r)^2 \, dr \right).$$

For the remaining elements in the first row of $\hat{R}_T$, using (4.2) and parts (a) and (b) of Lemma A.3, we have

$$T^{-1} \sum_{t=k+1}^{T} y_{t-1} \Delta y_{t-1} = T^{-1} \sum_{t=k+1}^{T} y_{t-1} \left(u_{t-1} + \frac{c}{T} y_{t-2} \right) \to \sigma_e^2,$$

$$T^{-1} \sum_{t=k+1}^{T} y_{t-1} \Delta y_{t-i} = T^{-1} \sum_{t=k+1}^{T} y_{t-1} \left(u_{t-i} + \frac{c}{T} y_{t-i-1} \right) \to 0$$

for $i = 2, \ldots, k$. Consider now the elements of the lower right $k \times k$ matrix of $\hat{R}_T$. We have, for $i, j = 1, \ldots, k$,

$$T^{-1} \sum_{t=k+1}^{T} \Delta y_{t-i} \Delta y_{t-j} = T^{-1} \sum_{t=k+1}^{T} \left(\frac{c}{T} y_{t-i-1} + u_{t-i} \right) \left(\frac{c}{T} y_{t-j-1} + u_{t-j} \right)$$

$$= T^{-1} \sum_{t=k+1}^{T} u_{t-i} u_{t-j} + c^2 T^{-3} \sum_{t=k+1}^{T} y_{t-i-1} y_{t-j-1}$$

$$+ cT^{-2} \sum_{t=k+1}^{T} y_{t-i-1} u_{t-j} + cT^{-2} \sum_{t=k+1}^{T} y_{t-j-1} u_{t-i}. \tag{A.3}$$

The last three terms converge to zero using parts (a) and (b) of Lemma A.3. Thus, by Lemma A.3(c), $T^{-1} \sum_{t=k+1}^{T} \Delta y_{t-i} \Delta y_{t-j} \Rightarrow 2\sigma_e^2$ if $i = j$, $-\sigma_e^2$ if $|i - j| = 1$, and 0 otherwise. This proves part (a).

For part (b), define the matrix $Q = \hat{R}_T - \tilde{R}_T$. We show that each element, $q_{ij}(i, j = 1, \ldots, k+1)$, is such that $TE(q_{ij}^2) \le C$, for some constant C. Consider first the (1,1) element. We have

$$T^{-1} \sum_{t=k+1}^{T} y_{t-1}^2 = T^{-1} \sum_{t=k+1}^{T} (a_T e_{t-1} + b_T X_{t-1})^2$$

$$= T^{-1} \sum_{t=k+1}^{T} (a_T^2 e_{t-1}^2 + b_T^2 X_{t-1}^2 + 2a_T b_T X_{t-1} e_{t-1})$$

and

$$TE(q_{11}^2) = 4a_T^2 T b_T^2 T^{-2} E \left[\sum_{t=k+1}^{T} X_{t-1} e_{t-1} \right]^2 \le C,$$

using (c) of Lemma A.1 and the fact that $a_T = O(1)$ and $Tb_T^2 = O(1)$. Consider next the (1,2) element. We have

$$TE(q_{12}^2) = E \left[T^{-1/2} \sum_{t=k+1}^{T} (y_{t-1} \Delta y_{t-1} - \sigma_e^2) \right]^2$$

$$\le E \left[T^{-1/2} \sum_{t=k+1}^{T} (y_{t-1} u_{t-1} - \sigma_e^2) \right]^2 + E \left[cT^{-3/2} \sum_{t=k+1}^{T} y_{t-1} y_{t-2} \right]^2$$

$$+ E \left[2cT^{-3/2} \sum_{t=k+1}^{T} y_{t-1} y_{t-2} (y_{t-1} u_{t-1} - \sigma_e^2) \right].$$

The second term after the inequality is $o(1)$ using Lemma A.4(e). Consider the first term:

$$E\left[T^{-1/2}\sum_{t=k+1}^{T}(y_{t-1}u_{t-1}-\sigma_e^2)\right]^2$$

$$\leq E\left[T^{-1/2}\sum_{t=k+1}^{T}(a_Te_{t-1}^2-\sigma_e^2)\right]^2+E\left[T^{1/2}b_TT^{-1}\sum_{t=k+1}^{T}X_{t-1}e_{t-1}\right]^2$$

$$+E\left[a_T\theta_TT^{-1/2}\sum_{t=k+1}^{T}e_{t-1}e_{t-2}\right]^2+E\left[\theta_TT^{1/2}b_TT^{-1}\sum_{t=k+1}^{T}X_{t-1}e_{t-2}\right]^2$$

$$\leq C,$$

using Lemma A.1(c), the fact that $a_T\to 1$, and standard arguments for i.i.d. random variables. For $i=2,\ldots,k$, we have

$$TE(q_{1,i+1}^2)=E\left[T^{-1/2}\sum_{t=k+1}^{T}y_{t-1}\Delta y_{t-i}\right]^2$$

$$=E\left[T^{-1/2}\sum_{t=k+1}^{T}y_{t-1}u_{t-i}+cT^{-3/2}\sum_{t=k+1}^{T}y_{t-1}y_{t-i-1}\right]^2$$

$$\leq E\left[T^{-1/2}\sum_{t=k+1}^{T}y_{t-1}u_{t-i}\right]^2+E\left[cT^{-3/2}\sum_{t=k+1}^{T}y_{t-1}y_{t-i-1}\right]^2$$

$$\leq C,$$

using (c) and (e) of Lemma A.4. Consider now the elements of the lower right submatrix of Q. We have, using (A.3), for $i=1,\ldots,k$,

$$q_{i+1,i+1}=T^{-1}\sum_{t=k+1}^{T}(u_{t-i}^2-2\sigma_e^2)+c^2T^{-3}\sum_{t=k+1}^{T}y_{t-i-1}^2+2cT^{-2}\sum_{t=k+1}^{T}y_{t-i-1}u_{t-i}$$

$$TE(q_{i+1,i+1}^2)\leq E\left[T^{-1/2}\sum_{t=k+1}^{T}(u_{t-i}^2-2\sigma_e^2)\right]^2+E\left[c^2T^{-5/2}\sum_{t=k+1}^{T}y_{t-i-1}^2\right]^2$$

$$+2E\left[cT^{-3/2}\sum_{t=k+1}^{T}y_{t-i-1}u_{t-i}\right]^2$$

$$\leq C,$$

using Lemma A.4 and the fact that

$$E\left[T^{-1/2}\sum_{t=k+1}^{T}(u_{t-i}^2-2\sigma_e^2)\right]^2$$

$$=E\left[T^{-1/2}\sum_{t=k+1}^{T}\{(e_{t-i}^2-\sigma_e^2)+(\theta_T^2e_{t-i-1}^2-\sigma_e^2)+2\theta_Te_{t-i}e_{t-i-1}\}\right]^2$$

$$\leq E\left[T^{-1/2}\sum_{t=k+1}^{T}(e_{t-i}^2-\sigma_e^2)\right]^2+E\left[T^{-1/2}\sum_{t=k+1}^{T}(\theta_T^2e_{t-i-1}^2-\sigma_e^2)\right]^2$$

$$+E\left[2\theta_TT^{-1/2}\sum_{t=k+1}^{T}e_{t-i}e_{t-i-1}\right]^2$$

$$\leq C,$$

where the last inequalities follow using standard arguments and the fact that $|\theta_T| \le 1$. For the elements with $|i - j| = 1$, we have

$$TE(q_{i+1,j+1}^2) \le E\left[T^{-1/2} \sum_{t=k+1}^{T} (u_{t-i}u_{t-j} + \sigma_e^2)\right]^2 + E\left[c^2 T^{-5/2} \sum_{t=k+1}^{T} y_{t-i-1} y_{t-j-1}\right]^2$$

$$+ E\left[cT^{-3/2} \sum_{t=k+1}^{T} y_{t-i-1} u_{t-j}\right]^2 + E\left[cT^{-3/2} \sum_{t=k+1}^{T} y_{t-j-1} u_{t-i}\right]^2$$

$$\le C,$$

and for $|i - j| > 1$,

$$TE(q_{i+1,j+1}^2) \le E\left[T^{-1/2} \sum_{t=k+1}^{T} u_{t-i}u_{t-j}\right]^2 + E\left[c^2 T^{-5/2} \sum_{t=k+1}^{T} y_{t-i-1} y_{t-j-1}\right]^2$$

$$+ E\left[cT^{-3/2} \sum_{t=k+1}^{T} y_{t-i-1} u_{t-j}\right]^2 + E\left[cT^{-3/2} \sum_{t=k+1}^{T} y_{t-j-1} u_{t-i}\right]^2$$

$$\le C.$$

Hence, we have $E(\|Q\|^2) \le C(k + 1)^2/T$ because Q is a matrix of dimension $(k + 1) \times (k + 1)$. ∎

We consider now results pertaining to the vector $\hat{V}_T$ stated in the following lemma.

LEMMA A.6. (a) $\|\hat{V}_T\| = O_p(k^{1/2})$; (b) $\|\hat{V}_T - V\| = O_p(k/T)^{1/2}$.

Proof. We start by showing that each element of $\hat{V}_T$ converges to the corresponding element of V. Consider the first element. From Lemma A.2:

$$T^{-1} \sum_{t=k+1}^{T} \Delta y_t \, y_{t-1} = cT^{-2} \sum_{t=k+1}^{T} y_{t-1}^2 + T^{-1} \sum_{t=k+1}^{T} y_{t-1} u_t \to -\sigma_e^2.$$

For the remaining elements, we have, using Lemma A.3,

$$T^{-1} \sum_{t=k+1}^{T} \Delta y_t \, \Delta y_{t-i} = T^{-1} \sum_{t=k+1}^{T} \left(\frac{c}{T} y_{t-1} + u_t\right)\left(\frac{c}{T} y_{t-i-1} + u_{t-i}\right)$$

$$= c^2 T^{-3} \sum_{t=k+1}^{T} y_{t-1} y_{t-i-1} + cT^{-2} \sum_{t=k+1}^{T} y_{t-1} u_{t-i}$$

$$+ cT^{-2} \sum_{t=k+1}^{T} y_{t-i-1} u_t + T^{-1} \sum_{t=k+1}^{T} u_t u_{t-i}$$

$$\to -\sigma_e^2 \quad \text{if } i = 1 \quad \text{and} \quad 0 \text{ otherwise.}$$

This proves part (a) because each of the $(k + 1)$ elements of the vector $\hat{V}_T$ is $O_p(1)$.

Let $v_i (i = 1, \ldots, k+1)$ be the ith element of the vector $\hat{V}_T - V$. To prove part (b), we show that $TE(v_i^2)$ is bounded. Consider the last $(k-1)$ elements. Using the preceding result and Lemma A.4, we have, for $2 \leq i \leq k$,

$$E\left(T^{-1/2} \sum_{t=k+1}^{T} \Delta y_t \Delta y_{t-i}\right)^2 \leq E\left(T^{-1/2} \sum_{t=k+1}^{T} u_t u_{t-i}\right)^2 + o(1) \leq C.$$

Consider now the first element of the vector $T^{1/2}(\hat{V}_T - V)$. We have

$$TE\left(T^{-1} \sum_{t=k+1}^{T} \Delta y_t \, y_{t-1} + \sigma_e^2\right)^2$$

$$= TE\left(cT^{-2} \sum_{t=k+1}^{T} y_{t-1}^2 + T^{-1} \sum_{t=k+1}^{T} y_{t-1} u_t + \sigma_e^2\right)^2$$

$$= TE\left(T^{-1} \sum_{t=k+1}^{T} y_{t-1} u_t + \sigma_e^2\right)^2 + o(1)$$

$$= E\left[T^{-1/2} a_T \sum_{t=k+1}^{T} e_t e_{t-1} + T^{-1/2} \sum_{t=k+1}^{T} (a_T \theta_T e_{t-1}^2 + \sigma_e^2)\right.$$

$$\left. + T^{1/2} b_T T^{-1} \sum_{t=k+1}^{T} X_{t-1} u_t\right]^2 + o(1)$$

$$\leq E\left[T^{-1/2} a_T \sum_{t=k+1}^{T} e_t e_{t-1}\right]^2 + E\left[T^{-1/2} a_T \sum_{t=k+1}^{T} (\sigma_e^2 - e_{t-1}^2)\right]^2$$

$$+ T b_T^2 E\left[T^{-1} \sum_{t=k+1}^{T} X_{t-1} u_t\right]^2 + o(1)$$

$$\leq C,$$

because $\theta_T \to -1$, $T b_T^2 \to \delta^2$, and $a_T \to 1$, and using Lemmas A.4 and A.1(c) and the fact that $T^{-1/2} \sum_{t=k+1}^{T} (e_{t-1}^2 - \sigma_e^2)$ is bounded in probability. For the second element, similar arguments show that

$$TE(v_2^2) = T^{1/2} E\left[T^{-1} \sum_{t=k+1}^{T} \Delta y_t \Delta y_{t-1} + \sigma_e^2\right]^2 \leq C,$$

and part (b) follows.   ∎

The next lemma concerns the inverse of the moment matrix $\hat{R}_T$.

LEMMA A.7. $\|\tilde{R}_T^{-1}\| = O_p(k^2)$ and if $k^{11}/T \to 0$ as $T \to \infty$, then $k^{1/2}\|\hat{R}_T^{-1} - \tilde{R}_T^{-1}\| \to 0$.

Proof. Note that $\tilde{R}_T = A + E$ where

$$A = \sigma_e^2 \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix},$$

$$E = \sigma_e^2 \begin{bmatrix} \sigma_e^{-2}T^{-1}\sum_{t=k+1}^{T}(a_T^2 e_{t-1}^2 + b_T^2 X_{t-1}^2) - 2 & 2 & & & & \\ 2 & & 0 & 0 & & \\ 0 & & 0 & 0 & 0 & \\ & & & 0 & 0 & 0 \\ & & & & 0 & 0 \\ & & & & & 0 \end{bmatrix}.$$

We have $\lambda_{\min}(E) = 0$ and $\lambda_{\min}(A) = 2\sigma_e^2(1 - \cos(\pi/k + 2)) = O_p(k^2)$ (see Dickey, 1990). By Corollary 8.1–3 of Golub and van Loan (1984), $\lambda_{\min}(A + E) \geq \lambda_{\min}(A) + \lambda_{\min}(E)$, because A and E are symmetric. This implies $\lambda_{\min}(\tilde{R}_T^{-1}) = \|\tilde{R}_T^{-1}\| = O_p(k^2)$ because the maximal eigenvalue of $\tilde{R}_T^{-1}$ is the reciprocal of the minimal eigenvalue of $\tilde{R}_T$.

For part (b), we follow developments similar to those in Said and Dickey's (1984) Theorem 4.1. Let $q = \|\hat{R}_T^{-1} - \tilde{R}_T^{-1}\|$ and $Q = \hat{R}_T - \tilde{R}_T$; we have

$$q = \|\hat{R}_T^{-1}(\tilde{R}_T - \hat{R}_T)\tilde{R}_T^{-1}\| \leq \|\hat{R}_T^{-1}\|\|\tilde{R}_T - \hat{R}_T\|\|\tilde{R}_T^{-1}\| \leq (q + \|\tilde{R}_T^{-1}\|)\|Q\|\|\tilde{R}_T^{-1}\|.$$

Upon rearrangement, we have

$$k^{1/2}q \leq k^{1/2}(1 - \|\tilde{R}_T^{-1}\|\|Q\|)^{-1}\|\tilde{R}_T^{-1}\|^2\|Q\|.$$

Note that because $E(\|Q\|^2) \leq C(k + 1)^2/(T - k)$, we have $k^{9/2}\|Q\| \to 0$ if $k^{11}/T \to 0$. Hence,

$$k^{1/2}q \leq (1 - k^{-5/2}k^{-2}\|\tilde{R}_T^{-1}\|k^{9/2}\|Q\|)^{-1}k^{-4}\|\tilde{R}_T^{-1}\|^2 k^{9/2}\|Q\| \to 0 \quad \text{as } T \to \infty,$$

which proves Lemma A.7. ∎

We are now in a position to prove the following result.

LEMMA A.8. *Suppose $k \to \infty$ and $k^{11}/T \to 0$ as $T \to \infty$, then $\|\hat{\beta} - \tilde{\beta}\| \to 0$, where $\tilde{\beta} = \tilde{R}_T^{-1}V$. Also, because $T^{-1}\sum_{t=k+1}^{T}(a_T^2 e_{t-1}^2 + b_T^2 X_{t-1}^2) \Rightarrow \sigma_e^2(1 + \delta^2\int_0^1 J_c(r)^2\, dr)$, we have $\|\hat{\beta} - \beta\| \to 0$, where $\beta = R^{-1}V$.*

Proof.

$$\|\hat{\beta} - \tilde{\beta}\| = \|\hat{R}_T^{-1}\hat{V}_T - \tilde{R}_T^{-1}V\| = \|\hat{R}_T^{-1}\hat{V}_T - \tilde{R}_T^{-1}\hat{V}_T + \tilde{R}_T^{-1}\hat{V}_T - \tilde{R}_T^{-1}V\|$$

$$\leq \|\hat{R}_T^{-1} - \tilde{R}_T^{-1}\|\|\hat{V}_T\| + \|\tilde{R}_T^{-1}\|\|\hat{V}_T - V\|$$

$$= k^{1/2}\|\hat{R}_T^{-1} - \tilde{R}_T^{-1}\|k^{-1/2}\|\hat{V}_T\| + k^{5/2}T^{-1/2}k^{-2}\|\tilde{R}_T^{-1}\|T^{1/2}k^{-1/2}\|\hat{V}_T - V\| \to 0$$

using Lemmas A.6 and A.7 and the fact that $k^{11}/T \to 0$. ∎

Because $\sum_{i=1}^{k} \hat{b}_i$ is a continuous function of the vector $\hat{\beta} = (\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_k)$, the limit of $\sum_{i=1}^{k} \hat{b}_i$ is the same as that of $\sum_{i=1}^{k} \beta_i$, where $\beta = (\beta_0, \beta_1, \ldots, \beta_k)$ is the solution to $\beta = R^{-1}V$. We have used *MAPLE* to verify that $\beta_i = (k - i + 1)S/(1 + (k + 1)S)$ ($i = 1, \ldots, k$), where $S = \sigma_e^2 \delta^2 \int_0^1 J_c(r)^2 \, dr$. Hence, $\sum_{i=1}^{k} \beta_i = k(k + 1)S/(2(1 + (k + 1)S))$ and $k^{-1} \sum_{i=1}^{k} \beta_i \to \frac{1}{2}$ as $k \to \infty$. This shows that $\sum_{i=1}^{k} \hat{b}_i \to \infty$ as $T \to \infty$ with $k \to \infty$.

Finally, note that the same result holds if $k^{11}/T \to \infty$ and $k/T \to 0$ as $T \to \infty$ because from Lemma A.8, $\|\hat{\beta} - \beta\|$ would be bounded below by the case $k^{11}/T \to 0$ and we still obtain $\sum_{i=1}^{k} \hat{b}_i \to \infty$ as $T \to \infty$. To complete the proof, it remains to show that s_{ek}^2 remains bounded, regardless of the rate at which k approaches infinity. Note first that $s_{ek}^2 = T^{-1} \sum_{t=k+1}^{T} \hat{e}_{tk}^2 \leq T^{-1} \sum_{t=k+1}^{T} e_{tk}^2$, using standard properties of projections. Hence, all we need to show is that $T^{-1} \sum_{t=k+1}^{T} e_{tk}^2$ remains bounded. This is straightforward using the following representation of e_{tk}:

$$e_{tk} = e_t + \sum_{i=k+1}^{\infty} b_i \Delta y_{t-i}$$

and using the fact that $b_i = -(1 + c/\delta\sqrt{T})(1 - \delta/\sqrt{T})^i$. Details are omitted. ■

The Nearly Twice Integrated Model

Proof of Theorem 4.2. It is useful first to state the following lemma proved in Nabeya and Perron (1994).

LEMMA A.9. *Let $\{y_t\}$ and $\{u_t\}$ be processes given by (4.1) and (4.3). Define $J_c(r) = \int_0^r \exp((r - s)c) \, dW(s)$, $J_\phi(r) = \int_0^r \exp((r - s)\phi) \, dW(s)$ and $Q_c(J_\phi(r)) = \int_0^r \exp((r - v)c)J_\phi(v) \, dv$. As $T \to \infty$:*

(a)

$$T^{-3/2} y_T \Rightarrow \sigma_e Q_c(J_\phi(1));$$

(b)

$$T^{-4} \sum_{t=1}^{T} y_t^2 \Rightarrow \sigma_e^2 \int_0^1 Q_c(J_\phi(r))^2 \, dr;$$

(c)

$$T^{-3} \sum_{t=1}^{T} y_{t-1} u_t \Rightarrow (\sigma_e^2/2)\left\{ Q_c(J_\phi(1))^2 - 2c \int_0^1 Q_c(J_\phi(r))^2 \, dr \right\};$$

(d)

$$T^{-2} \sum_{t=1}^{T} u_t^2 \Rightarrow \sigma_e^2 \int_0^1 J_\phi(r)^2 \, dr.$$

The autoregressive representation of the data-generating process is

$$\Delta y_t = (-c\phi/T^2)y_{t-1} + (1 + (c + \phi)/T + c\phi/T^2)\Delta y_{t-1} + e_t.$$

This implies $b_i = 0$ for $i > 1$. Thus, $\lim_{k \to \infty} \Sigma_{i=1}^{k} b_i = 1 + (c + \phi)/T + c\phi/T^2$. To derive the limit of $\hat{b}(1)$, it proves convenient to write the regression (A.1) as

$$\Delta y_t = \delta_0 y_{t-1} + \delta_1 \Delta y_{t-1} + \sum_{i=2}^{k} \delta_i \Delta^2 y_{t-i+1} + e_{tk},$$

where $\delta_1 = \Sigma_{i=1}^{k} b_i = b_1 = (1 + ((c + \phi)/T) + (c\phi/T^2))$ and $\delta_i = -\Sigma_{j=i}^{k} b_j = 0$ $(i = 2, \ldots, k)$ (note that $\delta_0 \equiv b_0$ and $e_{tk} = e_t$). We need to derive the limit of $T(\hat{\delta}_1 - \delta_1)$. Let $Z_t' = (\Delta^2 y_{t-1}, \ldots, \Delta^2 y_{t-k+1})$, $U_t' = (y_{t-1}, \Delta y_{t-1}, Z_t')$, and define the following $(k+1) \times (k+1)$ matrices:

$$\hat{R}_T = \sum_{t=k+1}^{T} U_t U_t' = \begin{bmatrix} \sum_{t=k+1}^{T} y_{t-1}^2 & \sum_{t=k+1}^{T} y_{t-1} \Delta y_{t-1} & \sum_{t=k+1}^{T} y_{t-1} Z_t' \\ \sum_{t=k+1}^{T} \Delta y_{t-1} y_{t-1} & \sum_{t=k+1}^{T} \Delta y_{t-1}^2 & \sum_{t=k+1}^{T} \Delta y_{t-1} Z_t' \\ \sum_{t=k+1}^{T} y_{t-1} Z_t & \sum_{t=k+1}^{T} \Delta y_t Z_t & \sum_{t=k+1}^{T} Z_t Z_t' \end{bmatrix},$$

$$\tilde{R}_T = \begin{bmatrix} H_T & K_T & & & \\ K_T & J_T & & & \\ & & \sigma_e^2 & & \\ & & & \sigma_e^2 & \\ & & & & \sigma_e^2 \end{bmatrix}$$

with $H_T = T^{-4} \Sigma_{t=k+1}^{T} V_{t-1}^2$, $K_T = cT^{-4} \Sigma_{t=k+1}^{T} V_{t-2}^2 + T^{-3} \Sigma_{t=k+1}^{T} V_{t-2} W_{t-1}$ and $J_T = T^{-2} \Sigma_{t=k+1}^{T} ((c/T) V_{t-2} W_{t-1})^2$ where $W_t = \Sigma_{j=1}^{t} \exp((t-j)\phi/T) e_j$ and $V_t = \Sigma_{i=1}^{t} \exp((t-i)c/T) W_i$. Also,

$$R = \sigma_e^2 \begin{bmatrix} \lambda_0 & \lambda_1 & & & \\ \lambda_1 & \lambda_2 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix},$$

where $\lambda_0 = \int_0^1 Q_c(J_\phi(r))^2 \, dr$, $\lambda_1 = \frac{1}{2} Q_c(J_\phi(1))^2$, and $\lambda_2 = c Q_c(J(1))^2 - c^2 \int_0^1 Q_c(J_\phi(r))^2 \, dr + \int_0^1 J_\phi(r)^2 \, dr$. Note that using standard results (see, e.g., Nabeya and Perron, 1994) we have

$$H_T = T^{-4} \sum_{t=k+1}^{T} V_{t-1}^2 \Rightarrow \sigma_e^2 \lambda_0,$$

$$K_T = cT^{-4} \sum_{t=k+1}^{T} V_{t-2}^2 + T^{-3} \sum_{t=k+1}^{T} V_{t-2} W_{t-1} \Rightarrow \sigma_e^2 \lambda_1,$$

$$J_T = T^{-2} \sum_{t=k+1}^{T} \left(\frac{c}{T} V_{t-2} + W_{t-1} \right)^2 \Rightarrow \sigma_e^2 \lambda_2.$$

Hence, $\|\tilde{R}_T - R\| \to 0$. Also define $D_T = \text{diag}\{T^{-2}, T^{-1}, T^{-1/2}, \ldots, T^{-1/2}\}$. We first show that each element of the matrix $D_T \hat{R}_T D_T$ converges in distribution to the corresponding element of R.

LEMMA A.10. *Let* $\{y_t\}$ *be generated by (4.1) and (4.3). Then for* $i, j = 1, \ldots, k - 1$, *as* $T \to \infty$:

(a)

$$T^{-3} \sum_{t=k+1}^{T} y_{t-1} \Delta y_{t-1} \Rightarrow (\sigma_e^2/2) Q_c (J_\phi(1))^2 \equiv \lambda_1;$$

(b)

$$T^{-2} \sum_{t=k+1}^{T} \Delta y_{t-1}^2 \Rightarrow \sigma_e^2 \left(c Q_c (J_\phi(1))^2 - c^2 \int_0^1 Q_c (J_\phi(r))^2 \, dr \right.$$

$$\left. + \int_0^1 J_\phi(r)^2 \, dr \right) \equiv \lambda_2;$$

(c)

$$T^{-1} \sum_{t=k+1}^{T} \Delta^2 y_{t-i} \Delta^2 y_{t-j} \Rightarrow \sigma_e^2 \quad \text{if } i = j \quad \text{and} \quad 0 \text{ otherwise};$$

(d)

$$T^{-5/2} \sum_{t=k+1}^{T} y_{t-1} \Delta^2 y_{t-i} \Rightarrow 0;$$

(e)

$$T^{-3/2} \sum_{t=k+1}^{T} \Delta y_{t-1} \Delta^2 y_{t-i} \Rightarrow 0.$$

Proof. To prove (a), note that $\Delta y_{t-1} = (c/T) y_{t-2} + u_{t-1}$. Therefore

$$T^{-3} \sum_{t=k+1}^{T} y_{t-1} \Delta y_{t-1}$$

$$= T^{-3} \sum_{t=k+1}^{T} ((1 + c/T) y_{t-2} + u_{t-1}) \left(\frac{c}{T} y_{t-2} + u_{t-1} \right)$$

$$= c(1 + c/T) T^{-4} \sum_{t=k+1}^{T} y_{t-2}^2 + T^{-3}(1 + c/T) \sum_{t=k+1}^{T} y_{t-2} u_{t-1} + o_p(1)$$

$$\Rightarrow \frac{\sigma_e^2}{2} Q_c (J_\phi(1))^2 \equiv \lambda_1 \sigma_e^2,$$

using Lemma A.9(b) and (c). The proof of part (b) follows similarly using Lemma A.9(b)–(d). To prove part (c), note that $\Delta^2 y_{t-i} = (c/T)^2 y_{t-i-2} + ((c + \phi)/T) u_{t-i-1} + e_{t-i}$. The result is immediate after expanding terms. Parts (d) and (e) follow using analogous arguments. ∎

Now let $Q = D_T \hat{R}_T D_T - \tilde{R}_T$. It is straightforward to show that a typical element of this matrix, q_{ij}, satisfies $TE(q_{ij}^2) \leq C$ for some constant C. Hence, $E\|Q\|^2 \leq C(k+1)^2/T$. Also, because $\tilde{R}_T$ is a block diagonal matrix with lower $(k-1) \times (k-1)$ block being the identity matrix (scaled), we have $\|\tilde{R}_T^{-1}\| = O_p(1)$. Using arguments as in Said and Dickey (1984), we have $k^{1/2}\|(D_T \hat{R}_T D_T)^{-1} - \tilde{R}_T^{-1}\| \to 0$ in probability if $k^3/T \to 0$.

Because $e_{tk} = e_t$ it remains to establish the limit of $\|D_T \sum_{t=k+1}^{T} U_t e_t\|$. We start with the following lemma.

LEMMA A.11. *Let $\{y_t\}$ be generated by (4.1) and (4.3). Then (a) $E[y_t y_s] \leq T^2 C$; (b) $E[\Delta y_t^2] \leq TC$; (c) $E[(\Delta^2 y_t)^2] \leq C$.*

Proof. For part (a), because $y_t = \sum_{r=1}^{t}(1+c/T)^{t-r}\sum_{j=1}^{r}(1+\phi/T)^{r-j}e_j$, and $E(e_i e_j) = 0$ if $i \neq j$,

$$
\begin{aligned}
E(y_t y_s) &= E\left[\left(\sum_{r=1}^{t}(1+c/T)^{t-r}\sum_{j=1}^{r}(1+\phi/T)^{r-j}e_j\right)\right.\\
&\qquad\left.\times\left(\sum_{u=1}^{s}(1+c/T)^{s-u}\sum_{i=1}^{u}(1+\phi/T)^{u-i}e_i\right)\right]\\
&\leq CT^2 E(e_t^2) = T^2 C,
\end{aligned}
$$

because $(1+c/T)$ and $(1+\phi/T)$ are bounded by some constant. Analogous arguments show that $E(u_t u_s) \leq TC$ and $E(y_{t-1} u_t) \leq T^2 C$. For parts (b) and (c), write $\Delta y_t = (c/T)y_{t-1} + u_t$, and $\Delta^2 y_t = (c/T)\Delta y_{t-1} + (\phi/T)u_{t-1} + e_t$. It follows that

$$
E[\Delta y_t^2] = E\left[\frac{c^2}{T^2}y_{t-1}^2\right] + E[u_t^2] + \frac{2c}{T}E[y_{t-1}u_t] \leq TC,
$$

$$
E[(\Delta^2 y_t)^2] = E\left[\left(\frac{c}{T}\Delta y_{t-1} + \frac{\phi}{T}u_{t-1} + e_t\right)^2\right] \leq \sigma_e^2 + o(1) = C. \qquad\blacksquare
$$

LEMMA A.12. $\|D_T \sum_{t=k+1}^{T} U_t e_t\| = O_p(k^{1/2})$.

Proof. Note first that

$$
E\left\|D_T\sum_{t=k+1}^{T}U_t e_t\right\|^2 = E\left(T^{-2}\sum_{t=k+1}^{T}y_{t-1}e_t\right)^2 + E\left(T^{-1}\sum_{t=k+1}^{T}\Delta y_{t-1}e_t\right)^2
$$

$$
+ \sum_{i=2}^{k}E\left(T^{-1/2}\sum_{t=k+1}^{T}\Delta^2 y_{t-i+1}e_t\right)^2
$$

and that y_{t-1}, Δy_{t-1}, and $\Delta^2 y_{t-i+1}$ are independent of e_t. The result follows using Lemma A.11, which allows us to derive the following:

$$
E\left(T^{-2}\sum_{t=k+1}^{T}y_{t-1}e_t\right)^2 = T^{-4}\sum_{t=k+1}^{T}E[y_{t-1}^2]E[e_t^2] = O(T^{-1}),
$$

$$
E\left(T^{-1}\sum_{t=k+1}^{T}\Delta y_{t-1}e_t\right)^2 = T^{-2}\sum_{t=k+1}^{T}E[(\Delta y_{t-1})^2]E[e_t^2] = O(1),
$$

$$
\sum_{i=2}^{k}E\left(T^{-1/2}\sum_{t=k+1}^{T}\Delta^2 y_{t-i+1}e_t\right)^2 = \sum_{i=2}^{k}T^{-1}\sum_{t=k+1}^{T}E[(\Delta^2 y_{t-i+1})^2]E(e_t^2) = O(k). \qquad\blacksquare
$$

We are now in a position to derive the limiting distribution of the estimates $\hat{\delta}$. We have

$$D_T^{-1}(\hat{\delta} - \delta) = ((D_T \hat{R}_T D_T)^{-1} - \tilde{R}_T^{-1})D_T \sum_{t=k+1}^{T} U_t e_t + \tilde{R}_T^{-1} D_T \sum_{t=k+1}^{T} U_t e_t. \tag{A.4}$$

Taking norms, the first and second terms are $o_p(1)$ and $O_p(k^{1/2})$ (because $\|\tilde{R}_T^{-1}\| = O_p(1)$), respectively, so that $\|D_T^{-1}(\hat{\delta} - \delta)\| = O_p(k^{1/2})$. Because $\|D_T^{-1}\| = O(T^2)$ and $k = o(T^{1/3})$, we have $\|\hat{\delta} - \delta\| \to 0$. Note that (A.4) implies $D_T^{-1}(\hat{\delta} - \delta) = O_p(1)$, and the limiting distribution of $D_T^{-1}(\hat{\delta} - \delta)$ is that of $\tilde{R}_T^{-1} D_T \sum_{t=k+1}^{T} U_t e_t$, which is the same as the limiting distribution of $R^{-1} D_T \sum_{t=k+1}^{T} U_t e_t$ in view of the fact that $\|\tilde{R}_T - R\| \to 0$. We are interested in the second element of this vector. By block diagonality, the limit of the first two elements is given by

$$\lim_{T \to \infty} \begin{pmatrix} T^2(\hat{\delta}_0 - \delta_0) \\ T(\hat{\delta}_1 - \delta_1) \end{pmatrix} = \begin{bmatrix} \lambda_0 & \lambda_1 \\ \lambda_1 & \lambda_2 \end{bmatrix}^{-1} \lim_{T \to \infty} \begin{bmatrix} T^{-2} \sum_{t=k+1}^{T} y_{t-1} e_t \\ T^{-1} \sum_{t=k+1}^{T} \Delta y_{t-1} e_t \end{bmatrix} \Big/ \sigma_e^2. \tag{A.5}$$

It is straightforward to show the following limits:

$$T^{-2} \sum_{t=k+1}^{T} y_{t-1} e_t \Rightarrow \sigma_e^2 \int_0^1 Q_c(J_\phi(r)) \, dW(r) \equiv \sigma_e^2 \lambda_3;$$

$$T^{-1} \sum_{t=k+1}^{T} \Delta y_{t-1} e_t \Rightarrow \sigma_e^2 \left(c \int_0^1 Q_c(J_\phi(r)) \, dW(r) + \int_0^1 J_\phi(r) \, dW(r) \right) \equiv \sigma_e^2 \lambda_4. \tag{A.6}$$

Let $T(\hat{\delta} - \delta_1) \to \eta$, where

$$\eta = (-\lambda_1 \lambda_3 + \lambda_0 \lambda_4)/(\lambda_0 \lambda_2 - \lambda_1^2). \tag{A.7}$$

Using this result and (A.5), we have $T(\hat{\delta}_1 - 1) = T(\hat{\delta}_1 - \delta_1) + T(\delta_1 - 1) \Rightarrow \eta + c + \phi$. Because $\|\hat{\delta} - \delta\| \to 0$, it is straightforward to show that $s_{ek}^2 \to \sigma_e^2$ and $T^{-2} s_{AR}^2 \Rightarrow \sigma_e^2/(\eta + c + \phi)^2$. This proves Theorem 4.2. ∎

Nearly Seasonally Integrated Model

Proof of Theorem 4.3. We first define the following variables to be used throughout. Let

$$A(r) = (\phi - c)[Q_c(J_{\phi,1}(r)) - Q_c(J_{\phi,2}(r))] + 2J_{c,1}(r),$$

$$B(r) = J_{\phi,1}(r) - J_{\phi,2}(r),$$

$$C(r) = A(r) - B(r),$$

where $J_{c,1}(s) = \int_0^s \exp((s-v)c) \, dW_1(v)$, $J_{\phi,i}(s) = \int_0^s \exp((s-v)\phi) \, dW_i(v)$, $Q_c(J_{\phi,i}(r)) = \int_0^r \exp((r-s)c)J_{\phi,i}(s) \, ds$ for $i = 1, 2$, $W_1(r)$ and $W_2(r)$ being independent Wiener processes.

The following lemma is proved in Nabeya and Perron (1994).

LEMMA A.13. *Let* $\{y_t\}$ *and* $\{u_t\}$ *be generated by* (4.1) *and* (4.4). *Then*

(a)

$$T^{-2} \sum_{t=1}^{T} y_{t-1}^2 \Rightarrow (\sigma_e^2/8) \int_0^1 (C(r)^2 + B(r)^2) \, dr;$$

(b)

$$T^{-2} \sum_{t=1}^{T} y_{t-1} u_t \Rightarrow -(\sigma_e^2/4) \int_0^1 B(r)^2 \, dr;$$

(c)

$$T^{-1} y_T^2 \Rightarrow (\sigma_e^2/8) A(1)^2;$$

(d)

$$T^{-2} \sum_{t=1}^{T} u_t^2 \Rightarrow (\sigma_e^2/2) \int_0^1 B(r)^2 \, dr.$$

The autoregressive representation of the model in terms of (A.1) is

$$\Delta y_t = (c\phi/T^2 + 2c/T) y_{t-1} - (1 + (c + \phi)/T + c\phi/T^2) \Delta y_{t-1} + e_t. \tag{A.8}$$

This implies $b_1 = -(1 + ((c + \phi)/T) + (c\phi/T^2))$ and $b_i = 0$ for $i > 1$. Thus,

$$\lim_{k \to \infty} \sum_{i=1}^{k} b_i = -\left(1 + \frac{c + \phi}{T} + \frac{c\phi}{T_2}\right)$$

which converges to -1 as $T \to \infty$. Hence, to prove the theorem it is sufficient to prove that the OLS estimates from (A.1) are consistent for the coefficients stated in (A.8). Consistency of s_{ek}^2 for σ_e^2 then follows immediately. It is convenient to write the regression (A.1) as

$$\Delta y_t = \delta_0 y_{t-1} + \delta_1 \Delta y_{t-1} + \sum_{i=2}^{k} \delta_i \Delta_2 y_{t-i+1} + e_{tk},$$

where $\Delta_2 y_{t-i+1} = y_{t-i+1} - y_{t-i-1}$ with $\delta_i = \sum_{j=i}^{k} (-1)^{j-i} b_j$ $(i = 1, \ldots, k)$. Note that $\delta_0 \equiv b_0$, $e_{tk} = e_t$, and $\sum_{i=1}^{k-1} (\delta_i + \delta_{i+1}) + \delta_k = \sum_{i=1}^{k} b_i$. Let $Z_t' = (\Delta_2 y_{t-1}, \ldots, \Delta_2 y_{t-k+1})$, $U_t' = (y_{t-1}, \Delta y_{t-1}, Z_t')$, and define the following $(k+1) \times (k+1)$ matrices:

$$\hat{R}_T = \sum_{t=k+1}^{T} U_t U_t' = \begin{bmatrix} \sum_{t=k+1}^{T} y_{t-1}^2 & \sum_{t=k+1}^{T} y_{t-1} \Delta y_{t-1} & \sum_{t=k+1}^{T} y_{t-1} Z_t' \\ \sum_{t=k+1}^{T} \Delta y_{t-1} y_{t-1} & \sum_{t=k+1}^{T} \Delta y_{t-1}^2 & \sum_{t=k+1}^{T} \Delta y_{t-1} Z_t' \\ \sum_{t=k+1}^{T} y_{t-1} Z_t & \sum_{t=k+1}^{T} \Delta y_{t-1} Z_t & \sum_{t=k+1}^{T} Z_t Z_t' \end{bmatrix},$$

$$\tilde{R}_T = \begin{bmatrix} H_T & \frac{1}{2}J_T & & & \\ \frac{1}{2}J_T & J_T & & & \\ & & \sigma_e^2 & & \\ & & & \sigma_e^2 & \\ & & & & \sigma_e^2 \end{bmatrix}$$

with $H_T = T^{-2} \sum_{t=k+1}^{[T/2]} [2W_{t-1}^2 + 2W_{t-1}V_{t-1} + V_{t-1}^2]$, $J_T = T^{-2} \sum_{t=k+1}^{[T/2]} V_{t-1}^2$ where $V_t = X_{2,t}^* - X_{1,t}^*$,

$$W_t = \sum_{j=1}^{t} [(\phi - c)\exp(2c(t-j)/T)(X_{1,j}^* - X_{2,j}^*) + \exp(2c(t-j)/T)e_{2j}]$$

with $X_{1,t}^* = \sum_{j=1}^{t} \exp(2\phi(t-j)/T)e_{2j}$ and $X_{2,t}^* = \sum_{j=1}^{t} \exp(2\phi(t-j)/T)e_{2j-1}$. Also,

$$R = \sigma_e^2 \begin{bmatrix} \lambda_0 & \lambda_1/2 & & & \\ \lambda_1/2 & \lambda_1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix},$$

where $\lambda_0 = (\frac{1}{8})\int_0^1 [C(r)^2 + B(r)^2] dr$ and $\lambda_1 = (\frac{1}{2})\int_0^1 B(r)^2 dr$. Note that using standard results (see, e.g., Nabeya and Perron, 1994) we have $H_T \Rightarrow \sigma_e^2 \lambda_0$ and $J_T \Rightarrow \sigma_e^2 \lambda_1$. Hence, $\|\tilde{R}_T - R\| \to 0$. Also define $D_T = \text{diag}\{T^{-1}, T^{-1}, T^{-1/2}, \ldots, T^{-1/2}\}$. We first show that each element of the matrix $D_T \hat{R}_T D_T$ converges in distribution to the corresponding element of R. The relevant results are stated in the following lemma.

LEMMA A.14. *Let* $\{y_t\}$ *be generated by* (4.1) *and* (4.4). *Then for* $i, j = 2, \ldots, k$, *as* $T \to \infty$:

(a)

$$T^{-2} \sum_{t=k+1}^{T} y_{t-1} \Delta y_{t-1} \Rightarrow (\sigma_e^2/4) \int_0^1 B(r)^2 dr = (\sigma_e^2/2)\lambda_1;$$

(b)

$$T^{-2} \sum_{t=k+1}^{T} \Delta y_{t-1}^2 \Rightarrow (\sigma_e^2/2) \int_0^1 B(r)^2 dr = \sigma_e^2 \lambda_1;$$

(c)

$$T^{-3/2} \sum_{t=k+1}^{T} y_{t-1} \Delta_2 y_{t-i+1} \Rightarrow 0;$$

(d)

$$T^{-3/2} \sum_{t=k+1}^{T} \Delta y_{t-1} \Delta_2 y_{t-i+1} \Rightarrow 0;$$

(e)

$$T^{-1} \sum_{t=k+1}^{T} (\Delta_2 y_{t-i+1})^2 \Rightarrow \sigma_e^2;$$

(f)

$$T^{-1} \sum_{t=k+1}^{T} \Delta_2 y_{t-i+1} \Delta_2 y_{t-j+1} \Rightarrow 0.$$

Proof. To prove (a), note that because $y_{t-1} = \alpha_T y_{t-2} + u_{t-1}$ (with $\alpha_T = 1 + c/T$) and $\Delta y_{t-1} = (c/T) y_{t-2} + u_{t-1}$,

$$T^{-2} \sum_{t=k+1}^{T} y_{t-1} \Delta y_{t-1}$$

$$= T^{-3} c \alpha_T \sum_{t=k+1}^{T} y_{t-2}^2 + T^{-3} c \alpha_T \sum_{t=k+1}^{T} y_{t-2} u_{t-1} + T^{-2} \alpha_T \sum_{t=k+1}^{T} y_{t-2} u_{t-1}$$

$$+ T^{-2} \sum_{t=k+1}^{T} u_{t-1}^2 \Rightarrow (\sigma_e^2/4) \int_0^1 B(r)^2 \, dr = (\sigma_e^2/2)\lambda_1,$$

because the first two terms are $o_p(1)$, and using the results of Lemma A.13. For part (b),

$$T^{-2} \sum_{t=k+1}^{T} \Delta y_{t-1}^2 = T^{-2} \sum_{t=k+1}^{T} \left(\frac{c}{T} y_{t-2} + u_{t-1} \right) \left(\frac{c}{T} y_{t-2} + u_{t-1} \right)$$

$$= T^{-2} \sum_{t=k+1}^{T} u_{t-1}^2 + o_p(1) \Rightarrow (\sigma_e^2/2) \int_0^1 B(r)^2 \, dr = \sigma_e^2 \lambda_2.$$

Parts (c) to (f) follow analogously using the representation

$$\Delta_2 y_{t-i+1} = \frac{(c + \phi)}{T} y_{t-i} + \left[\frac{(c + \phi)}{T} + \frac{c\phi}{T^2} \right] y_{t-i-1} + e_{t-i+1},$$

along with the results that $T^{-2} \sum_{t=k+1}^{T} y_{t-i+1} y_{t-j+1} = O_p(1)$ and $T^{-1} \sum_{t=k+1}^{T} y_{t-i+1} \times e_{t-j+1} = O_p(1)$. ∎

Now let $Q = D_T \hat{R}_T D_T - \tilde{R}_T$. It is straightforward to show that a typical element of this matrix, q_{ij}, satisfies $TE(q_{ij}^2) \le CT$ for some constant C. Hence, $E\|Q\|^2 \le C(k+1)^2/T$. Also, because $\tilde{R}_T$ is a block diagonal matrix with lower $(k-1) \times (k-1)$ block being the identity matrix (scaled), we have $\|\tilde{R}_T^{-1}\| = O_p(1)$. Using arguments as in Said and Dickey (1984), we have $k^{1/2} \|(D_T \hat{R}_T D_T)^{-1} - \tilde{R}_T^{-1}\| \to 0$ in probability if $k^3/T \to 0$.

Because $e_{tk} = e_t$ it remains to establish the limit of $\|D_T \sum_{t=k+1}^T U_t e_t\|$. We start with the following lemma.

LEMMA A.15. *Let $\{y_t\}$ be defined as in (4.1) and (4.4). Then (a) $E[u_t^2] \leq TC$; (b) $E[y_{t-i}u_t] \leq TC$; (c) $E[y_t^2] \leq TC$; (d) $E[\Delta y_t^2] \leq TC$; (e) $E[(\Delta_2 y_{t-i+1})^2] \leq C$.*

Proof. Part (a) follows because u_t is a near-integrated process, hence Lemma A.1(b) applies. Following Nabeya and Perron (1994), define, for $s = 1,\dots,[T/2]$,

$$X_{1,s} = \sum_{j=1}^s (\rho_T^2)^{s-j} e_{2j} \quad \text{and} \quad X_{2,s} = \sum_{j=1}^s (\rho_T^2)^{s-j} e_{2j-1},$$

$$y_{2s} = (1 - \gamma_T) \sum_{j=1}^s (\alpha_T^2)^{s-j} X_{1,j} + (\alpha_T - \rho_T) \sum_{j=1}^s (\alpha_T^2)^{s-j} X_{2,j} + \gamma_T \sum_{j=1}^s (\alpha_T^2)^{s-j} e_{2j},$$

$$u_{2s} = X_{1,s} - \rho_T X_{2,s} \quad \text{and} \quad u_{2s-1} = X_{2,s} - (1/\rho_T) X_{1,s} + (1/\rho_T) e_{2s},$$

where $\rho_T = (1 + \phi/T)$, $\alpha_T = (1 + c/T)$, and $\gamma_T = \alpha_T/\rho_T$, and note that $1 - \gamma_T = O(T^{-1})$. Using these definitions, we have, for t odd and $t - 1 = 2s$,

$$E[y_{t-1}u_t] = E[y_{2s}u_{2(s+1)}]$$

$$\leq E\left| \frac{-1}{\rho_T}(1 - \gamma_T) \sum_{j=1}^s (\alpha_T^2)^{s-j} X_{1,j} X_{1,s+1} \right| + E\left| \frac{1}{\rho_T}(1 - \gamma_T) \sum_{j=1}^s (\alpha_T^2)^{s-j} X_{1,j} e_{2s+1} \right|$$

$$+ E\left| (\alpha_T - \rho_T) \sum_{j=1}^s (\alpha_T^2)^{s-j} X_{2,j} X_{2,s+1} \right| + E\left| (\gamma_T/\rho_T) \sum_{j=1}^t (\alpha_T^2)^{s-j} e_{2j} e_{2s+1} \right|$$

$$+ E\left| (\gamma_T/\rho_T) \left(\sum_{j=1}^s (\alpha_T^2)^{s-j} e_{2j} \right) X_{1,s+1} \right|.$$

It is easy to deduce that the second and the fourth terms are $o(1)$. For the first term,

$$E\left| \frac{-1}{\rho_T}(1 - \gamma_T) \sum_{j=1}^s (\alpha_T^2)^{s-j} X_{1,j} X_{1,s+1} \right| \leq \frac{1}{\rho_T} TC_1 T(1 - \gamma_T)\left(\frac{(\alpha_T^2)^s - 1)}{T(\alpha_T^2 - 1)} \right) \leq TC.$$

(A.9)

Similar derivations can be made to show that the third and the fifth terms are also $O(T)$. Analogous arguments hold for t even. Thus, $E|y_{t-1}u_t| \leq TC$. The arguments for the general case $E|y_{t-i}u_t|$ are similar. Consider now part (c). We have, for t even,

$$E[y_t^2] = (1 - \gamma_T)^2 \sum_{j=1}^s \sum_{i=1}^s (\alpha_T^2)^{s-j}(\alpha_T^2)^{s-i} E(X_{1,j} X_{1,i})$$

$$+ (\alpha_T - \rho_T)^2 \sum_{j=1}^s \sum_{i=1}^s (\alpha_T^2)^{s-j}(\alpha_T^2)^{s-i} E(X_{2,j} X_{2,i}) + \gamma_T^2 \text{Var}\left(\sum_{j=1}^s (\alpha_T^2)^{s-j} e_{2j} \right)$$

$$+ \gamma_T(\alpha_T - \rho_T) \sum_{j=1}^s \sum_{i=1}^s (\alpha_T^2)^{s-j}(\alpha_T^2)^{s-i} E(X_{2j} e_{2j}).$$

Using Lemma A.1(b), the first and second terms are bounded by TC. Lemma A.1 can be used to show the third and fourth terms are also $O(T)$. Hence, $E[y_t^2] \leq TC$ for t even. Because $E[y_t^2] = \alpha_T^2 E[y_{t-1}^2] + 2\alpha_T E[y_{t-1}u_t] + E[u_t^2]$, it follows that $E[y_t^2] \leq TC$ for t odd also. For part (d),

$$E[\Delta y_t \, \Delta y_s] = E\left[\left(\frac{c}{T} y_{t-1} + u_t\right)\left(\frac{c}{T} y_{s-1} + u_s\right)\right]$$

$$= E\left[\frac{c}{T^2} y_{t-1} y_{s-1}\right] + E[u_t u_s] + \frac{c}{T} E[y_{s-1} u_t] + \frac{c}{T} [y_{t-1} u_t] \leq TC,$$

using parts (a) to (c). Now for part (e), we have

$$\Delta_2 y_{t-i+1} = \frac{(c + \phi)}{T} y_{t-i} + \left[\frac{(c + \phi)}{T} + \frac{c\phi}{T^2}\right] y_{t-i-1} + e_{t-i+1}$$

and $E[(\Delta_2 y_{t-i+1})^2]$ is less than the sum of the expectations of the square of each term,

$$E[(\Delta_2 y_{t-i+1})^2] \leq (c + \phi)^2 T^{-2} E|y_{t-i}^2| + \left(\frac{c + \phi}{T} + \frac{c\phi}{T^2}\right)^2 E|y_{t-i-1}^2| + E|e_{t-1}^2| \leq C$$

using part (c).   ∎

The results in the next lemma follow.

LEMMA A.16. *Let $\{y_t\}$ be generated by (4.1) and (4.4). Then*

(a)

$$E\left(T^{-1} \sum_{t=k+1}^{T} y_{t-1} e_t\right)^2 = T^{-2} \sum_{t=k+1}^{T} E[y_{t-1}^2]E[e_t^2] = O(1);$$

(b)

$$E\left(T^{-1} \sum_{t=k+1}^{T} \Delta y_{t-1} e_t\right)^2 = T^{-2} \sum_{t=k+1}^{T} E[\Delta y_{t-1}^2]E[e_t^2] = O(1);$$

(c)

$$E \sum_{i=1}^{k} \left(T^{-1/2} \sum_{t=k+1}^{T} \Delta_2 y_{t-i+1} e_t\right)^2 = T^{-1} \sum_{i=1}^{k} \sum_{t=k+1}^{T} E[(\Delta_2 y_{t-i+1})^2]E[e_t^2] = O(k).$$

We are now in a position to prove the following result.

LEMMA A.17. $\|D_T \sum_{t=k+1}^{T} U_t e_t\| = O_p(k^{1/2}).$

Because

$$E \left\| D_T \sum_{t=k+1}^{T} U_t e_t \right\|^2 = E \left(T^{-1} \sum_{t=k+1}^{T} y_{t-1} e_t \right)^2 + E \left(T^{-1} \sum_{t=k+1}^{T} \Delta y_{t-1} e_t \right)^2$$

$$+ \sum_{i=2}^{k} E \left(T^{-1/2} \sum_{t=k+1}^{T} \Delta_2 y_{t-i+1} e_t \right)^2,$$

the result follows directly from Lemma A.16.

We are now in a position to show consistency of the estimates $\hat{\delta}$. We have

$$D_T^{-1}(\hat{\delta} - \delta) = ((D_T \hat{R}_T D_T)^{-1} - \tilde{R}_T^{-1}) D_T \sum_{t=k+1}^{T} U_t e_t + \tilde{R}_T^{-1} D_T \sum_{t=k+1}^{T} U_t e_t.$$

Taking norms, the first and second terms are $o_p(1)$ and $O_p(k^{1/2})$ (because $\|\tilde{R}_T^{-1}\| = O_p(1)$), respectively, so that $\|D_T^{-1}(\hat{\delta} - \delta)\| = O_p(k^{1/2})$. Because $\|D_T^{-1}\| = O(T)$ and $k = o(T^{1/3})$, we have $\|\hat{\delta} - \delta\| \to 0$, and $\hat{b}(1) = \hat{\delta}_1 \to -1$. Thus, $s_{AR}^2 = s_{ek}^2/(1 - \hat{\delta}_1)^2 \to \sigma_e^2/4$, because $s_{ek}^2 \to \sigma_e^2$.   ∎