ECONOMETRICA
JOURNAL OF THE ECONOMETRIC SOCIETY

LAG LENGTH SELECTION AND THE CONSTRUCTION OF UNIT ROOT TESTS WITH GOOD SIZE AND POWER

BY SERENA NG AND PIERRE PERRON[1]

It is widely known that when there are errors with a moving-average root close to −1, a high order augmented autoregression is necessary for unit root tests to have good size, but that information criteria such as the *AIC* and the *BIC* tend to select a truncation lag (*k*) that is very small. We consider a class of Modified Information Criteria (*MIC*) with a penalty factor that is sample dependent. It takes into account the fact that the bias in the sum of the autoregressive coefficients is highly dependent on *k* and adapts to the type of deterministic components present. We use a local asymptotic framework in which the moving-average root is local to −1 to document how the *MIC* performs better in selecting appropriate values of *k*. In Monte-Carlo experiments, the *MIC* is found to yield huge size improvements to the DF^{GLS} and the feasible point optimal P_T test developed in Elliott, Rothenberg, and Stock (1996). We also extend the *M* tests developed in Perron and Ng (1996) to allow for *GLS* detrending of the data. The *MIC* along with *GLS* detrended data yield a set of tests with desirable size and power properties.

KEYWORDS: Integrated processes, truncation lag, *GLS* detrending, information criteria.

1. INTRODUCTION

MANY UNIT ROOT TESTS have been developed for testing the null hypothesis of a unit root against the alternative of stationarity. While the presence or absence of a unit root has important implications, many remain skeptical about the conclusions drawn from such tests. This concern is justifiable, as these tests generally suffer from two problems. First, many tests have low power when the root of the autoregressive polynomial is close to but less than unity (e.g., DeJong et al. (1992)). Second, the majority of the tests suffer from severe size distortions when the moving-average polynomial of the first differenced series has a large negative root (e.g., Schwert (1989), Perron and Ng (1996)). Although less severe, the problem also arises when there is a large negative autoregressive root in the residuals. The consequence is over-rejections of the unit root hypothesis. While few economic time series are found to have negative serial correlation of the autoregressive type, many do exhibit a large negative moving-average root. The inflation rate is an example, and omitted outliers have also been shown to induce

a negative moving average root in the error process (e.g., Franses and Haldrup (1994)). It is therefore desirable to have powerful unit root tests that are robust to size distortions.

The implementation of unit root tests often necessitates the selection of an autoregressive truncation lag, say, k. This is required in the autoregression used to form the Said-Dickey-Fuller test. It is also required in constructing an autoregressive estimate of the spectral density at frequency zero. However, simulations have repeatedly shown a strong association between k and the severity of size distortions and/or the extent of power loss (e.g., Ng and Perron (1995), and Lopez (1997)). To date, there exists no operational procedure for selecting k that yields robust and satisfactory results. The problem is that while a small k is adequate for finite order autoregressive errors and $ARMA$ processes with small moving-average components, a large k is generally necessary for noise functions with a moving-average root that is large. For the latter class of models, selection rules such as the Schwarz or Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC) tend to select values of k that are generally too small for unit root tests to have good sizes.

This paper sets out to resolve two problems. One is to enhance the power of several tests that have been shown to have small size distortions. More precisely, in Perron and Ng (1996), we analyzed a class of modified tests, originally proposed by Stock (1990), and showed that these tests are far more robust to size distortions than other unit root tests in the literature when the residuals have negative serial correlation. Exploiting the findings in Elliott, Rothenberg, and Stock (1996) (hereafter ERS) and Dufour and King (1991) that local GLS detrending of the data yields substantial power gains, we apply the idea of GLS detrending to the modified tests and show that non-negligible size and power gains can be made when used in conjunction with an autoregressive spectral density estimator at frequency zero provided the truncation lag is appropriately selected.

Our second task is to provide an improved procedure for choosing the truncation lag. We argue that the penalty of k assigned to overfitting in information criteria such as the AIC under-estimates the cost of a low order model when the unit root process has a negative moving-average root and, hence, tends to select a lag length that is too small. We suggest a class of modified information criteria that takes better account of the cost of underfitting. The modified AIC is shown to lead to substantial size improvements over standard information criteria in all the unit root tests considered. Taking the two steps together, GLS detrending along with the selection rule for k provide procedures which allow for the construction of unit root tests with much improved size and power. Additionally, using GLS detrended data to estimate the spectral density at frequency zero is found to have favorable size and power implications.

The paper proceeds as follows. After setting up the preliminaries in Section 2, Section 3 considers the class of M tests with GLS detrending. Section 4 explains why information criteria will underestimate k and Section 5 proposes an alternative penalty factor in the context of integrated data. Section 6 presents the properties of the modified information criteria using a local asymptotic framework

and shows why the *AIC* version is preferable. Simulation results are presented in Section 7, and the procedures are illustrated using inflation rates for the G7 countries in Section 8. Section 9 offers brief concluding remarks. Mathematical derivations are given in the Appendix.

2. PRELIMINARIES

We assume that we have $T + 1$ observations generated by (for $t = 0, \ldots, T$):

$$(1) \qquad y_t = d_t + u_t, \quad u_t = \alpha u_{t-1} + v_t,$$

where $E(u_0^2) < \infty$ (even as $T \to \infty$), $v_t = \delta(L)e_t = \sum_{j=0}^{\infty} \delta_j e_{t-j}$ with $\sum_{j=0}^{\infty} j|\delta_j| < \infty$ and $\{e_t\} \sim$ i.i.d. $(0, \sigma_e^2)$. The non-normalized spectral density at frequency zero of v_t is given by $\sigma^2 = \sigma_e^2 \delta(1)^2$. Also, $T^{-1/2} \sum_{t=1}^{[sT]} v_t \Rightarrow \sigma W(s)$, where $\Rightarrow$ denotes weak convergence in distribution and $W(s)$ is a Wiener process defined on $C[0, 1]$. In (1), $d_t = \psi' z_t$, where z_t is a set of deterministic components. We consider $d_t = \sum_{i=0}^{p} \psi_i t^i$, with special focus on $p = 0, 1$, though the analysis remains valid for more general cases.[2] Of interest is testing the null hypothesis $\alpha = 1$ against $\alpha < 1$. The *DF* test due to Dickey and Fuller (1979) and Said and Dickey (1984) is the t statistic for $\beta_0 = 0$ in the autoregression:

$$(2) \qquad \Delta y_t = d_t + \beta_0 y_{t-1} + \sum_{j=1}^{k} \beta_j \Delta y_{t-j} + e_{tk}.$$

In Perron and Ng (1996), we analyzed the properties of three tests: MZ_α, MZ_t, and *MSB*, collectively referred to as the *M* tests. These are defined as (for the no deterministic term case):

$$(3) \qquad MZ_\alpha = (T^{-1} y_T^2 - s_{AR}^2) \left(2T^{-2} \sum_{t=1}^{T} y_{t-1}^2 \right)^{-1},$$

$$(4) \qquad MSB = \left(T^{-2} \sum_{t=1}^{T} y_{t-1}^2 / s_{AR}^2 \right)^{1/2},$$

and $MZ_t = MZ_\alpha \times MSB$. All three tests are based on s_{AR}^2, an autoregressive estimate of the spectral density at frequency zero of v_t. Let $\hat{\beta}(1) = \sum_{i=1}^{k} \hat{\beta}_i, \hat{\sigma}_k^2 = (T-k)^{-1} \sum_{t=k+1}^{T} \hat{e}_{tk}^2$, with $\hat{\beta}_i$ and $\{\hat{e}_{tk}\}$ obtained from (2) estimated by *OLS*. Then,

$$(5) \qquad s_{AR}^2 = \hat{\sigma}_k^2 / (1 - \hat{\beta}(1))^2.$$

The *M* tests for $p = 0$ and 1 can be obtained upon replacing y_{t-1} and y_T by the residuals obtained from least squares detrending. The MZ_α and MZ_t tests can be

---

[2] Most of the results presented here, in particular Theorem 1 below, hold with the less restrictive assumption that $\{e_t\}$ is a martingale difference sequence with respect to the sigma algebra $\mathscr{F}_t = \{e_s; s \le t\}$ with $\sigma_e^2 = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E(e_t^2 | \mathscr{F}_{t-1})$.

viewed as modified versions of the Phillips (1987) and Phillips-Perron (1988) Z_α and Z_t tests, referred to as the Z tests below (see Perron and Ng (1998)). The Z tests suffer from severe size distortions when v_t has a negative moving-average root. When the root is close to minus one (e.g., $-.8$), the rejection rate can be as high as 100% (see Schwert (1989)). Perron and Ng (1996) showed that the M tests have dramatically smaller size distortions than most (if not all) unit root tests in the literature in cases of negative moving-average errors *if the autoregressive spectral density estimator defined above is used in conjunction with a suitably chosen* k. The three M tests have similar properties and we use MZ_α for illustration.

3. THE TESTS UNDER *GLS* DETRENDING

To improve power, ERS proposed local to unity *GLS* detrending of the data. For any series $\{x_t\}_{t=0}^T$, define $(x_0^{\bar\alpha}, x_t^{\bar\alpha}) \equiv (x_0, (1 - \bar\alpha L)x_t), t = 1, \ldots, T$, for some chosen $\bar\alpha = 1 + \bar c/T$. The *GLS* detrended series is defined as

$$(6) \qquad \tilde y_t \equiv y_t - \hat\psi' z_t$$

where $\hat\psi$ minimizes $S(\bar\alpha, \psi) = (y^{\bar\alpha} - \psi' z^{\bar\alpha})'(y^{\bar\alpha} - \psi' z^{\bar\alpha})$. If v_t is i.i.d. normal, the point optimal test of the null hypothesis $\alpha = 1$ against the alternative hypothesis $\alpha = \bar\alpha$ is the likelihood ratio statistic, $L = S(\bar\alpha) - S(1)$, where $S(\bar\alpha) = \min_\psi S(\bar\alpha, \psi)$. ERS considered a feasible point optimal test that takes into account that v_t may be serially correlated. The statistic is

$$(7) \qquad P_T = [S(\bar\alpha) - \bar\alpha S(1)]/s_{AR}^2,$$

where s_{AR}^2 is defined as in (5). The value of $\bar c$ is chosen such that the asymptotic local power function of the test is tangent to the power envelope at 50% power. For $p = 0$ this is -7.0 and for $p = 1$, it is -13.5. The P_T test did not perform well in simulations when k was selected by the *BIC*. Large size distortions were recorded. ERS then recommended the DF^{GLS} statistic (whose local asymptotic power function also lies close to the Gaussian local power envelope) as the t-statistic for testing $\beta_0 = 0$ from the following regression estimated by *OLS*:

$$(8) \qquad \Delta \tilde y_t = \beta_0 \tilde y_{t-1} + \sum_{j=1}^k \beta_j \Delta \tilde y_{t-j} + e_{tk}.$$

3.1. *The* M^{GLS} *Tests*

While the power gains of the DF^{GLS} from using *GLS* detrended data are impressive, simulations also show that the test exhibits strong size distortions when v_t is an $MA(1)$ with a negative coefficient. For $T = 100$, ERS report rejection rates of 30 and 50% for $p = 0$ and 1, respectively, when the nominal rate is 5%. An explanation is provided in Section 4. Size distortions, however, are less of an issue with the M tests in theory as we showed in Perron and Ng (1996). In practice, it does require us to have a way to find the appropriate k. Suppose

for the moment that this lag length selection problem is solved. Since the power gains from the DF^{GLS} over the DF come from the use of GLS detrended data, it is natural to consider the M tests under GLS detrending. We now state the asymptotic properties of the M^{GLS} tests.

THEOREM 1: *Let y_t be generated as in* (1) *with $\alpha = 1 + c/T$ and consider data obtained from local GLS detrending at $\bar{\alpha} = 1 + \bar{c}/T$. Also, let s^2_{AR} be a consistent estimate of σ^2. If $p = 0$, $MZ^{GLS}_\alpha \Rightarrow 0.5(J_c(1)^2 - 1)(\int_0^1 J_c(r)^2 dr)^{-1}$ and $MSB^{GLS} \Rightarrow (\int_0^1 J_c(r)^2 dr)^{1/2}$. If $p = 1$, $MZ^{GLS}_\alpha \Rightarrow 0.5(V_{c,\bar{c}}(1)^2 - 1)(\int_0^1 V_{c,\bar{c}}(r)^2 dr)^{-1}$ and $MSB^{GLS} \Rightarrow (\int_0^1 V_{c,\bar{c}}(r)^2 dr)^{1/2}$, where $J_c(r)$ is an Ornstein-Uhlenbeck process defined by $dJ_c(r) = cJ_c(r) dr + dW(r)$ with $J_c(0) = 0$, and $V_{c,\bar{c}}(r) = J_c(r) - r[\lambda J_c(1) + 3(1 - \lambda) \int_0^1 sJ_c(s) ds]$, $\lambda = (1 - \bar{c})/(1 - \bar{c} + \bar{c}^2/3)$. Results for MZ^{GLS}_t can be obtained in view of the fact that $MZ^{GLS}_t = MZ^{GLS}_\alpha \times MSB^{GLS}$.*

The proof is based on standard results for sample moments of GLS detrended data and therefore omitted. The MZ^{GLS}_α, MZ^{GLS}_t, and MSB^{GLS} have similar size and power properties and will be referred to as the M^{GLS} tests. We have computed, via simulations, the local asymptotic power functions of the M tests with OLS and GLS detrending. These showed the M^{GLS} tests to be uniformly more powerful than the M tests and also to have local asymptotic power functions that are indistinguishable from those of the DF^{GLS} and the Gaussian asymptotic local power envelope. A graph is therefore not presented. The important point is that the M^{GLS} tests, like the DF^{GLS}, achieve local asymptotic power that is approximately the same as the feasible point-optimal test. We also consider two modified feasible point optimal tests and derive their limiting distributions:

$$(9) \qquad p = 0: \quad MP^{GLS}_T = \left[\bar{c}^2 T^{-2} \sum_{t=1}^T \tilde{y}^2_{t-1} - \bar{c} T^{-1} \tilde{y}^2_T \right] \Big/ s^2_{AR}$$

$$\Rightarrow \bar{c}^2 \int_0^1 J_c(r)^2 dr - \bar{c} J_c(1)^2,$$

$$p = 1: \quad MP^{GLS}_T = \left[\bar{c}^2 T^{-2} \sum_{t=1}^T \tilde{y}^2_{t-1} + (1 - \bar{c}) T^{-1} \tilde{y}^2_T \right] \Big/ s^2_{AR}$$

$$\Rightarrow \bar{c}^2 \int_0^1 V_{c,\bar{c}}(r)^2 dr + (1 - \bar{c}) V_{c,\bar{c}}(1)^2.$$

These tests are based on the same motivation that leads to the definition of the M tests in Stock (1990), namely, to provide functionals of sample moments that have the same asymptotic distributions as well known unit root tests. The MP_T is important because its limiting distribution coincides with that of the feasible point optimal test considered in ERS.

Critical values for the limiting distributions of the M^{GLS} tests were obtained by numerical simulations that approximate the Wiener process $W(r)$ by partial sums of $N(0, 1)$ random variables with 5,000 steps and 20,000 simulations. These are summarized in Table I.

TABLE I

ASYMPTOTIC CRITICAL VALUES[a]

Percentile	MZ_α^{GLS}	MZ_t^{GLS}, ADF^{GLS}	MSB^{GLS}	P_T^{GLS}, MP_T^{GLS}
		Case: $p = 0, \bar{c} = -7.0$		
.01	-13.8	-2.58	.174	1.78
.05	-8.1	-1.98	.233	3.17
.10	-5.7	-1.62	.275	4.45
		Case: $p = 1, \bar{c} = -13.5$		
.01	-23.8	-3.42	.143	4.03
.05	-17.3	-2.91	.168	5.48
.10	-14.2	-2.62	.185	6.67

[a]For MZ_α^{GLS}, MZ_t^{GLS}, and ADF^{GLS} with $p = 0$, the critical values are from Fuller (1976). The other entries were obtained using 20,000 simulations and 5,000 steps to approximate the Wiener processes.

3.2. *The Construction of the Autoregressive Spectral Density Estimator and the $\bar{M}^{GLS}$*

All that is required for Theorem 1 and the results of ERS to hold is that s_{AR}^2 converges to σ^2 under the null hypothesis of a unit root, and for the tests to be consistent, that Ts_{AR}^2 diverges under the alternative of stationarity. The attractiveness of GLS detrending is that it estimates the deterministic function with more precision and leads to reduced bias in $\hat{\beta}_i$. Since this should lead to a more precise estimate of $\hat{\beta}(1)$, we also consider estimating s_{AR}^2 using the autoregression (8) based on GLS detrended data. An additional advantage is that this estimator is invariant to the slope of the trend function.[3] Experimentation with different values of $\bar{c}$ when constructing the autoregressive spectral density gave similar results. Hence, $\bar{c} = -7.0$ and -13.5 will continue to be used. The tests so constructed are denoted $\bar{Z}^{GLS}$ and $\bar{M}^{GLS}$. These are to be distinguished from the Z^{GLS} and M^{GLS} statistics that use least squares detrended data to estimate s_{AR}^2.

4. THE SIZE ISSUE, THE SELECTION OF k, AND INFORMATION CRITERIA

To see the influence of the lag order k, we consider the finite sample size of the tests $\bar{M}Z_\alpha^{GLS}$, DF^{GLS}, and $\bar{Z}_\alpha^{GLS}$. The setup, which will be used throughout, is to consider data generated by (1) with $\alpha = 1$ and v_t either an $MA(1)$, i.e. $v_t = e_t + \theta e_{t-1}$, or an $AR(1)$, i.e. $v_t = \rho v_{t-1} + e_t$, with $e_t \sim$ i.i.d. $N(0, 1)$. The errors are generated from the rndn() function in Gauss with 999 as the seed, and 5,000 replications are used. We use the nominal size of 5% as the benchmark.

---

[3] When $p = 1$, Perron and Ng (1996), Stock (1990), and ERS estimate equation (2) with only a constant by ordinary least-squares. This still ensures a consistent estimate under the null hypothesis but the omission of the trend term implies a noninvariance with respect to the slope of the trend function in finite samples under the alternative hypothesis. This noninvariance is such that the tests lose power as the slope increases.

TABLE II.A

Exact Size of the Tests at Selected Values of k: $p = 0$

MA Case	T	θ	$k = 0$	1	2	3	4	5	6	7	8	9	10
$\overline{Z}_\alpha^{GLS}$	100	−0.8	0.994	0.904	0.808	0.740	0.698	0.672	0.654	0.644	0.636	0.633	0.627
		−0.5	0.618	0.322	0.224	0.193	0.176	0.181	0.182	0.192	0.197	0.207	0.216
		0.0	0.072	0.082	0.089	0.104	0.112	0.125	0.137	0.152	0.164	0.179	0.190
		0.5	0.010	0.143	0.056	0.116	0.100	0.129	0.135	0.153	0.158	0.179	0.187
		0.8	0.006	0.172	0.036	0.147	0.071	0.155	0.108	0.166	0.138	0.188	0.169
	250	−0.8	0.987	0.861	0.725	0.641	0.592	0.559	0.539	0.524	0.515	0.507	0.504
		−0.5	0.553	0.258	0.166	0.124	0.112	0.110	0.106	0.112	0.111	0.115	0.116
		0.0	0.060	0.062	0.063	0.066	0.069	0.075	0.077	0.081	0.087	0.087	0.092
		0.5	0.008	0.111	0.047	0.080	0.065	0.077	0.076	0.083	0.084	0.091	0.095
		0.8	0.004	0.145	0.030	0.108	0.049	0.099	0.062	0.096	0.074	0.100	0.085
$\overline{M}Z_\alpha^{GLS}$	100	−0.8	0.990	0.828	0.543	0.319	0.187	0.129	0.095	0.086	0.079	0.077	0.078
		−0.5	0.595	0.241	0.126	0.097	0.084	0.092	0.097	0.106	0.120	0.132	0.141
		0.0	0.065	0.070	0.082	0.095	0.104	0.119	0.130	0.145	0.158	0.172	0.185
		0.5	0.008	0.139	0.054	0.113	0.097	0.126	0.132	0.150	0.156	0.176	0.186
		0.8	0.004	0.168	0.035	0.143	0.068	0.153	0.106	0.164	0.136	0.186	0.168
	250	−0.8	0.987	0.828	0.597	0.407	0.260	0.166	0.107	0.075	0.055	0.044	0.035
		−0.5	0.545	0.227	0.121	0.081	0.067	0.063	0.062	0.066	0.070	0.071	0.074
		0.0	0.058	0.059	0.061	0.064	0.066	0.072	0.076	0.079	0.084	0.084	0.088
		0.5	0.008	0.110	0.045	0.078	0.064	0.076	0.075	0.082	0.083	0.090	0.095
		0.8	0.004	0.145	0.029	0.108	0.048	0.099	0.061	0.095	0.074	0.100	0.084
DF^{GLS}	100	−0.8	0.991	0.843	0.624	0.455	0.322	0.243	0.187	0.157	0.129	0.114	0.104
		−0.5	0.605	0.263	0.147	0.108	0.089	0.087	0.081	0.079	0.076	0.078	0.075
		0.0	0.073	0.070	0.072	0.074	0.072	0.072	0.070	0.070	0.071	0.074	0.073
		0.5	0.010	0.124	0.046	0.081	0.063	0.075	0.074	0.077	0.078	0.075	0.071
		0.8	0.005	0.150	0.030	0.107	0.043	0.089	0.054	0.082	0.059	0.080	0.064
	250	−0.8	0.985	0.823	0.613	0.449	0.332	0.247	0.194	0.155	0.128	0.109	0.099
		−0.5	0.535	0.227	0.123	0.084	0.068	0.064	0.062	0.062	0.061	0.058	0.060
		0.0	0.057	0.054	0.054	0.053	0.056	0.055	0.058	0.058	0.056	0.055	0.054
		0.5	0.009	0.103	0.043	0.070	0.055	0.061	0.057	0.058	0.057	0.060	0.060
		0.8	0.004	0.133	0.027	0.090	0.039	0.074	0.046	0.068	0.049	0.065	0.051

AR Case	T	ρ	$k = 0$	1	2	3	4	5	6	7	8	9	10
$\overline{Z}_\alpha^{GLS}$	100	−0.8	0.802	0.286	0.286	0.295	0.300	0.308	0.314	0.322	0.328	0.340	0.346
		0.8	0.000	0.102	0.115	0.135	0.143	0.158	0.172	0.191	0.201	0.217	0.226
	250	−0.8	0.737	0.170	0.171	0.171	0.173	0.176	0.179	0.182	0.185	0.184	0.183
		0.8	0.000	0.069	0.071	0.077	0.082	0.084	0.090	0.094	0.093	0.099	0.100
$\overline{M}Z_\alpha^{GLS}$	100	−0.8	0.784	0.027	0.035	0.050	0.065	0.081	0.089	0.113	0.120	0.140	0.149
		0.8	0.000	0.100	0.114	0.134	0.142	0.157	0.170	0.190	0.199	0.216	0.225
	250	−0.8	0.730	0.043	0.041	0.047	0.048	0.050	0.058	0.062	0.064	0.071	0.071
		0.8	0.000	0.069	0.071	0.077	0.082	0.084	0.090	0.094	0.092	0.099	0.100
DF^{GLS}	100	−0.8	0.791	0.073	0.074	0.077	0.079	0.080	0.076	0.081	0.075	0.083	0.078
		0.8	0.000	0.066	0.070	0.069	0.068	0.069	0.071	0.072	0.071	0.077	0.079
	250	−0.8	0.718	0.057	0.054	0.056	0.054	0.055	0.057	0.057	0.057	0.059	0.058
		0.8	0.000	0.059	0.061	0.063	0.062	0.060	0.060	0.060	0.058	0.058	0.054

TABLE II.B

EXACT SIZE OF THE TESTS AT SELECTED VALUES OF k: $p = 1$

MA Case	T	θ	k=0	1	2	3	4	5	6	7	8	9	10
$\overline{Z}_\alpha^{GLS}$	100	−0.8	1.000	0.994	0.965	0.940	0.918	0.906	0.893	0.888	0.882	0.882	0.883
		−0.5	0.810	0.470	0.340	0.299	0.290	0.309	0.323	0.358	0.376	0.413	0.450
		0.0	0.053	0.076	0.100	0.136	0.169	0.222	0.259	0.313	0.346	0.400	0.430
		0.5	0.001	0.169	0.059	0.163	0.155	0.226	0.253	0.303	0.343	0.397	0.432
		0.8	0.001	0.219	0.023	0.229	0.091	0.284	0.190	0.350	0.297	0.422	0.389
	250	−0.8	1.000	0.992	0.960	0.927	0.899	0.875	0.859	0.848	0.839	0.833	0.830
		−0.5	0.801	0.404	0.253	0.202	0.181	0.174	0.176	0.179	0.188	0.199	0.209
		0.0	0.047	0.057	0.064	0.077	0.081	0.094	0.104	0.113	0.129	0.142	0.154
		0.5	0.002	0.127	0.037	0.092	0.071	0.097	0.095	0.120	0.122	0.143	0.153
		0.8	0.001	0.176	0.014	0.129	0.038	0.131	0.068	0.139	0.099	0.162	0.136
$\overline{M}Z_\alpha^{GLS}$	100	−0.8	1.000	0.925	0.592	0.345	0.220	0.185	0.159	0.163	0.163	0.185	0.183
		−0.5	0.750	0.252	0.104	0.093	0.099	0.132	0.159	0.204	0.234	0.285	0.317
		0.0	0.032	0.055	0.080	0.115	0.152	0.206	0.243	0.300	0.332	0.390	0.419
		0.5	0.001	0.161	0.054	0.155	0.148	0.219	0.246	0.299	0.338	0.392	0.427
		0.8	0.001	0.211	0.021	0.224	0.087	0.277	0.186	0.346	0.294	0.417	0.386
	250	−0.8	1.000	0.978	0.800	0.514	0.276	0.147	0.087	0.058	0.048	0.040	0.036
		−0.5	0.786	0.311	0.126	0.074	0.057	0.059	0.065	0.069	0.082	0.100	0.106
		0.0	0.040	0.049	0.057	0.070	0.077	0.088	0.099	0.108	0.124	0.136	0.149
		0.5	0.002	0.124	0.034	0.090	0.069	0.095	0.094	0.117	0.121	0.142	0.150
		0.8	0.001	0.174	0.014	0.128	0.036	0.130	0.066	0.137	0.098	0.160	0.135
DF^{GLS}	100	−0.8	1.000	0.977	0.815	0.600	0.415	0.293	0.203	0.167	0.125	0.116	0.098
		−0.5	0.838	0.375	0.176	0.112	0.087	0.084	0.074	0.070	0.068	0.069	0.065
		0.0	0.070	0.072	0.067	0.072	0.069	0.073	0.072	0.071	0.069	0.068	0.061
		0.5	0.003	0.160	0.045	0.091	0.064	0.073	0.062	0.066	0.061	0.067	0.062
		0.8	0.002	0.209	0.021	0.127	0.036	0.099	0.046	0.085	0.053	0.076	0.061
	250	−0.8	1.000	0.984	0.875	0.685	0.504	0.362	0.259	0.194	0.150	0.120	0.099
		−0.5	0.809	0.352	0.161	0.101	0.072	0.061	0.059	0.053	0.053	0.056	0.052
		0.0	0.053	0.055	0.055	0.056	0.056	0.054	0.054	0.054	0.055	0.055	0.054
		0.5	0.002	0.126	0.034	0.071	0.048	0.057	0.053	0.054	0.052	0.056	0.050
		0.8	0.001	0.172	0.015	0.098	0.024	0.075	0.032	0.063	0.038	0.058	0.042

AR Case	T	ρ	k=0	1	2	3	4	5	6	7	8	9	10
$\overline{Z}_\alpha^{GLS}$	100	−0.8	0.960	0.497	0.509	0.526	0.549	0.574	0.593	0.616	0.640	0.669	0.690
		0.8	0.000	0.123	0.162	0.213	0.252	0.312	0.351	0.410	0.441	0.489	0.514
	250	−0.8	0.956	0.319	0.327	0.330	0.335	0.345	0.352	0.360	0.369	0.375	0.379
		0.8	0.000	0.070	0.077	0.089	0.099	0.119	0.129	0.150	0.159	0.176	0.183
$\overline{M}Z_\alpha^{GLS}$	100	−0.8	0.941	0.001	0.008	0.030	0.058	0.104	0.140	0.191	0.233	0.290	0.329
		0.8	0.000	0.121	0.160	0.211	0.249	0.309	0.349	0.408	0.439	0.487	0.513
	250	−0.8	0.952	0.011	0.015	0.020	0.028	0.037	0.050	0.062	0.076	0.092	0.106
		0.8	0.000	0.070	0.076	0.089	0.099	0.119	0.128	0.150	0.159	0.175	0.183
DF^{GLS}	100	−0.8	0.967	0.069	0.067	0.067	0.060	0.068	0.059	0.065	0.063	0.065	0.065
		0.8	0.000	0.072	0.067	0.070	0.071	0.069	0.069	0.074	0.066	0.072	0.071
	250	−0.8	0.959	0.052	0.051	0.052	0.050	0.049	0.050	0.050	0.051	0.052	0.053
		0.8	0.000	0.057	0.055	0.057	0.054	0.055	0.058	0.056	0.055	0.059	0.058

Reported in Table II are the sizes of the tests for $p = 0, 1$ at selected values of θ and ρ. We report results for $T = 100$ and 250 to highlight the fact that the size issue in the negative moving average case is not a small sample problem.

Several features of the results for MA errors are of note. First, for a θ of the same absolute value, a negative θ always requires a larger lag to obtain a more accurate size. Second, for a positive θ, the size of the tests are significantly better when k is even. Third, the larger the number of deterministic terms, the more distant are the exact from the nominal sizes. Fourth, there is always a value of k that will yield an $\overline{M}Z_\alpha^{GLS}$ with a reasonably accurate size when $\theta = -0.80$. For example, for $p = 0$ when $T = 100$ and 250, exact sizes of .077 and .055, respectively, are achievable with lags of 9 and 8. In contrast, there does not exist a k that will result in an acceptable size for $\overline{Z}_\alpha^{GLS}$, and increasing the sample size will not reduce the size distortions adequately. For the sample sizes reported, the exact sizes for $\overline{Z}_\alpha^{GLS}$ are well above 0.5 at $\theta = -.8$, and size distortions are non-negligible even when $\theta = -0.5$. Fifth, the DF^{GLS} also exhibits a dependence on k. At $T = 100$, the size is .104 at $k = 10$, but is .624 when $k = 2$.

For AR errors, size discrepancies between $\overline{Z}_\alpha^{GLS}$ and $\overline{M}_\alpha^{GLS}$ also exist when ρ is large and negative, albeit not as dramatic as in the MA case. The results reveal that the $\overline{M}Z_\alpha^{GLS}$ is inferior to the DF^{GLS} when $p = 1$, the sample size is small, and $|\rho|$ is large. For large negative AR coefficients, the $\overline{M}Z_\alpha^{GLS}$ has a more accurate size at k greater than 1. For large positive AR coefficients, the $\overline{M}Z_\alpha^{GLS}$ is oversized for small T, though the size improves as T increases.

Clearly, the choice of k is crucial especially in the case of negative moving average errors. Various practical solutions have been used against this problem. In ERS, the BIC is used to select k but they set the lower bound to be 3, because even larger size distortions would have resulted if zero was the lower bound. An alternative method is the sequential t test for the significance of the last lag considered in Ng and Perron (1995). It has the ability to yield higher k's than the BIC when there are negative moving-average errors and reduce size distortions. But, the sequential test tends to overparameterize in other cases. This, as does ERS's implementation of the BIC, leads to less efficient estimates and subsequently to power losses. Neither approach is fully satisfactory.

The AIC and the BIC belong to the class of information based rules (hereafter, IC) where the chosen value of k is $k_{ic} = \arg\min_{k \in [0, \ldots, k_{\max}]} IC(k)$ where[4]

$$(10) \qquad IC(k) = \ln(\hat{\sigma}_k^2) + (k + p + 1)C_T/(T - k_{\max}),$$

$\hat{\sigma}_k^2 = (T - k_{\max})^{-1} \sum_{t=k_{\max}+1}^{T} \hat{e}_{tk}^2$, $C_T/T \to 0$ as $T \to \infty$, and $C_T > 0$. The various criteria differ in C_T, the weight applied to overfitting, but all use k as the penalty to overfitting. We argue below that, with integrated data, this penalty may be a poor approximation to the cost of underfitting.

---

[4] See Ng and Perron (2000) for an analysis pertaining to the adequacy of the particular formulation of the information criterion stated.

5. AN ALTERNATIVE PENALTY FOR INTEGRATED DATA

To motivate the main idea, we consider the derivation of the AIC for data generated by a finite order $AR(k_0)$ with normal errors and a unit root (i.e., the DGP is (2) with $k = k_0$, $\beta_0 = 0$, and $e_{tk} = e_t \sim$ i.i.d. $N(0, 1)$). The family of parametric models fitted is given by (8). For notation, we let $\beta^0(k) = (0, \beta_1, \dots, \beta_k)'$, $\hat{\beta}(k) = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)'$, $\beta^0_{-0}(k) = (\beta_1, \dots, \beta_k)'$, and $\hat{\beta}_{-0}(k) = (\hat{\beta}_1, \dots, \hat{\beta}_k)'$. In what follows, we adopt the treatment of Gourieroux and Monfort (1995, pp. 307–309). The goal is to select an autoregressive order between 0 and some upper bound $k_{\max}$. We assume this upper bound satisfies $k_{\max}/T \to 0$ as $T \to \infty$. Let $f(\Delta y \mid \beta^0(k))$ be the likelihood function of the data $(\Delta y_{k_{\max}+1}, \dots, \Delta y_T)$ conditional on the initial observations $(y_0, \dots, y_{k_{\max}})$. This ensures that each competing model is evaluated with the same number of effective observations, namely $T - k_{\max}$. The Kullback distance between the true probability distribution and the estimated parametric model is $Q = E_0[\ln(f(\Delta y \mid \beta^0(k))) - \ln(f(\Delta y \mid \hat{\beta}(k)))]$ with sample analog:

$$\widetilde{Q} = (T - k_{\max})^{-1} \sum_{t=k_{\max}+1}^{T} \ln(f(\Delta y_t \mid \beta^0(k)))$$

$$- (T - k_{\max})^{-1} \sum_{t=k_{\max}+1}^{T} \ln(f(\Delta y_t \mid \hat{\beta}(k))).$$

As discussed in Gourieroux-Monfort (1995), Akaike's suggestion was to find a Q^* such that $\lim_{T \to \infty} E[T(Q - Q^*)] = 0$ so that Q^* is unbiased for Q to order T^{-1}. Let $X_t = (\tilde{y}_{t-1}, Z_t)$ with $Z_t = (\Delta \tilde{y}_{t-1}, \dots, \Delta \tilde{y}_{t-k})$ and

$$\Phi_T(k) = (1/\hat{\sigma}_k^2)(\hat{\beta}(k) - \beta^0(k))' \sum_{t=k_{\max}+1}^{T} X_t X_t'(\hat{\beta}(k) - \beta^0(k)),$$

where $\hat{\sigma}_k^2 = (T - k_{\max})^{-1} \sum_{t=k_{\max}+1}^{T} \hat{e}_{tk}^2$. Using Taylor series expansions, we have $TQ = \Phi_T(k)/2 + o_p(1)$ and $T\widetilde{Q} = -\Phi_T(k)/2 + o_p(1)$. Since $T(Q - \widetilde{Q}) = \Phi_T(k) + o_p(1)$, $\lim_{T \to \infty} E[T(Q - Q^*)] = 0$ if $Q^* = \widetilde{Q} + \Phi_T(k)$ and the remainder term is uniformly integrable. Now consider the term $\Phi_T(k)$ in the context of integrated data. Given the asymptotic block diagonality of the matrix $D_T \sum_{t=k_{\max}+1}^{T} X_t X_t' D_T$ with $D_T = \text{diag}((T - k_{\max})^{-1}, (T - k_{\max})^{-1/2}, \dots, (T - k_{\max})^{-1/2})$, we have

$$(11) \quad \Phi_T(k) = \left(\frac{1}{\hat{\sigma}_k^2}\right) \hat{\beta}_0^2 \sum_{t=k_{\max}+1}^{T} \tilde{y}_{t-1}^2$$

$$+ \left(\frac{1}{\hat{\sigma}_k^2}\right)(\hat{\beta}_{-0}(k) - \beta^0_{-0}(k))' \sum_{t=k_{\max}+1}^{T} Z_t Z_t'(\hat{\beta}_{-0}(k) - \beta^0_{-0}(k)) + o_p(1)$$

$$= (1/\hat{\sigma}_k^2) \hat{\beta}_0^2 \sum_{t=k_{\max}+1}^{T} \tilde{y}_{t-1}^2 + \chi_k^2 + o_p(1)$$

where χ_k^2 is a chi-square random variable with k degrees of freedom that is asymptotically independent of the first term. Hence, assuming the remainder term is uniformly integrable, a Q^* that will satisfy $\lim_{T\to\infty} E[T(Q-Q^*)]=0$ is

$$Q^* = (T-k_{\max})^{-1} \sum_{t=k_{\max}+1}^{T} \ln(f(\Delta y_t|\beta^0(k)))$$

$$-(T-k_{\max})^{-1} \sum_{t=k_{\max}+1}^{T} \ln(f(\Delta y_t|\hat\beta(k)))$$

$$+\left[(1/\hat\sigma_k^2)\hat\beta_0^2 \sum_{t=k_{\max}+1}^{T} \tilde y_{t-1}^2 + k\right].$$

Since the first term is common to all models and the second term is proportional to $-((T-k_{\max})/2)\ln(\hat\sigma_k^2)$ under normality, minimizing Q^* is equivalent to minimizing

$$(12) \qquad MAIC(k) = \ln(\hat\sigma_k^2) + \frac{2(\tau_T(k)+k)}{T-k_{\max}},$$

where $\tau_T(k) = (\hat\sigma_k^2)^{-1}\hat\beta_0^2 \sum_{t=k_{\max}+1}^{T} \tilde y_{t-1}^2$ and $\hat\sigma_k^2 = (T-k_{\max})^{-1} \sum_{t=k_{\max}+1}^{T} \hat e_{tk}^2$. Note that the same result holds in the general case where the data are generated by (1). The important step is the relation given by (11), which can be shown to hold in the general case, in the sense that the remainder term is $o_p(1)$ uniformly in k for $0 < k \le K_T$, where $K_T = o(T^{1/3})$. This follows using the results of Berk (1974) and Said and Dickey (1984) provided $k \to \infty$, $k^3/T \to 0$, and $T^{-1/2}\sum_{j=k+1}^{\infty}|\beta_j| \to 0$.

As a natural generalization of the $MAIC(k)$, we propose a class of *Modified Information Criteria (MIC)* that selects k as $k_{mic} = \arg\min_k MIC(k)$ where

$$(13) \qquad MIC(k) = \ln(\hat\sigma_k^2) + \frac{C_T(\tau_T(k)+k)}{T-k_{\max}},$$

with $C_T > 0$ and $C_T/T \to 0$ as $T \to \infty$. The $MBIC$ obtains with $C_T = \ln(T-k_{\max})$.

There are two important elements that distinguish (13) from standard information criteria. The first is the imposition of the null hypothesis $\beta_0 = 0$. This idea of imposing the null hypothesis in model selection when the ultimate objective is hypothesis testing appears new and may have implications beyond the unit root issue.[5] This is an interesting avenue of research, but is beyond the scope of this paper. The second element is that we retain a term in the penalty function that is stochastic. Now, as $T \to \infty$, $(1/\hat\sigma_k^2)\hat\beta_0^2 \sum_{t=k_{\max}+1}^{T} \tilde y_{t-1}^2 \Rightarrow (\sigma^2/\sigma_e^2)[\widetilde W(1)^2 - 1]^2/[2\int_0^1 \widetilde W(r)^2 dr] \equiv \tau(\widetilde W)$, say, where $\widetilde W$ is a detrended Wiener process, whose exact form depends on the deterministic components and the method of detrending. In principle, we could use the mean of $\tau(\widetilde W)$, which is independent of k, instead of $\tau_T(k)$, to construct Q^* and the objective function would then reduce

---

[5] We thank a referee for bringing this feature to our attention.

to the standard AIC. A central theme of our argument for retaining $\tau_T(k)$ is that, unless both T and k are very large, it varies substantially with k, especially when a negative MA component is present. In a sense, the imposition of the null hypothesis, $\beta_0 = 0$, allows us to avoid using asymptotic expected values to approximate the penalty factor. We can therefore hope that $\tau_T(k)$ will better capture the relevant cost of selecting different orders in finite samples since it depends not only on k, but also on the nature of the deterministic components and the detrending procedure. Hence, the penalty term will also depend on these factors.

To get an idea of the dependence of $\tau_T(k)$ on k, Figure 1 graphs the average over 5,000 replications of $\tau_T(k)$ against k for $p = 0, 1$ using the DGP $\Delta y_t = e_t + \theta e_{t-1}$ with $e_t \sim$ i.i.d. $N(0, 1)$. For positive θ, $\tau_T(k)$ is indeed fairly invariant to k (as is the case for AR processes, not reported). But, there is a clear inverse and nonlinear relation between $\tau_T(k)$ and k for negative values of θ. This relation exists even when θ is $-.5$ for both $p = 0$ and $p = 1$. The basic reason for the dependence of $\tau_T(k)$ on k is that the bias in the sum of the estimated autoregressive coefficients (i.e., $\hat{\beta}_0$) is very high for small values of k. This suggests that $\tau_T(k)$ will play a role in determining the appropriate k when there are negative
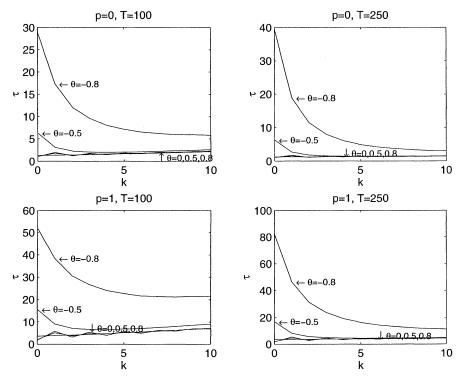

FIGURE 1: $\tau_T(k)$.

moving-average errors. To make this claim precise, the next subsection uses a local asymptotic framework to analyze the properties of $\tau_T(k)$ and the *MIC*.

6. LOCAL ASYMPTOTIC PROPERTIES

As discussed above, $\tau_T(k)$ will likely influence the choice of k when there is a large negative moving-average root, a situation for which standard unit root tests exhibit severe size distortions. To understand the issue of lag length selection in this setting, a useful device is an asymptotic framework in which the moving average coefficient is parameterized as local to -1.

$$(14) \qquad y_t - \alpha_T y_{t-1} = e_t + \theta_T e_{t-1},$$

$$\alpha_T = 1 + c/T, \qquad \theta_T = -1 + \delta/\sqrt{T},$$

$e_t \sim$ i.i.d. $(0, \sigma_e^2)$, $y_0 = e_0 = 0$, $\delta > 0$, and $c < 0$. As T increases to infinity, $\alpha_T \to 1, \theta_T \to -1$, and y_t becomes a white noise process. However, y_t is a *nearly integrated nearly white noise process* in finite samples. Nabeya and Perron (1994) used this asymptotic framework to derive approximations to the finite sample properties of the least-squares estimator in a first-order autoregressive model with negative moving-average errors and showed these approximations not only to be superior to those obtained from standard asymptotic distributions but also to be quite accurate even for small sample sizes. Perron and Ng (1996) used the same setup to analyze the size and power of various unit root tests based upon augmented autoregressions. Our aim here is to use the same device to first obtain useful qualitative features about the relative properties of the *IC* and *MIC*, and second, to guide us to an appropriate choice of the penalty weight C_T.

6.1. *Local Asymptotic Properties of $\tau_T(k)$*

The following Theorem (proved in the Appendix) summarizes the general properties of $\tau_T(k)$ in this local asymptotic framework.

THEOREM 2: *Let* y_t *be generated by* (14), $Z_1 = (1 + \delta^2 \int_0^1 J_c(r)^2 dr)/ [\delta^2 \int_0^1 J_c(r)^2 dr]^2$ *when* $p = 0$, *and* $Z_1 = (1 + \delta^2 \int_0^1 V_{c,\bar{c}}(r)^2 dr)/[\delta^2 \int_0^1 V_{c,\bar{c}}(r)^2 dr]^2$ *when* $p = 1$. *Let* $k \to \infty, K_{1T} = o(T^{1/2}), T^{1/2}/K_{2T} = o_p(1)$, *and* $K_{3T} = o(T)$. *Then,* (i) *uniformly in* $0 < k \le K_{1T}$: $(k^2/T)\tau_T(k) \Rightarrow Z_1$; (ii) *if* $k/T^{1/2} \to \kappa, \tau_T(k) \Rightarrow \vartheta(\kappa)^2 Z_2$, *where* $E(Z_2) = E(Z_1)$ *and* $\vartheta(\kappa) = 2\delta \exp(-2\kappa\delta)/(1 - \exp(-2\kappa\delta))$; *and* (iii) *uniformly in* $K_{2T} \le k \le K_{3T}, \tau_T(k) \Rightarrow Z_3$, *where* Z_3 *is* $O_p(1)$.

Theorem 2 shows that the local limit of $\tau_T(k)$ is unbounded unless k increases at least at rate $T^{1/2}$. Since $\tau_T(0)$ diverges at rate T (see Nabeya and Perron (1994)), and $\tau_T(k)$ diverges at rate T/k^2 if $k^2/T \to 0$, the cost to overfitting is not linear in k as is assumed for the *IC*, and Figure 1 bears this out. Theorem 2 also highlights the fact that $\tau_T(k)$ will depend, via Z_1, on the nature of the deterministic components and the method of detrending.

Given Theorem 2 and the fact that $\hat{\sigma}_k^2$ is consistent for σ_e^2 irrespective of the rate of increase of k (shown in the Appendix), it can be shown that, in this local asymptotic framework, the limiting distribution of the DF test diverges to $-\infty$ unless k grows at least at rate $T^{1/2}$.[6] Hence, the unit root will always be rejected if standard critical values are used. This explains why the DF test tends to be oversized when k is small. In Perron and Ng (1996, 1998), we showed that $k^2 s_{AR}^2 = O_p(1)$ but we also need $T s_{AR}^2$ to be $O_p(1)$ for the M tests to be bounded under the null hypothesis in this local framework. But $k^2 s_{AR}^2$ and $T s_{AR}^2$ can both be bounded only if $k/T^{1/2} \to \kappa$ for some $0 < \kappa < \infty$. In the next section, we will consider model selection procedures with this requirement in mind.

6.2. Local Asymptotic Properties of the MIC and IC

In the standard asymptotic framework, with data generated by (1), $\tau_T(k)$ is invariant to k for T and k large. It follows that k_{mic} grows at a logarithmic rate from Hannan and Deistler (1988). We now consider the properties of k chosen by the MIC and the IC in the local asymptotic framework with the additional technical requirement that $C_T^2/T \to 0$. Using Theorem 2 and an expansion for $\ln(\hat{\sigma}_k^2)$ derived in the Appendix, we have the following expansions for the MIC and IC.

LEMMA 1: *Let y_t be generated by* (14) *and K_{1T}, K_{2T}, and K_{3T} as defined in Theorem 2; then* (i) *uniformly in $0 < k \le K_{1T}$:*

$$(15) \qquad MIC(k) = \ln(\sigma_e^2) + \frac{1}{k} + \frac{C_T}{k^2} Z_1 + C_T \frac{k}{T} + O_p(T^{-1/2});$$

(ii) *for $k/T^{1/2} \to \kappa$, with $\vartheta(\kappa)$ defined in Theorem 2:*

$$(16) \qquad MIC(k) = \ln(\sigma_e^2) + \frac{\vartheta(\kappa)}{\sqrt{T}} + \frac{C_T}{T} \vartheta(\kappa)^2 Z_2 + C_T \frac{k}{\sqrt{T}} + O_p(T^{-1/2});$$

(iii) *uniformly in $K_{2T} \le k \le K_{3T}$:*

$$(17) \qquad MIC(k) = \ln(\sigma_e^2) + \frac{2\delta}{\sqrt{T} \exp(2k\delta/\sqrt{T})} + \frac{C_T}{T} Z_3 + C_T \frac{k}{T} + O_p(T^{-1/2}).$$

The expansions for the $IC(k)$ are the same except that the terms involving the Z's are absent.

The objects of ultimate interest are the values k_{ic} and k_{mic} chosen by the IC and the MIC, respectively, over all possible values of k.

---

[6] This is a more general result than Pantula (1991) who used a different local asymptotic framework for modeling the negative moving average component in unit root tests. Pantula's result applies only when $k = O_p(T^{1/4})$.

THEOREM 3: *Let y_t be generated by* (14) *and consider the rate of increase of k as $T \to \infty$. Define*:

$$F_{MIC(1)}(\kappa) = \frac{-4\delta^2 \exp(-2\kappa\delta)}{\sqrt{T}[1 - \exp(-2\kappa\delta)]^2}$$

$$+ C_T \left(\frac{-16\delta^3}{T} \frac{\exp(-2\kappa\delta)^2 Z_2}{[1 - \exp(-2\kappa\delta)]^3} + \frac{1}{\sqrt{T}} \right);$$

$$F_{MIC(2)}(\kappa) = -\frac{\sqrt{C_T/T}}{\kappa^2} + \sqrt{C_T/T} - \frac{2C_T^2}{\kappa^3 T} Z_1;$$

$$F_{IC}(\kappa) = \frac{-4\delta^2 \exp(-2\kappa\delta)}{[1 - \exp(-2\kappa\delta)]^2} + C_T.$$

When C_T is fixed, $k_{mic}/(\kappa_{mic(1)} T^{1/2}) \to 1$ where $\kappa_{mic(1)}$ solves $F_{MIC(1)}(\kappa) = 0$, and $(k_{ic}/\kappa_{ic} T^{1/2}) \to 1$, where κ_{ic} solves $F_{IC}(\kappa) = 0$. When $C_T \to \infty, k_{ic}/\sqrt{T/C_T} \to 1$. When $C_T \to \infty$ and $C_T^3/T \to d(\geq 0), k_{mic}/(\kappa_{mic(2)} \sqrt{T/C_T}) \to 1$, where $\kappa_{mic(2)}$ solves $F_{MIC(2)}(\kappa) = 0$. If $C_T \to \infty$ and $C_T^3/T \to \infty, k_{mic}/(2Z_1 T)^{1/3} \to 1$.

Provided C_T is constant (as is the case with Akaike's value of 2), both the *IC* and the *MIC* will choose k proportional to $T^{1/2}$ asymptotically, much faster than the logarithmic rate attained under the standard asymptotic framework. The rate of $\sqrt{T}$ is optimal for unit root tests because this is the rate that will render the class of *DF* as well as the *M* tests nondivergent under the null hypothesis in this local asymptotic framework. However, when C_T increases with T, this optimal rate is not achieved.

The proportionality factor for the rate of increase of k_{ic} when C_T increases with T is one and highlights a major shortcoming of the *IC* in such cases, namely a complete independence to δ, the parameter of the local asymptotic framework. For the k_{mic} when C_T increases with T, the proportionality factor depends on the parameters of the model via Z_1. This effect, however, still vanishes as T increases. This lack of flexibility does not arise when C_T is fixed since κ_{mic} and κ_{ic} always maintain a dependence on δ. Although k_{ic} increases at the appropriate rate of $\sqrt{T}$, it too has drawbacks because κ_{ic} is invariant to T and it is the same whether $p = 0$ or $p = 1$. In contrast, κ_{mic} adapts to the sample size and the trend function via a second order factor involving Z_1. Thus, the *MIC* with C_T fixed maintains a proportionality factor that is the most flexible with respect to T, the regression model, as well as parameters of the DGP.

If the weight C_T on the penalty factor is too large, that is $C_T^3/T \to \infty$, the *MIC* chooses a rate that is independent of C_T, namely $T^{1/3}$, while the *IC* continues to select a k that is proportional to $\sqrt{T/C_T}$. This, however, is a result of theoretical rather than practical interest since almost all information criteria that have been suggested are such that $C_T^3/T \to 0$.

To illustrate these properties, we consider the *MAIC* and *AIC* with $C_T = 2$ and the *MBIC* and *BIC* with $C_T = \ln(T)$. We denote the proportionality factors by $\kappa_{maic}, \kappa_{aic}, \kappa_{mbic}$, and κ_{bic}, respectively. Numerical values of κ are reported in

TABLE III.A

VALUES OF THE PROPORTIONALITY FACTOR κ

$T\backslash\delta$		$p = 0$						$p = 1$					
		0.50	1.00	1.50	2.00	5.00	8.00	0.50	1.00	1.50	2.00	5.00	8.00
100	bic	0.99	0.96	0.93	0.89	0.67	0.54	0.99	0.96	0.93	0.89	0.67	0.54
	aic	0.69	0.65	0.61	0.57	0.39	0.30	0.69	0.65	0.61	0.57	0.39	0.30
	mbic	3.59	1.73	1.24	1.04	0.68	0.54	6.34	2.89	1.87	0.40	0.71	0.54
	maic	1.70	0.88	0.68	0.59	0.39	0.30	2.96	1.37	0.90	0.70	0.39	0.30
150	bic	0.99	0.96	0.93	0.90	0.69	0.55	0.99	0.96	0.93	0.90	0.69	0.55
	aic	0.69	0.65	0.61	0.57	0.39	0.30	0.69	0.65	0.61	0.57	0.39	0.30
	mbic	3.58	1.73	1.24	1.04	0.69	0.55	6.41	2.92	1.88	1.42	0.72	0.56
	maic	1.63	0.85	0.67	0.59	0.39	0.30	2.87	1.32	0.88	0.69	0.39	0.30
250	bic	0.99	0.97	0.94	0.90	0.70	0.57	0.99	0.67	0.94	0.90	0.70	0.57
	aic	0.69	0.65	0.61	0.57	0.39	0.30	0.69	0.65	0.61	0.57	0.39	0.30
	mbic	3.55	1.71	1.24	1.04	0.71	0.57	6.46	2.93	1.89	1.42	0.73	0.57
	maic	1.55	0.82	0.66	0.58	0.39	0.30	2.75	1.27	0.85	0.67	0.39	0.30
500	bic	0.99	0.97	0.94	0.91	0.72	0.58	0.99	0.97	0.94	0.91	0.72	0.58
	aic	0.69	0.65	0.61	0.57	0.39	0.30	0.69	0.65	0.61	0.57	0.39	0.30
	mbic	3.47	1.68	1.23	1.04	0.72	0.58	6.46	2.92	1.89	1.42	0.75	0.59
	maic	1.44	0.79	0.65	0.58	0.39	0.30	2.60	1.20	0.81	0.65	0.39	0.30
5000	bic	0.99	0.98	0.96	0.93	0.76	0.63	0.99	0.98	0.96	0.93	0.76	0.63
	aic	0.69	0.65	0.61	0.57	0.39	0.30	0.69	0.65	0.61	0.57	0.39	0.30
	mbic	3.06	1.52	1.16	1.02	0.76	0.63	6.12	2.74	1.78	1.36	0.78	0.63
	maic	1.14	0.71	0.62	0.57	0.39	0.30	2.12	1.00	0.72	0.60	0.39	0.30
50000	bic	0.99	0.98	0.96	0.94	0.79	0.66	0.99	0.98	0.96	0.94	0.79	0.66
	aic	0.69	0.65	0.61	0.57	0.39	0.30	0.69	0.65	0.61	0.57	0.39	0.30
	mbic	2.56	1.34	1.08	0.99	0.79	0.67	5.43	2.42	1.59	1.25	0.80	0.66
	maic	0.92	0.67	0.61	0.57	0.39	0.30	1.69	0.85	0.66	0.58	0.39	0.30

Table III.A for selected values of δ and T under the null hypothesis of a unit root with $c = 0$. In the case of the *MIC* these values are random variables because of the presence of Z_1. The numerical calculations reported use its expected value.[7] Table III.A shows first that the κ_{mic}'s are drastically larger than the κ_{ic}'s when δ is small. Second, the κ_{ic}'s vary much less with δ than the κ_{mic}'s. Third, the κ_{mic}'s are larger when $p = 1$ than when $p = 0$, but the κ_{ic}'s are invariant to the order of the deterministic terms in the regression.

Values of k selected by the *IC* and the *MIC* are in Table III.B. These are constructed using the expected value of the *MIC* given by (for the $IC(k)$ the term

[7] To get the expected value we use the following crude approximation. For $p = 0, E[Z_1] \approx$ $(E[1 + \delta^2 \int_0^1 W(r)^2 dr])/\delta^4 E([\int_0^1 W(r)^2 dr]^2) = (4 + 2\delta^2)/\delta^4$. For $p = 1$, $E[Z_1] \approx E[1 + \delta^2 \int_0^1 V_{0,\bar{c}}(r)^2 dr]/\delta^4 E(\int_0^1 V_{0,\bar{c}}(r)^2 dr)^2$. For $\bar{c} = -13.5$, tedious but straightforward algebra gives $E[Z_1] \approx (1 + \delta^2(.102475)/(1 + \delta^4(.102475)^2)$.

TABLE III.B

k Chosen by Minimizing the Asymptotic Approximation of the *MIC* and *IC* (18)

$T\backslash\delta$		$p=0$						$p=1$					
		0.50	1.00	1.50	2.00	5.00	8.00	0.50	1.00	1.50	2.00	5.00	8.00
100	bic	4	4	4	4	3	2	4	4	4	4	3	2
	aic	6	6	6	5	3	3	6	6	6	5	3	3
	mbic	16	8	5	4	3	2	29	13	8	6	3	2
	maic	17	8	6	5	3	3	29	13	9	7	3	3
150	bic	5	5	5	4	3	3	5	5	5	4	3	3
	aic	8	8	7	7	4	3	8	8	7	7	4	3
	mbic	19	9	6	5	3	3	35	15	10	7	3	3
	maic	20	10	8	7	4	3	35	16	10	8	4	3
250	bic	6	6	6	6	4	3	6	6	6	6	4	3
	aic	10	10	9	9	6	4	10	10	9	9	6	4
	mbic	23	11	8	7	4	3	43	19	12	9	4	3
	maic	24	13	10	9	6	4	43	20	13	10	6	4
500	bic	8	8	8	8	6	5	8	8	8	8	6	5
	aic	15	14	13	12	8	6	15	14	13	12	8	6
	mbic	31	15	11	9	6	5	57	26	16	12	6	5
	maic	32	17	14	13	8	6	58	27	18	14	8	6
5000	bic	24	23	23	22	18	15	24	23	23	22	18	15
	aic	49	46	43	40	27	21	49	46	43	40	27	21
	mbic	74	37	28	24	18	15	148	66	43	33	19	15
	maic	80	50	44	40	27	21	150	71	50	42	27	21
50000	bic	67	66	65	63	53	45	67	66	65	63	53	45
	aic	154	147	137	128	88	68	154	147	137	128	88	68
	mbic	174	91	74	67	54	45	369	164	108	85	54	45
	maic	207	151	138	128	88	68	379	190	147	130	88	68

with Z_1 is absent)

$$(18) \qquad E[MIC(k)] = \ln(\sigma_e^2) + \frac{(1-\theta_T^2)(-\theta_T)^{2(k+1)}}{1-(-\theta_T)^{2(k+1)}}$$

$$+ C_T\left(\left[\frac{(1-\theta_T^2)(-\theta_T)^{2k}}{1-(-\theta_T)^{2(k+1)}}\right]^2 E(Z_1) + \frac{k}{T}\right).$$

This expression is valid uniformly in $0 < k \le K_{3T}$ for $K_{3T} = o(T)$ and is obtained using equation (A.4) and Lemma A.5. When δ is small and especially when $p = 1$, the k_{mic} are several folds larger than the k_{ic}. As well, differences between k_{aic} and k_{maic} are nontrivial even for sample sizes deemed unrealistically large. Table III.B also provides some guidance for the k's that would likely be selected in practice. When $\theta = -.9$, for example, the local asymptotic results for $\delta = 1$ and $T = 100$ suggest that $k_{bic} = 4$ for both $p = 0$ and 1, but that the k_{maic} are 8 and 13, respectively. The local asymptotic results therefore support the earlier claim that the *IC* lacks the flexibility to handle DGP's with a large negative moving

average root and/or alternative deterministic components. The issue of practical interest is whether in finite samples and for realistic values of the moving average parameter, such large discrepancies remain and what are the implications for unit root tests. The rest of the analysis presents evidence to this effect.

6.3. The Finite Sample Properties of k_{mic}

To evaluate the finite sample properties of the MIC vis-a-vis the IC, we performed the following simulation experiment. For a given DGP, we construct the $\overline{M}Z_{\alpha}^{GLS}$ and DF^{GLS} tests at each $k \in [0, 10]$, and record the exact sizes. We then find the so-called optimal values of k^*, denoted k_{mz}^* and k_{dfgls}^*, as the first k with a size closest to within .03 of the nominal size of 0.05. If no such k exists, k^* is the k for which the exact size of the test has the smallest absolute deviation from the nominal size of 5%. We then obtain k_{bic} as the median value selected by the BIC over the range 0 and 10. A similar procedure is used to obtain k_{mbic}, k_{aic}, and k_{maic}. The setup for the simulations is as outlined in Section 4.

In Table IV.A ($p = 0$) and IV.B ($p = 1$), we report the values for $k_{mz}^*, k_{dfgls}^*, k_{ic}$, and k_{mic} for selected parameters for $C_T = 2$ and $C_T = \ln(T - k_{max})$. While the k_{ic}'s are in the same range as the k^*'s for many cases, there are important differences between the two when $\theta < 0$. The k_{mic} is much closer to k_{mz}^* and k_{dfgls}^* in negative MA models than the k_{ic}. For example, when $\theta = -.8$, k_{mz}^* and k_{dfgls}^* are 10 at $T = 250$ and $p = 1$; the BIC chooses 3 over half the time whereas k_{maic} has 8 as the median value. The k_{aic} is closer to k_{mz}^* and k_{dfgls}^* than the k_{bic}, but still not as close as the k_{maic}. For the AR case, the k_{maic} indeed selects the appropriate order of one, but when $p = 1$, this sometimes differs from k_{mz}^*. This does not reflect a weakness of the MIC but rather that the $\overline{M}^{GLS}$ tests are undersized when the autoregressive coefficient is large and negative. The results show that k_{maic} corresponds closely to the k_{dfgls}^* in the AR case.

7. SIZE AND POWER OF THE TESTS IN FINITE SAMPLES

In this section, we need to establish two things: first, that the MIC chooses values of k that are appropriate in the sense of minimizing size distortions under the null hypothesis and not overparameterizing under the alternative; second, that the M^{GLS} and $\overline{M}^{GLS}$ statistics have power comparable to the DF^{GLS} and dominate the M tests. Whenever an estimate of σ^2 is required, s_{AR}^2 based upon the regression (8) is used. Because the $MAIC$ dominates all other criteria from both theoretical and numerical perspectives we only consider the M^{GLS} tests constructed using the $MAIC$. In the simulations, the lower bound is always zero to reduce the chance of overparameterizing when a large k is not necessary. The upper bound is $k_{max} = \text{int}(12(T/100)^{1/4})$.[8] The results are tabulated for $T = 100, 150$, and 250. The ten statistics being considered are detailed in the notes to Table V.A and Table VI.A.

---

[8] In practice, the upper bound should be set such that the constraint does not bind. In simulations, the upper bound occasionally binds when $T = 100$ but not for larger sample sizes (e.g. $T = 150$).

For $p = 0$, the results for size are reported in Table V.A. The M and M^{GLS} tests still have inflated sizes at $T = 100$ when $\theta = -0.8$. However, use of GLS detrended data to construct s^2_{AR} produces substantial size improvements. The DF^{GLS} based on the BIC yields an exact size around 0.4 at $\theta = -0.8$, but

TABLE IV.A

SELECTED k FOR $p = 0$

						MA Case: $v_t = (1 + \theta L)e_t$						
			$T = 100$						$T = 250$			
θ	k^*_{mz}	k^*_{dfgls}	k_{bic}	k_{aic}	k_{mbic}	k_{maic}	k^*_{mz}	k^*_{dfgls}	k_{bic}	k_{aic}	k_{mbic}	k_{maic}
-0.8	8	8	2	4	5	6	7	10	4	6	5	7
-0.7	6	10	2	3	3	4	6	7	3	5	4	5
-0.6	5	7	1	2	2	3	5	6	2	4	3	4
-0.5	3	7	1	2	2	2	4	4	2	3	2	3
-0.4	2	4	1	2	1	2	2	2	1	2	2	2
-0.3	2	3	1	1	1	1	2	2	1	2	1	2
-0.2	1	2	0	1	1	1	1	1	1	1	1	1
-0.1	1	1	0	0	0	0	1	1	0	1	0	1
0.0	0	0	0	0	0	0	0	0	0	0	0	0
0.1	0	0	0	0	0	0	0	0	0	1	0	1
0.2	0	0	0	1	0	1	1	1	1	1	1	1
0.3	2	2	1	1	1	1	2	1	1	2	1	2
0.4	2	2	1	2	1	2	2	2	1	2	2	2
0.5	2	2	1	2	2	2	2	2	2	3	2	3
0.6	2	2	2	3	2	3	2	2	2	4	2	4
0.7	2	2	2	4	2	4	2	2	3	5	3	5
0.8	2	2	3	5	3	5	2	2	4	7	4	7

						AR Case: $(1 - \rho L)v_t = e_t$						
			$T = 100$						$T = 250$			
ρ	k^*_{mz}	k^*_{dfgls}	k_{bic}	k_{aic}	k_{mbic}	k_{maic}	k^*_{mz}	k^*_{dfgls}	k_{bic}	k_{aic}	k_{mbic}	k_{maic}
-0.8	1	1	1	1	1	1	1	1	1	1	1	1
-0.7	1	1	1	1	1	1	1	1	1	1	1	1
-0.6	1	1	1	1	1	1	1	1	1	1	1	1
-0.5	1	1	1	1	1	1	1	1	1	1	1	1
-0.4	1	1	1	1	1	1	1	1	1	1	1	1
-0.3	1	1	1	1	1	1	1	1	1	1	1	1
-0.2	1	1	0	1	1	1	1	1	1	1	1	1
-0.1	1	1	0	0	0	0	1	1	0	1	0	1
0.0	0	0	0	0	0	0	0	0	0	0	0	0
0.1	0	0	0	0	0	0	0	0	0	1	0	1
0.2	1	0	0	1	0	1	1	1	1	1	1	1
0.3	1	1	1	1	1	1	1	1	1	1	1	1
0.4	0	1	1	1	1	1	1	1	1	1	1	1
0.5	1	1	1	1	1	1	1	1	1	1	1	1
0.6	1	1	1	1	1	1	1	1	1	1	1	1
0.7	1	1	1	1	1	1	1	1	1	1	1	1
0.8	1	1	1	1	1	1	1	1	1	1	1	1

Note: The entries are the median values of the selected k using 5,000 replications.

TABLE IV.B

SELECTED k FOR $p = 1$

MA Case: $v_t = (1 + \theta L)e_t$

			$T = 100$						$T = 250$			
θ	k^*_{mz}	k^*_{dfgls}	k_{bic}	k_{aic}	k_{mbic}	k_{maic}	k^*_{mz}	k^*_{dfgls}	k_{bic}	k_{aic}	k_{mbic}	k_{maic}
−0.8	6	10	0	2	6	6	7	10	3	5	7	8
−0.7	4	8	1	2	4	5	5	8	3	4	5	6
−0.6	4	6	1	2	3	4	4	6	2	3	4	4
−0.5	3	6	1	2	2	3	3	4	2	3	3	3
−0.4	2	3	1	1	2	2	2	3	1	2	2	2
−0.3	2	2	0	1	1	1	2	2	1	1	1	2
−0.2	1	4	0	1	1	1	1	1	1	1	1	1
−0.1	0	1	0	0	0	1	1	1	0	1	0	1
0.0	0	0	0	0	0	0	0	0	0	0	0	0
0.1	1	0	0	1	0	0	1	0	0	1	0	1
0.2	1	2	0	1	0	1	1	1	1	1	0	1
0.3	0	2	1	1	0	1	2	2	1	2	1	2
0.4	2	2	1	2	1	2	2	2	1	2	2	2
0.5	2	2	1	2	2	2	2	2	2	3	2	3
0.6	2	2	2	3	2	2	2	2	3	4	2	4
0.7	2	2	2	4	2	4	2	2	3	5	3	5
0.8	2	2	3	5	2	4	4	4	4	7	4	6

AR Case: $(1 - \rho L)v_t = e_t$

			$T = 100$						$T = 250$			
ρ	k^*_{mz}	k^*_{dfgls}	k_{bic}	k_{aic}	k_{mbic}	k_{maic}	k^*_{mz}	k^*_{dfgls}	k_{bic}	k_{aic}	k_{mbic}	k_{maic}
−0.8	3	1	1	1	1	1	4	1	1	1	1	1
−0.7	3	1	1	1	1	1	1	1	1	1	1	1
−0.6	2	1	1	1	1	1	1	1	1	1	1	1
−0.5	1	1	1	1	1	1	1	1	1	1	1	1
−0.4	1	1	1	1	1	1	1	1	1	1	1	1
−0.3	1	1	1	1	1	1	1	1	1	1	1	1
−0.2	1	1	0	1	1	1	1	1	1	1	1	1
−0.1	0	1	0	0	0	0	0	1	0	1	0	1
0.0	0	0	0	0	0	0	0	0	0	0	0	0
0.1	1	0	0	1	0	0	1	0	0	1	0	1
0.2	1	1	1	1	0	1	1	1	1	1	0	1
0.3	1	1	1	1	0	1	1	1	1	1	1	1
0.4	1	1	1	1	1	1	1	1	1	1	1	1
0.5	1	1	1	1	1	1	1	1	1	1	1	1
0.6	1	1	1	1	1	1	1	1	1	1	1	1
0.7	1	1	1	1	1	1	1	1	1	1	1	1
0.8	1	1	1	1	1	1	1	1	1	1	1	1

Note: The entries are the median values of the selected k using 5,000 replications.

the *MAIC* improves the size to 0.107. The *MAIC* also yields a substantial size improvement over $DF^{GLS}(t)$, which has a size of .35 when $\theta = -0.8$. The results also confirm that the P_T test constructed with k chosen by *BIC* suffers from sub-

TABLE V.A
SIZE OF THE TESTS; $p = 0$

MA Case: $v_t = e_t + \theta e_{t-1}$, $e_t \sim N(0,1)$											
T	θ	M_{mic}	M_{mic}^{GLS}	$\bar{M}_{mic}^{GLS}$	DF_{bic}	DF_{bic}^{GLS}	DF_t^{GLS}	DF_{mic}^{GLS}	P_{Tbic}	$\bar{P}_{Tmic}^{GLS}$	$\bar{M}P_{Tmic}^{GLS}$
100	−0.8	0.058	0.041	0.036	0.394	0.411	0.356	0.107	0.429	0.034	0.036
	−0.5	0.028	0.013	0.056	0.082	0.104	0.095	0.073	0.109	0.048	0.051
	0.0	0.018	0.005	0.043	0.072	0.077	0.072	0.050	0.094	0.036	0.040
	0.5	0.027	0.013	0.062	0.081	0.082	0.080	0.048	0.118	0.048	0.054
	0.8	0.042	0.018	0.067	0.095	0.097	0.091	0.039	0.145	0.054	0.062
150	−0.8	0.034	0.018	0.026	0.407	0.390	0.304	0.095	0.420	0.025	0.027
	−0.5	0.025	0.009	0.054	0.085	0.093	0.086	0.066	0.091	0.047	0.049
	0.0	0.017	0.005	0.049	0.062	0.075	0.070	0.051	0.085	0.040	0.046
	0.5	0.028	0.011	0.054	0.066	0.072	0.071	0.046	0.090	0.047	0.050
	0.8	0.040	0.013	0.060	0.079	0.077	0.070	0.034	0.108	0.047	0.052
250	−0.8	0.016	0.006	0.022	0.370	0.323	0.236	0.088	0.380	0.021	0.021
	−0.5	0.036	0.012	0.058	0.075	0.089	0.079	0.063	0.086	0.049	0.054
	0.0	0.021	0.006	0.047	0.059	0.062	0.062	0.045	0.069	0.039	0.042
	0.5	0.034	0.009	0.056	0.059	0.068	0.069	0.045	0.078	0.047	0.051
	0.8	0.044	0.014	0.059	0.066	0.062	0.058	0.042	0.081	0.050	0.053
AR Case: $v_t = \rho v_{t-1} + e_t$, $e_t \sim N(0,1)$											
T	ρ	M_{mic}	M_{mic}^{GLS}	$\bar{M}_{mic}^{GLS}$	DF_{bic}	DF_{bic}^{GLS}	DF_t^{GLS}	DF_{mic}^{GLS}	P_{Tbic}	$\bar{P}_{Tmic}^{GLS}$	$\bar{M}P_{Tmic}^{GLS}$
100	−0.8	0.002	0.000	0.015	0.067	0.082	0.076	0.049	0.063	0.013	0.014
	−0.5	0.010	0.003	0.037	0.067	0.076	0.071	0.046	0.087	0.028	0.033
	0.5	0.029	0.012	0.064	0.082	0.087	0.082	0.053	0.108	0.051	0.057
	0.8	0.047	0.025	0.090	0.091	0.081	0.077	0.053	0.125	0.068	0.077
150	−0.8	0.003	0.000	0.021	0.062	0.069	0.067	0.048	0.051	0.019	0.020
	−0.5	0.013	0.004	0.039	0.060	0.074	0.069	0.045	0.072	0.031	0.035
	0.5	0.030	0.010	0.055	0.063	0.069	0.066	0.047	0.081	0.046	0.051
	0.8	0.043	0.017	0.070	0.067	0.071	0.069	0.047	0.092	0.055	0.063
250	−0.8	0.007	0.001	0.029	0.056	0.057	0.054	0.040	0.049	0.026	0.028
	−0.5	0.022	0.006	0.046	0.052	0.057	0.058	0.047	0.062	0.040	0.042
	0.5	0.028	0.008	0.052	0.056	0.061	0.061	0.045	0.070	0.043	0.046
	0.8	0.037	0.010	0.060	0.060	0.061	0.063	0.051	0.073	0.051	0.056

Notes: The results reported for the M_{mic} tests are those corresponding to MZ_α. Since the numerical results for MZ_α, MZ_t, and MSB are similar, we use the label M in the tables. Three variants of the M tests are considered. The M_{mic} is the original MZ_α, the M_{mic}^{GLS} uses GLS detrended data to construct the statistic but uses least squares detrending in the estimation of s_{AR}^2, and $\bar{M}_{mic}^{GLS}$ uses GLS detrending at $\bar{c} = -7.0$ when constructing s_{AR}^2. When the BIC is used, $k_{max} = 6$ and $k_{min} = 3$ as in ERS. For the MIC, $k_{max} = \text{int}(12(T/100)^{1/4})$, $k_{min} = 0$, and $C_T = 2$, giving the $MAIC$. The DF_t^{GLS} is based on the 10% sequential t test for the significance of the last lag.

stantial size distortions. However, the $\bar{P}_T$ and $\bar{M}P_T^{GLS}$ show remarkably little size distortions when the $MAIC$ is used to select k.

To assess the relative power of the tests, we tabulate the rejection rates at $\bar{\alpha} = 1 + \bar{c}/T$. For $p = 0$ and $\bar{c} = -7.0$, the values of $\bar{\alpha}$ are 0.93, 0.953, and 0.972

TABLE V.B

SIZE-ADJUSTED POWER OF THE TESTS; $p = 0$

					MA Case: $v_t = e_t + \theta e_{t-1}, e_t \sim N(0,1)$						
T	θ	M_{mic}	M_{mic}^{GLS}	$\bar{M}_{mic}^{GLS}$	DF_{bic}	DF_{bic}^{GLS}	DF_t^{GLS}	DF_{mic}^{GLS}	P_{Tbic}	$\bar{P}_{Tmic}^{GLS}$	$\bar{M}P_{Tmic}^{GLS}$
100	−0.8	0.193	0.237	0.311	0.241	0.419	0.412	0.385	0.313	0.326	0.311
	−0.5	0.253	0.351	0.344	0.191	0.448	0.429	0.385	0.414	0.362	0.355
	0.0	0.315	0.452	0.458	0.131	0.432	0.436	0.454	0.418	0.471	0.442
	0.5	0.259	0.332	0.324	0.122	0.414	0.402	0.370	0.402	0.351	0.347
	0.8	0.180	0.257	0.264	0.134	0.366	0.335	0.343	0.348	0.296	0.281
150	−0.8	0.180	0.252	0.296	0.262	0.448	0.417	0.415	0.381	0.307	0.292
	−0.5	0.267	0.387	0.384	0.195	0.450	0.430	0.393	0.421	0.398	0.396
	0.0	0.339	0.463	0.460	0.152	0.422	0.405	0.454	0.413	0.467	0.453
	0.5	0.263	0.382	0.380	0.160	0.441	0.418	0.405	0.429	0.393	0.398
	0.8	0.233	0.346	0.363	0.147	0.382	0.362	0.404	0.387	0.377	0.374
250	−0.8	0.190	0.319	0.334	0.287	0.441	0.420	0.437	0.419	0.335	0.327
	−0.5	0.264	0.404	0.396	0.201	0.477	0.455	0.428	0.458	0.425	0.409
	0.0	0.338	0.481	0.477	0.152	0.464	0.446	0.483	0.455	0.485	0.478
	0.5	0.258	0.411	0.411	0.169	0.439	0.424	0.431	0.441	0.422	0.413
	0.8	0.246	0.385	0.395	0.143	0.432	0.413	0.419	0.434	0.406	0.405

					AR Case: $v_t = \rho v_{t-1} + e_t, e_t \sim N(0,1)$						
T	ρ	M_{mic}	M_{mic}^{GLS}	$\bar{M}_{mic}^{GLS}$	DF_{bic}	DF_{bic}^{GLS}	DF_t^{GLS}	DF_{mic}^{GLS}	P_{Tbic}	$\bar{P}_{Tmic}^{GLS}$	$\bar{M}P_{Tmic}^{GLS}$
100	−0.8	0.246	0.349	0.350	0.154	0.433	0.429	0.404	0.384	0.362	0.342
	−0.5	0.330	0.437	0.432	0.145	0.428	0.420	0.455	0.405	0.452	0.437
	0.5	0.279	0.365	0.368	0.115	0.338	0.330	0.385	0.379	0.385	0.374
	0.8	0.196	0.251	0.241	0.084	0.297	0.292	0.298	0.261	0.276	0.269
150	−0.8	0.275	0.398	0.416	0.161	0.477	0.437	0.432	0.442	0.419	0.409
	−0.5	0.332	0.453	0.454	0.161	0.404	0.390	0.468	0.423	0.463	0.458
	0.5	0.307	0.419	0.411	0.140	0.417	0.410	0.446	0.407	0.436	0.430
	0.8	0.232	0.333	0.330	0.103	0.341	0.324	0.362	0.325	0.349	0.353
250	−0.8	0.325	0.479	0.472	0.157	0.485	0.480	0.493	0.468	0.477	0.471
	−0.5	0.300	0.452	0.447	0.179	0.493	0.477	0.466	0.469	0.459	0.460
	0.5	0.336	0.443	0.436	0.158	0.447	0.434	0.456	0.428	0.468	0.456
	0.8	0.279	0.375	0.366	0.131	0.386	0.382	0.381	0.381	0.387	0.383

Note: Power is evaluated at $\bar{\alpha} = 1 + \bar{c}/T, \bar{c} = -7.0$. These are 0.93, 0.953, and 0.972 for $T = 100$, $T = 150$, and $T = 250$, respectively.

for $T = 100, 150,$ and 250, respectively.[9] The size-adjusted power of the tests are reported in Table V.B. Although the tests should reject the null hypothesis 50% of the time in large samples, none of the tests quite achieve the 50% (asymptotic) target for sample sizes less than 200. Thus, in finite samples, the need to account for serial correlation reduces the power of the tests across the board. However, the *GLS* based tests have more power than tests that do not implement *GLS*

[9] We also performed extensive power simulations with other values of α and T. The results are qualitatively similar and do not affect the conclusions concerning the relative performance of the different tests.

TABLE VI.A

SIZE OF THE TESTS; $p = 1$

					MA Case: $v_t = e_t + \theta e_{t-1}, e_t \sim N(0,1)$						
T	θ	M_{mic}	M^{GLS}_{mic}	$\bar{M}^{GLS}_{mic}$	DF_{bic}	DF^{GLS}_{bic}	DF^{GLS}_t	DF^{GLS}_{mic}	P_{Tbic}	$\bar{P}^{GLS}_{Tmic}$	$\bar{M}P^{GLS}_{Tmic}$
100	−0.8	0.140	0.123	0.059	0.509	0.567	0.528	0.123	0.562	0.061	0.061
	−0.5	0.037	0.025	0.035	0.092	0.107	0.111	0.069	0.145	0.036	0.036
	0.0	0.011	0.005	0.018	0.075	0.074	0.078	0.039	0.161	0.017	0.018
	0.5	0.032	0.022	0.048	0.084	0.085	0.094	0.038	0.190	0.046	0.047
	0.8	0.048	0.033	0.056	0.096	0.105	0.109	0.017	0.250	0.053	0.056
150	−0.8	0.074	0.061	0.027	0.581	0.586	0.477	0.082	0.570	0.028	0.028
	−0.5	0.032	0.017	0.028	0.087	0.103	0.104	0.049	0.114	0.028	0.029
	0.0	0.017	0.008	0.023	0.066	0.067	0.072	0.038	0.111	0.023	0.024
	0.5	0.025	0.016	0.032	0.068	0.084	0.085	0.029	0.142	0.032	0.034
	0.8	0.058	0.035	0.058	0.080	0.088	0.085	0.018	0.181	0.055	0.057
250	−0.8	0.025	0.016	0.012	0.566	0.531	0.381	0.064	0.554	0.012	0.012
	−0.5	0.031	0.016	0.033	0.094	0.105	0.099	0.048	0.103	0.034	0.035
	0.0	0.023	0.011	0.029	0.056	0.059	0.065	0.038	0.083	0.030	0.031
	0.5	0.032	0.012	0.037	0.062	0.066	0.070	0.028	0.097	0.036	0.037
	0.8	0.059	0.032	0.059	0.062	0.067	0.060	0.026	0.119	0.059	0.060

					AR Case: $v_t = \rho v_{t-1} + e_t, e_t \sim N(0,1)$						
T	ρ	M_{mic}	M^{GLS}_{mic}	$\bar{M}^{GLS}_{mic}$	DF_{bic}	DF^{GLS}_{bic}	DF^{GLS}_t	DF^{GLS}_{mic}	P_{Tbic}	$\bar{P}^{GLS}_{Tmic}$	$\bar{M}P^{GLS}_{Tmic}$
100	−0.8	0.000	0.000	0.000	0.072	0.079	0.084	0.034	0.067	0.001	0.000
	−0.5	0.006	0.003	0.012	0.071	0.074	0.077	0.037	0.122	0.013	0.013
	0.5	0.033	0.021	0.042	0.076	0.083	0.088	0.038	0.178	0.040	0.043
	0.8	0.081	0.056	0.093	0.081	0.086	0.091	0.043	0.210	0.080	0.085
150	−0.8	0.001	0.000	0.003	0.061	0.062	0.069	0.028	0.043	0.001	0.004
	−0.5	0.017	0.006	0.024	0.059	0.070	0.075	0.038	0.090	0.014	0.023
	0.5	0.037	0.018	0.044	0.056	0.070	0.074	0.039	0.127	0.041	0.043
	0.8	0.045	0.022	0.056	0.063	0.066	0.073	0.043	0.142	0.056	0.055
250	−0.8	0.001	0.000	0.003	0.057	0.054	0.058	0.028	0.009	0.004	0.004
	−0.5	0.017	0.006	0.024	0.046	0.047	0.053	0.038	0.025	0.023	0.023
	0.5	0.037	0.018	0.044	0.059	0.063	0.068	0.039	0.046	0.042	0.043
	0.8	0.045	0.022	0.056	0.054	0.056	0.061	0.043	0.054	0.052	0.055

Note: The results reported for the M_{mic} tests are those corresponding to MZ_α. Since the numerical results for MZ_α, MZ_t, and MSB are similar, we use the label M in the tables. Three variants of the M tests are considered. The M_{mic} is the original MZ_α, the M^{GLS}_{mic} uses GLS detrended data to construct the statistic but uses least squares detrending in the estimation of s^2_{AR}, and $\bar{M}^{GLS}_{mic}$ uses GLS detrending at $\bar{c} = -13.5$ when constructing s^2_{AR}. When the BIC is used, $k_{max} = 6$ and $k_{min} = 3$ as in ERS. For the MIC, $k_{max} = \text{int}(12(T/100)^{1/4})$, $k_{min} = 0$, and $C_T = 2$, giving the $MAIC$. The DF^{GLS}_t is based on the 10% t test for the significance of the last lag.

detrending. Furthermore, the power increase from $T = 100$ to $T = 150$ is quite substantial, especially when we take into account the fact that we are evaluating at a closer alternative at $T = 150$ than at $T = 100$.

For $p = 1$, the results for size are reported in Table VI.A. The size distortions are somewhat larger when there is a linear trend in the regression, but there are clearly gains from using GLS detrended data to construct s^2_{AR} and selecting k

TABLE VI.B

SIZE-ADJUSTED POWER OF THE TESTS; $p = 1$

MA Case: $v_t = e_t + \theta e_{t-1}$, $e_t \sim N(0,1)$

T	θ	M_{mic}	M^{GLS}_{mic}	$\bar{M}^{GLS}_{mic}$	DF_{bic}	DF^{GLS}_{bic}	DF^{GLS}_{t}	DF^{GLS}_{mic}	P_{Tbic}	$\bar{P}^{GLS}_{Tmic}$	$\bar{M}P^{GLS}_{Tmic}$
100	−0.8	0.296	0.266	0.348	0.283	0.398	0.414	0.357	0.355	0.350	0.348
	−0.5	0.320	0.246	0.300	0.258	0.421	0.420	0.324	0.324	0.311	0.305
	0.0	0.450	0.266	0.439	0.175	0.362	0.341	0.432	0.488	0.448	0.442
	0.5	0.196	0.147	0.207	0.161	0.316	0.310	0.259	0.320	0.224	0.218
	0.8	0.149	0.166	0.154	0.169	0.282	0.269	0.254	0.260	0.163	0.155
150	−0.8	0.236	0.269	0.316	0.309	0.444	0.416	0.354	0.326	0.316	0.314
	−0.5	0.283	0.348	0.325	0.291	0.434	0.415	0.358	0.384	0.333	0.330
	0.0	0.407	0.470	0.457	0.200	0.384	0.352	0.462	0.369	0.466	0.465
	0.5	0.312	0.347	0.348	0.212	0.365	0.365	0.375	0.354	0.344	0.348
	0.8	0.171	0.209	0.217	0.201	0.317	0.310	0.309	0.302	0.227	0.225
250	−0.8	0.247	0.285	0.296	0.371	0.460	0.430	0.375	0.406	0.300	0.297
	−0.5	0.308	0.370	0.366	0.278	0.459	0.441	0.392	0.422	0.374	0.371
	0.0	0.380	0.458	0.451	0.250	0.437	0.418	0.447	0.432	0.455	0.456
	0.5	0.311	0.374	0.365	0.258	0.426	0.406	0.388	0.441	0.367	0.363
	0.8	0.225	0.268	0.271	0.233	0.364	0.354	0.310	0.388	0.280	0.282

AR Case: $v_t = \rho v_{t-1} + e_t$, $e_t \sim N(0,1)$

T	ρ	M_{mic}	M^{GLS}_{mic}	$\bar{M}^{GLS}_{mic}$	DF_{bic}	DF^{GLS}_{bic}	DF^{GLS}_{t}	DF^{GLS}_{mic}	P_{Tbic}	$\bar{P}^{GLS}_{Tmic}$	$\bar{M}P^{GLS}_{Tmic}$
100	−0.8	0.257	0.308	0.296	0.204	0.380	0.372	0.385	0.256	0.301	0.299
	−0.5	0.332	0.396	0.382	0.205	0.367	0.366	0.431	0.303	0.398	0.393
	0.5	0.235	0.265	0.270	0.133	0.262	0.264	0.286	0.247	0.279	0.280
	0.8	0.114	0.127	0.129	0.098	0.167	0.165	0.206	0.144	0.150	0.144
150	−0.8	0.311	0.363	0.349	0.240	0.435	0.409	0.409	0.345	0.351	0.347
	−0.5	0.388	0.433	0.423	0.240	0.404	0.376	0.437	0.385	0.438	0.429
	0.5	0.296	0.338	0.345	0.207	0.338	0.313	0.366	0.333	0.358	0.355
	0.8	0.189	0.211	0.218	0.133	0.239	0.228	0.251	0.211	0.228	0.224
250	−0.8	0.379	0.452	0.441	0.245	0.472	0.456	0.468	0.464	0.446	0.442
	−0.5	0.372	0.453	0.451	0.300	0.509	0.484	0.445	0.445	0.452	0.457
	0.5	0.331	0.387	0.385	0.233	0.391	0.372	0.392	0.401	0.395	0.388
	0.8	0.259	0.280	0.288	0.189	0.314	0.292	0.289	0.312	0.285	0.284

Note: Power is evaluated at $\bar{\alpha} = 1 + \bar{c}/T$, $\bar{c} = -13.5$. There are 0.865, 0.910, and 0.946 for $T = 100$, $T = 150$, and $T = 250$, respectively.

with the *MAIC*. The resulting $\bar{M}^{GLS}$ tests have size close to the nominal size for all parameter values, while the M and M^{GLS} tests continue to have size close to .15. As well, the *MAIC* yields exact sizes for the DF^{GLS} and the P_T tests that are far more accurate than those obtained using the sequential t test and the *BIC*.

The size-adjusted power of the tests for $p = 1$ are reported in Table VI.B. Power is again evaluated at $\bar{\alpha} = 1 + \bar{c}/T$. For $\bar{c} = -13.5$, $\bar{\alpha}$ is 0.865 for $T = 100$, 0.910 for $T = 150$, and 0.946 for $T = 250$. Compared to the size-adjusted

power reported in Table V.B, all *GLS* based tests have lower power when $p = 1$. However, there are non-negligible power gains from *GLS* detrending.

A feature of interest is that the *MIC* allows the $\overline{M}^{GLS}$ tests to have an exact size that is closer to 5% than fixing k at k_{mz}^* assuming a known DGP. For example, in the *MA* case with $\theta = -.8$ and $p = 1$, the size is .059 using the *MIC* and .159 using a value of k fixed at $k_{mz}^* = 6$. When $T = 150$, the corresponding values are .027 with the *MAIC* and .093 at $k_{mz}^* = 8$ (not reported). Hence, the *MAIC* chooses a k that adapts to the particular sample analyzed. Size improvements obtained using the *MIC* compared to fixing k at k^* are even larger when the number of deterministic components increases (see Perron and Rodríguez (1998) for the case of a trend with a change in slope).

The size reported for $\overline{M}^{GLS}$ is remarkable, given that when $\theta < 0$, the *Z* tests based on kernel estimates of σ^2 reject with probability close to one under the null hypothesis. The *DF* tests have higher size-unadjusted power in the case of negative *MA* errors. Such high probabilities of rejections are misleading, however, as they reflect size distortions. Although the DF_{bic}^{GLS} has higher size-adjusted power than $\overline{M}_{mic}^{GLS}$ for some parameter values, this result must also be put into perspective. Consider $\theta = -0.8$, $T = 100$, and $p = 1$. While the size-adjusted power of the DF_{bic}^{GLS} is higher than the $\overline{M}_{mic}^{GLS}$ by 5 basis points, its size is more distorted by 45 basis points. This is a substantial increase in the probability of a Type I error. The DF_{mic}^{GLS} fares better, but there is still a size advantage in favor of the $\overline{M}_{mic}^{GLS}$.

Overall, the $\overline{M}_{mic}^{GLS}$ tests have better sizes, while the DF_{mic}^{GLS} has better power. A drawback of the $\overline{M}_{mic}^{GLS}$ test is that it is substantially undersized for pure *AR* processes with a coefficient close to -1. Although economic time series rarely display such a property, this nonetheless translates into lower (size unadjusted) power. As discussed in Section 6.2, this is not due to a weakness of the *MIC* in selecting k, but rather is an inherent property of the $\overline{M}^{GLS}$ tests. The trade-offs between the DF_{mic}^{GLS} and the $\overline{M}_{mic}^{GLS}$ are to be determined by the practitioner.

8. EMPIRICAL ILLUSTRATIONS

To illustrate the ease of application of our procedures and the differences in inference that can result, we consider inflation series from the G7 countries. Data for the *GDP* deflator are taken from the OECD International Sector Database. We constructed annualized quarterly inflation rate series covering the period 1960:2 to 1997:2 ($T + 1 = 149$ observations) as (400 times) the logged differences of successive quarters. All estimation results are presented in Table VII. As a preliminary step, we first estimated the simple $ARMA(1, 1)$ model for each series. The results show autoregressive roots that are indeed close to one (ranging from .92 for Germany to .98 for France). Also, the *MA* coefficients are negative for all series, ranging from $-.27$ for Italy to $-.72$ for France. This suggests that such series are likely affected by the types of problems we discussed.

The first step to constructing the tests is to generate *GLS* detrended series as defined by (6) with $\bar{c} = -7$ as the noncentrality parameter (since inflation series are nontrending; otherwise use $\bar{c} = -13.5$). The next step is to estimate by

SERENA NG AND PIERRE PERRON

TABLE VII

EMPIRICAL RESULTS FOR INFLATION SERIES FROM THE G7 COUNTRIES

		USA	Canada	UK	Japan	Italy	France	Germany
	$AR(\alpha)$:	.95	.95	.93	.87	.92	.98	.92
	$MA(\theta)$:	−.32	−.51	−.55	−.42	−.27	−.72	−.54
BIC	k	2	1	1	1	0	3	3
	$\overline{Z}_\alpha^{GLS}$	−4.04	−9.63^b	−26.81^a	−34.12^a	−7.39^c	−5.70^c	−26.58^a
	$\overline{M}Z_\alpha^{GLS}$	−3.71	−8.13^b	−18.05^a	−23.95^a	−7.17^c	−3.49	−16.57^a
	DF^{GLS}	−1.37	−2.07^b	−3.31^a	−3.85^a	−1.95^c	−1.36	−3.00^a
	$\overline{P}_T^{GLS}$	7.39	3.36^c	1.40^a	1.01^a	4.06^c	7.92	1.96^b
	$\overline{M}P_T^{GLS}$	6.60	3.12^b	1.41^a	1.02^a	3.42^c	7.01	1.98^b
MAIC	k	2	2	8	12	5	7	5
	$\overline{Z}_\alpha^{GLS}$	−4.04	−6.98^c	−14.75^a	−13.35^a	−2.71	−3.27	−18.83^a
	$\overline{M}Z_\alpha^{GLS}$	−3.71	−5.48	−6.00^c	−3.17	−2.49	−1.07	−8.82^b
	DF^{GLS}	−1.37	−1.69^c	1.77^c	−1.33	−1.08	−0.77	−2.21^b
	$\overline{P}_T^{GLS}$	7.39	4.92	4.13^c	7.64	11.68	25.85	3.42^c
	$\overline{M}P_T^{GLS}$	6.60	4.58	4.17^c	7.73	9.83	22.89	3.44^c

Note: a, b, and c denotes a statistic significant at the 1%, 5%, and 10% level, respectively.

OLS the autoregression (8) for all values of k ranging from 0 to some maximal order, say k_{max}. We use $k_{max} = \text{int}(12(T/100)^{1/4})$ but other values are valid. In the model selection stage, each autoregression uses the same number of effective observations, namely $T - k_{max}$. The selected k_{maic} is obtained as the value of k that minimizes (12). Given k_{maic}, the DF_{maic}^{GLS} is obtained as the t-statistic on $\beta_0 = 0$ from re-estimation of (8) with $T - k_{maic}$ effective observations. The same regression is used to construct s_{AR}^2 according to (5). Then, the $\overline{M}^{GLS}$ tests are constructed from (3) and (4), the $\overline{P}_T^{GLS}$ test from (7), and the $\overline{M}P_T^{GLS}$ test from (9), all with *GLS* detrended series $\tilde{y}_t$. Using the *BIC*, the steps are the same except that the term $\tau_T(k)$ is omitted from (12).

The results show interesting differences between using the *MAIC* or the *BIC* to select the lag length. First, for all countries, except the U.S., the selected lag length is higher with *MAIC* than with *BIC*. Second, in all cases (except the U.S.), the evidence in favor of stationarity is weaker with the *MAIC* than with the *BIC*. For example, with the *BIC*, the $\overline{M}Z_{\alpha,bic}^{GLS}$ and DF_{bic}^{GLS} tests suggest a rejection at the 1% significance level for Japan and Germany, at the 5% level for Canada, and at the 10% level for Italy. When using the *MAIC*, there are only rejections at the 5% level for Germany (showing that the tests still have power) and at the 10% level for the U.K. For Canada, the DF_{maic}^{GLS} is significant at the 10% level but barely. For the U.S. and France, there is no difference in outcome. Of interest also is the fact that the $\overline{Z}_{\alpha,maic}^{GLS}$ still shows strong rejections at the 1% level for Japan, the U.K., and Germany. This accords with the simulation results that the $\overline{Z}_\alpha^{GLS}$ test has high size distortions even when constructed using the *MAIC*.

9. CONCLUSION

The analysis of ERS shows that detrending data by *GLS* yields power gains for unit root tests. We find that *GLS* detrending also allows for a more precise autoregressive spectral density estimate and ensures that it is invariant to the parameters of the trend function. We also show that the *BIC* or *AIC* are not sufficiently flexible for unit root tests. We suggest a modification, the *MAIC*, which is evidently more robust when there are negative moving-average errors. Use of the *MAIC* in conjunction with *GLS* detrended data results in a class of $\overline{M}^{GLS}$ tests that have good size and power. The key distinction between the *MAIC* and standard information criteria is that the former takes account of the fact that the bias in the estimate of the sum of the autoregressive coefficients is highly dependent on k. In this paper, we show that the *MAIC* is useful for the DF^{GLS} test and the construction of the autoregressive estimate of the spectral density at frequency zero. We also expect the *MAIC*, or the more general *MIC*, to be useful in a broader range of applications because macroeconomic data are known to be highly persistent, and in such cases, the bias in the estimate of the sum of the autoregressive coefficients should depend on the order of the autoregression. However, it should also be noted that while the suggested information criterion is useful in the context of unit root tests as we have shown, it may not be appropriate when selecting the lag order in other contexts. The general usefulness of the *MIC* also awaits further investigation.

Dept. of Economics, Johns Hopkins University, Baltimore, MD 21218, U.S.A.; Serena.Ng@jhu.edu

and

Dept. of Economics, Boston University, 270 Bay State Rd., Boston, MA 02215, U.S.A.; perron@bu.edu

APPENDIX

PROOF OF THEOREM 2: In this Appendix, for simplicity, we consider the derivations in the case where the DGP is (14) and no deterministic component is present in the regression (8), in which case y_t is used instead of $\tilde{y}_t$. All the results carry over to the more general case with least-squares or *GLS* detrending. With *GLS* detrending, the stated results remain exactly the same when $p = 0$ and are as stated in the text with $V_{c,\bar{c}}(r)$ replacing $J_c(r)$ when $p = 1$. As a matter of notation, we let $\Rightarrow$ denote weak convergence in distribution and $\rightarrow$ convergence in probability. Also, to alleviate notational burden, we let $N = T - k_{\max}$. Note that we can write the DGP as

$$\Delta y_t = -\sum_{i=1}^{k}(-\theta_T)^i \Delta y_{t-i} + \frac{c}{T}\sum_{i=0}^{k}(-\theta_T)^i y_{t-i-1} + e_t - (-\theta_T)^{k+1}e_{t-k-1}.$$

Let

$$z_{t-j} = \sum_{i=0}^{k-j}(-\theta_T)^i \Delta y_{t-i-j} = \frac{c}{T}\sum_{i=0}^{k-j}(-\theta_T)^i y_{t-i-j-1} + e_{t-j} - (-\theta_T)^{k+1-j}e_{t-k-1};$$

the DGP can also be written as

$$(A.1) \qquad \Delta y_t = \frac{c}{T} y_{t-1} - (-\theta_T) z_{t-1} + \frac{c}{T} \sum_{i=1}^{k} (-\theta_T)^i y_{t-i-1} + e_t - (-\theta_T)^{k+1} e_{t-k-1}$$

$$= \beta_0 y_{t-1} + \sum_{i=1}^{k} \beta_i z_{t-i} + e_{tk}$$

with $\beta_0 = c/T$, $\beta_1 = \theta_T$, $\beta_i = 0$ for $i > 1$, and

$$(A.2) \qquad e_{tk} = \frac{c}{T} \sum_{i=1}^{k} (-\theta_T)^i y_{t-i-1} + e_t - (-\theta_T)^{k+1} e_{t-k-1}.$$

Since $\{z_{t-i}, i = 1, \ldots, k\}$ is a linear function of $\{\Delta y_{t-i}, i = 1, \ldots, k\}$, the OLS estimate of β_0 and the sum of squared residuals from the regression

$$(A.3) \qquad \Delta y_t = \beta_0 y_{t-1} + \sum_{i=1}^{k} \beta_i \Delta y_{t-i} + e_{tk}$$

have the same properties as the corresponding quantities from estimating (A.1). We start with a Lemma that will be used extensively.

LEMMA A.1: *Let* $J_c(r) = \int_0^r \exp((r-s)c) dW(s)$, $K_{3T} = o(T)$, *and* $\delta_{ij} = 1$ *if* $i = j$ *and* 0 *otherwise with* $i, j = 1, \ldots, k$. *Then, uniformly for* $i, j \leq k_{\max} \leq K_{3T} = o(T)$: (a) $N^{-1} \sum_{t=k_{\max}+1}^{T} y_{t-1}^2 = \sigma_e^2 (1 + \delta^2 \int_0^1 J_c(r)^2 dr) + O_p(T^{-1/2})$; (b) $N^{-1} \sum_{t=k_{\max}+1}^{T} y_{t-1} z_{t-j} = \delta_{1j} \sigma_e^2 + O_p(T^{-1/2})$; (c) $N^{-1} \sum_{t=k_{\max}+1}^{T} z_{t-i} z_{t-j} = \sigma_e^2 (\delta_{ij} + (-\theta_T)^{2(k+1)-i-j}) + O_p(T^{-1/2})$; (d) $N^{-1} \sum_{t=k_{\max}+1}^{T} \Delta y_t y_{t-1} = -\sigma_e^2 + O_p(T^{-1/2})$; (e) $N^{-1} \sum_{t=k_{\max}+1}^{T} \Delta y_t z_{t-j} = -\delta_{1j} \sigma_e^2 + O_p(T^{-1/2})$.

PROOF: From Nabeya and Perron (1994), we can write $y_t = a_T e_t + b_T X_t$ where $X_t = (1 + c/T) X_{t-1} + e_t$, $a_T = (-\theta_T)(1 + c/T)$, $b_T = 1 - (1 + c/T)(-\theta_T)$, $a_T \to 1$, $T^{1/2} b_T \to \delta$ as $T \to \infty$. Also, $\sum_{t=1}^{T} X_{t-1} e_t = O_p(T)$. For part (a), we have:

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

$$= \sigma_e^2 \left(1 + \delta^2 \int_0^1 J_c(r)^2 dr \right) + O_p(T^{-1/2}).$$

For part (b), we have

$$N^{-1} \sum_{t=k_{\max}+1}^{T} y_{t-1} z_{t-j}$$

$$= N^{-1} \sum_{t=k_{\max}+1}^{T} (a_T e_{t-1} + b_T X_{t-1}) \left(\frac{c}{T} \sum_{i=0}^{k-j} (-\theta_T)^i y_{t-i-j-1} - (-\theta_T)^{k+1-j} e_{t-k-1} + e_{t-j} \right).$$

Since $X_t = \sum_{j=1}^{t} \exp((t-j)c/T) e_j$, $X_{t-1} = X_{t-k-1} + \sum_{j=1}^{k} \exp((t-j)c/T) e_j$. We have $\sum_{t=k_{\max}+1}^{T} e_t e_s = O_p(T^{1/2})$ for $t \neq s$, $T^{1/2} b_T = O_p(1)$, and $\sum_{t=k_{\max}+1}^{T} X_{t-1} e_t = O_p(T)$, hence, $N^{-1} \sum_{t=k_{\max}+1}^{T} y_{t-1} z_{t-1} - \sigma_e^2 = O_p(T^{-1/2})$ and $N^{-1} \sum_{t=k_{\max}+1}^{T} y_{t-1} z_{t-j} = O_p(T^{-1/2})$ for $j > 1$ uniformly in $0 < i, j \leq k_{\max} \leq K_{3T}$. For part (c),

$$N^{-1} \sum_{t=k_{\max}+1}^{T} z_{t-i} z_{t-j}$$

$$= N^{-1} \sum_{t=k_{\max}+1}^{T} [e_{t-i} - e_{t-k-1}(-\theta_T)^{k+1-i}][e_{t-j} - e_{t-k-1}(-\theta_T)^{k+1-j}] + O_p(T^{-1/2})$$

$$= \sigma_e^2 (\delta_{ij} + (-\theta_T)^{2(k+1)-i-j}) + O_p(T^{-1/2}).$$

Part (d) follows using the fact that

$$N^{-1} \sum_{t=k_{\max}+1}^{T} \Delta y_t y_{t-1} = cN^{-2} \sum_{t=k_{\max}+1}^{T} y_{t-1}^2 + N^{-1} \sum_{t=k_{\max}+1}^{T} (a_T e_{t-1} + b_T X_{t-1})(e_t + \theta_T e_{t-1}),$$

and part (e) using the fact that

$$N^{-1} \sum_{t=k_{\max}+1}^{T} \Delta y_t z_{t-i} = N^{-1} \sum_{t=k_{\max}+1}^{T} \left(\frac{c}{T} y_{t-1} + e_t - (-\theta_T) e_{t-1} \right)$$

$$\times \left(\frac{c}{T} \sum_{j=0}^{k-i} (-\theta_T)^j y_{t-i-j-1} + e_{t-i} - e_{t-k-1} (-\theta_T)^{k+1-i} \right).$$

The proof of Theorem 2 follows directly from the following Lemma concerned with the limit of $\hat{\beta}_0$, the OLS estimate of β_0 from the regression (A.1), which is equivalent to the OLS estimate of β_0 from the autoregression (8).

LEMMA A.2: *Consider the OLS estimate of β_0 from the autoregression* (8). *Let $x = \delta^2 \int_0^1 J_c(r)^2 dr$ when $p = 0$ and $x = \delta^2 \int_0^1 V_{c,\bar{c}}(r)^2 dr$ when $p = 1$. Then,* (i) *uniformly in $0 < k \le K_{1T}$: $k\hat{\beta}_0 \Rightarrow -x^{-1}$;* (ii) *if $k/T^{1/2} \to \kappa$, $T^{1/2}\hat{\beta}_0 \Rightarrow A$, where A is a random variable with $E(A) = -\vartheta(\kappa)E(x^{-1})$, $\vartheta(\kappa) = 2\delta \exp(-2\kappa\delta)/(1 - \exp(-2\kappa\delta))$);* (iii) *uniformly in $K_{2T} \le k \le K_{3T}$, $T^{1/2}\hat{\beta}_0 = O_p(1)$.*

(i) *Case* $0 < k \le K_{1T}$. Let $X_t' = (y_{t-1}, Z_t')$, $Z_t = (z_{t-1}, \ldots, z_{t-k})'$. Define the following two $(k+1)$ by $(k+1)$ matrices:

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

$$\tilde{R} = \sigma_e^2 \begin{bmatrix} 1+x & e' \\ e & I_k + \iota\iota' \end{bmatrix},$$

where $x = \delta^2 \int_0^1 J_c(r)^2 dr$, $e' = (1, 0_{1 \times k})$ and ι is a $k \times 1$ vector of ones. The inverse of $\tilde{R}$ is

$$\tilde{R}^{-1} = \sigma_e^{-2} \frac{1}{1+(k+1)x} \begin{bmatrix} k+1 & -k & 1 & 1 & & 1 \\ -k & k(1+x) & -(1+x) & -(1+x) & & -(1+x) \\ 1 & -(1+x) & k(1+x) & -x & & -x \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & -(1+x) & & -x & k(1+x) & -x \\ 1 & -(1+x) & -x & -x & -x & k(1+x) \end{bmatrix}.$$

In what follows, the structure of $\tilde{R}^{-1}$ will be exploited. Let

$$\hat{V}_T = \left(N^{-1} \sum_{t=k_{\max}+1}^{T} \Delta y_t y_{t-1}, N^{-1} \sum_{t=k_{\max}+1}^{T} \Delta y_t Z_t' \right)', \qquad \tilde{V} = (-1, -1, 0_{k-1}')'.$$

Note that $\hat{\theta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)' = \hat{R}_T^{-1} \hat{V}_T$.

LEMMA A.3: (a) $\max_{0 < k \le K_{1T}} \|\hat{R}_T - \tilde{R}\| = O_p(K_{1T}/\sqrt{T})$,   (b) $\max_{0 < k \le K_{1T}} \|\hat{R}_T^{-1} - \tilde{R}^{-1}\| = O_p(K_{1T}/\sqrt{T})$.

PROOF: Using Lemma A.1, we have $\widehat{R}_{ij} - \widetilde{R}_{ij} = C_{ij}$ where $C_{ij} = O_p(T^{-1/2})$ uniformly in $0 < k \leq K_{1T}$ for $i, j = 1, \ldots, k+1$ and part (a) follows. To prove part (b), note that the lower $k \times k$ submatrix of $\widetilde{R}$ is, say, $\widetilde{R}_k = (I_k + \iota\iota')$. Then $(I_k + \iota\iota')^{-1} = I_k - (1 + \iota'\iota)^{-1}\iota\iota'$ has eigenvalues of 1 and $(1 + \iota'\iota)^{-1}$. Hence $\max_{0 < k \leq K_{1T}} \|\widetilde{R}_k^{-1}\| = O_p(1)$. Using arguments as in Perron and Ng (1998), the $(k+1) \times (k+1)$ matrix $\widetilde{R}$ also has eigenvalues of the same order as $\widetilde{R}_k$. Hence $\max_{0 < k \leq K_{1T}} \|\widetilde{R}^{-1}\| = O_p(1)$. Let $Q_T = \|\widehat{R}_T^{-1} - \widetilde{R}^{-1}\|$. Then $\max_{0 < k \leq K_{1T}} Q_T \leq \max_{0 < k \leq K_{1T}} \|\widehat{R}_T^{-1}\|\|\widehat{R}_T - \widetilde{R}\|\|\widetilde{R}^{-1}\| = O_p(K_{1T}/\sqrt{T})$ by arguments analogous to Berk (1974).

LEMMA A.4: (a) $\max_{0 < k \leq K_{1T}} \|\widehat{V}_T - \widetilde{V}\| = O_p(\sqrt{K_{1T}/T})$; (b) $\max_{0 < k \leq K_{1T}} \|\widehat{V}_T\| = O_p(1)$.

PROOF: Part (a) follows directly from Lemma A.1 (d, e). Since $\widetilde{V} = (-1, -1, 0_{1 \times k-1})'$, $\max_{0 < k \leq K_{1T}} \|\widetilde{V}\| = O_p(1)$ and $\max_{0 < k \leq K_{1T}} \|\widehat{V}_T\| = O_p(1)$ in view of part (a) of the Lemma.

Let $e' = (1, 0_{1 \times k})$. We have

$$\hat{\beta}_0 = e'\widehat{R}_T^{-1}\widehat{V}_T = e'(\widehat{R}_T^{-1} - \widetilde{R}^{-1})(\widehat{V}_T - \widetilde{V}) + e'(\widehat{R}_T^{-1} - \widetilde{R}^{-1})\widetilde{V} + e'\widetilde{R}^{-1}(\widehat{V}_T - \widetilde{V}) + e'\widetilde{R}^{-1}\widetilde{V}$$

$$= A + B + C + D.$$

For a matrix M, denote the i, jth element of M^{-1} by m^{ij}. We now consider the limit of each of the four terms. For A,

$$\max_{0 < k \leq K_{1T}} \|ke'(\widehat{R}_T^{-1} - \widetilde{R})(\widehat{V}_T - \widetilde{V})\|^2 \leq K_{1T}^2 \sum_{i=1}^{k} \|\hat{r}^{1i} - \tilde{r}^{1i}\|^2 \|\widehat{V}_{i,T} - \widetilde{V}_i\|^2$$

$$\leq c\frac{K_{1T}^2}{T} \sum_{i=1}^{k} \|\hat{r}^{1i} - \tilde{r}^{1i}\|^2 \leq c\left(\frac{K_{1T}^2}{T}\right)^2 \to 0$$

since $\sum_{i=1}^{k} \|\hat{r}^{1i} - \tilde{r}^{1i}\|^2 \leq \|\widehat{R}_T^{-1} - \widetilde{R}^{-1}\|^2 = O_p(k^2/T) \leq O_p(K_{1T}^2/T)$. For the term B, we need $\max_{0 < k \leq K_{1T}} \|ke'(\widehat{R}_T^{-1} - \widetilde{R}^{-1})\widetilde{V}\|$. Recall that $\widehat{R}_T = \widetilde{R} + C$, where $C_{ij} = O_p(T^{-1/2})$, $i, j = 1, \ldots, k+1$. Then $\widehat{R}_T^{-1} = \widetilde{R}^{-1} + \widetilde{R}^{-1}C\widetilde{R}^{-1} + O_p(T^{-1})$. We have (ignoring, without loss, the $O_p(T^{-1})$ term):

$$\|e'(\widehat{R}_T^{-1} - \widetilde{R}^{-1})\widetilde{V}\| = \|e'\widetilde{R}^{-1}C\widetilde{R}^{-1}\widetilde{V}\| = \left\|\sum_{j=1}^{k+1}(\tilde{r}^{1j} + \tilde{r}^{2j})\sum_{i=1}^{k+1}\tilde{r}^{1i}C_{ij}\right\|$$

$$\leq \sum_{j=1}^{k+1}\|(\tilde{r}^{1j} + \tilde{r}^{2j})\|\sum_{i=1}^{k+1}\|\tilde{r}^{1i}\|\|C_{ij}\| \leq c^*\sum_{j=1}^{k}\|(\tilde{r}^{1j} + \tilde{r}^{2j})\|\sum_{i=1}^{k+1}\|\tilde{r}^{1i}\|,$$

where $c^* = \max_{i,j} C_{ij}$. But the sum of the norms of the elements of the first row of $\widetilde{R}^{-1}$ is $[2k + (k-1)][1 + (k+1)x]^{-1} = O_p(1)$, and the sum of the norms of the sum of the first two elements of each row of $\widetilde{R}^{-1}$ is given by $(1 + kx + (k-1)x)/(1 + (k+1)x) = O_p(1)$. Since $c^* = O_p(T^{-1/2})$, $\max_{0 < k \leq K_{1T}} \|ke'(\widehat{R}_T^{-1} - \widetilde{R}^{-1})\widetilde{V}\| = O_p(K_{1T}/T^{1/2})$. For the term C, first note that

$$(\widehat{V}_T - \widetilde{V})' = N^{-1}\left[\left(\sum_{t=k_{max}+1}^{T}\Delta y_t y_{t-1} + 1\right), \left(\sum_{t=k_{max}+1}^{T}\Delta y_t z_{t-1} + 1\right),\right.$$

$$\left.\sum_{t=k_{max}+1}^{T}\Delta y_t z_{t-2}, \ldots, \sum_{t=k_{max}+1}^{T}\Delta y_t z_{t-k}\right].$$

Thus,

$$\max_{0 < k \le K_{1T}} \|ke'\widetilde{R}^{-1}(\widehat{V}_T - \widetilde{V})\|$$

$$\le \frac{K_{1T}}{1 + (K_{1T} + 1)x} \left[\left\| (K_{1T} + 1)\left(N^{-1}\sum_{t=k_{\max}+1}^{T} \Delta y_t y_{t-1} + 1 \right) \right\| \right.$$

$$\left. - \left\| K_{1T}\left(N^{-1}\sum_{t=k_{\max}+1}^{T} \Delta y_t z_{t-1} + 1 \right) \right\| + N^{-1}\sum_{j=2}^{k} \left\| \sum_{t=k_{\max}+1}^{T} \Delta y_t z_{t-j} \right\| \right]$$

$$\le (cK_{1T}/N^{1/2}) \to 0.$$

For D,

$$\max_{0 < k \le K_{1T}} (ke'\widetilde{R}^{-1}\widetilde{V}) = \frac{-K_{1T}}{1 + (K_{1T} + 1)x} \to \frac{-1}{x} \equiv \frac{-1}{\delta^2 \int_0^1 J_c(r)^2 dr}.$$

Combining the results, we have $k\hat{\beta}_0 = o_p(1) + o_p(1) + o_p(1) + O_p(1) = O_p(1)$. Thus, $k\hat{\beta}_0 \Rightarrow (\delta^2 \int_0^1 J_c(r)^2 dr)^{-1}$ uniformly in $0 < k \le K_{1T}$.

(ii) *Case $k/T^{1/2} \to \kappa$:*

For the case where $k/T^{1/2} \to \kappa$, we first note that the relevant matrix $\widetilde{R}$ is given by

$$\sigma_e^2 \begin{bmatrix} 1+x & e' \\ e & I_k + \mu\mu' \end{bmatrix}$$

where $\mu' = [(-\theta_T)^k, \ldots, (-\theta_T)]$. The limiting vector $\widetilde{V}$ remains unchanged. Now we have $\sqrt{T}\hat{\beta}_0 = \sqrt{T}e'\widehat{R}^{-1}\widetilde{V} + \sqrt{T}W$, where $W = A + B + C$. It is easy to see that $E(\sqrt{T}W) = 0$ in large samples. So, we need to evaluate $\sqrt{T}e'\widetilde{R}^{-1}\widetilde{V}$, (minus) the sum of the first two elements of $\widetilde{R}^{-1}$. By partitioned matrix inversion, these are given by

$$\widetilde{R}_{11}^{-1}: [(1+x) - e'(I_k + \mu\mu')^{-1}e]^{-1},$$

$$\widetilde{R}_{12}^{-1}: -[(1+x) - e'(I_k + \mu\mu')^{-1}e]^{-1}e'(I_k + \mu\mu')^{-1}e.$$

Since $(I_k + \mu\mu')^{-1} = I_k - \mu\mu'/\lambda$, $\lambda = 1 + \mu'\mu$, $e'(I_k + \mu\mu')^{-1}e = e'e - (e'\mu\mu'e/\lambda) = 1 - [(-\theta_T)^{2k}/\lambda]$. Hence, $1 - e'(I_k + \mu\mu')^{-1}e = (-\theta_T)^{2k}/\lambda$ and substituting for $\lambda = (1 - (-\theta_T)^{2(k+1)})/(1 - (-\theta_T)^2)$,

$$(A.4) \qquad \sqrt{T}e'\widetilde{R}^{-1}\widetilde{V} = \frac{-(-\theta)^{2k}(1-\theta_T)^2}{[1 - (-\theta)^{2(k+1)}]x + (-\theta)^{2k}(1-\theta_T)^2} = \frac{-(-\theta)^{2k}(1-\theta_T)^2}{[1 - (-\theta)^{2(k+1)}]x} + O_p(T^{-1/2}),$$

which converges to $-\vartheta(\kappa)/x$ setting $k = \kappa T^{1/2}$ and taking the limit as $T \to \infty$.

(iii) *Case $K_{2T} \le k \le K_{3T}$.*

Note that $(-\theta_T)^k \to 0$ if $k/\sqrt{T} \to \infty$. Let

$$\widetilde{R} = \sigma_e^2 \begin{bmatrix} 1+x & e' \\ e & I_k \end{bmatrix}$$

and note that

$$\widetilde{R}^{-1} = \sigma_e^2 \begin{bmatrix} x^{-1} & -x^{-1} & 0_{1 \times k-1} \\ -x^{-1} & (1+x)x^{-1} & 0 \\ 0_{k-1 \times 1} & 0 & I_{k-1} \end{bmatrix}.$$

Thus, $\|\widetilde{R}\| = O_p(1)$ and $\|\widetilde{R}^{-1}\| = O_p(1)$. From Lemma A.1 (a, b, c), $\max_{K_{2T} \leq k \leq K_{3T}} \|\widehat{R}_T - \widetilde{R}\| = O_p(K_{3T}/T^{1/2})$. Also, $\max_{K_{2T} \leq k \leq K_{3T}} \|\widehat{R}_T\| \leq \max_{K_{2T} \leq k \leq K_{3T}} \|\widehat{R}_T - \widetilde{R}\| + \max_{K_{2T} \leq k \leq K_{3T}} \|\widetilde{R}\| = O_p(K_{3T}/T^{1/2})$ since $\max_{K_{2T} \leq k \leq K_{3T}} \|\widetilde{R}\| = O_p(1)$ and $K_{3T}/T^{1/2} \to \infty$. Since $\max_{K_{2T} \leq k \leq K_{3T}} \|\widehat{R}_T^{-1}\| \leq 1/\max_{K_{2T} \leq k \leq K_{3T}} \|\widehat{R}_T\|$, $\max_{K_{2T} \leq k \leq K_{3T}} \|\widehat{R}_T^{-1}\| \leq O_p(T^{1/2}/K_{3T})$. It follows that

$$\max_{K_{2T} \leq k \leq K_{3T}} \|\widehat{R}_T^{-1} - \widetilde{R}^{-1}\| \leq \max_{K_{2T} \leq k \leq K_{3T}} \|\widehat{R}_T^{-1}\| \|\widehat{R}_T - \widetilde{R}\| \|\widetilde{R}^{-1}\|$$

$$\leq O_p(T^{1/2}/K_{3T}) O_p(K_{3T}/T^{1/2}) O_p(1) = O_p(1).$$

We also have $\max_{K_{2T} \leq k \leq K_{3T}} \|\widehat{V}_T - \widetilde{V}\| = O_p(K_{3T}^{1/2}/T^{1/2})$ and $\|\widetilde{V}\| = O_p(1)$. Thus,

$$T^{1/2}\widehat{\beta}_0 = T^{1/2} e' \widehat{R}_T^{-1} \widehat{V}_T$$

$$= T^{1/2} e' (\widehat{R}_T^{-1} - \widetilde{R}^{-1})(\widehat{V}_T - \widetilde{V}) + T^{1/2} e' (\widehat{R}_T^{-1} - \widetilde{R}^{-1})\widetilde{V}$$

$$+ T^{1/2} e' \widetilde{R}^{-1}(\widehat{V}_T - \widetilde{V}) + T^{1/2} e' \widetilde{R}^{-1}\widetilde{V}$$

$$= A + B + C + D.$$

For A, we have

$$\max_{K_{2T} \leq k \leq K_{3T}} \|T^{1/2} e'(\widehat{R}_T^{-1} - \widetilde{R}^{-1})(\widehat{V}_T - \widetilde{V})\|^2 \leq \max_{K_{2T} \leq k \leq K_{3T}} T \sum_{i=1}^{k} \|\widehat{r}^{1i} - \widetilde{r}^{1i}\|^2 \|\widehat{V}_{i,T} - \widetilde{V}_i\|^2$$

$$\leq \max_{K_{2T} \leq k \leq K_{3T}} c \sum_{i=1}^{k} \|\widehat{r}^{1i} - \widetilde{r}^{1i}\|^2$$

$$\leq \max_{K_{2T} \leq k \leq K_{3T}} c \|\widehat{R}_T^{-1} - \widetilde{R}^{-1}\|^2 = O_p(1).$$

For B, $\|T^{1/2} e'(\widehat{R}_T^{-1} - \widetilde{R}^{-1})\widetilde{V}\| = \|T^{1/2} e' \widetilde{R}^{-1} C \widetilde{R}^{-1} \widetilde{V}\| + O_p(T^{-1/2})$ for some $\widetilde{R}$ such that $\widehat{R}_T = \widetilde{R} + C$, where the matrix C has elements that are $O_p(T^{-1/2})$. Now

$$e' \widetilde{R}^{-1} = [x^{-1}, -x^{-1}, 0, 0, 0, \ldots, 0],$$

$$\widetilde{R}^{-1} \widetilde{V} = [0, -1, 0, 0, \ldots, 0]'.$$

Therefore $C \widetilde{R}^{-1} \widetilde{V} = -C_{\cdot 2}$ where $C_{\cdot 2}$ is the second column of the matrix C. Then $e' \widetilde{R}^{-1} C \widetilde{R}^{-1} \widetilde{V} = x^{-1}(-C_{12} + C_{22})$. Since C_{12} and $C_{22} = O_p(T^{-1/2})$, $\max_{K_{2T} \leq k \leq K_{3T}} \|T^{1/2} e' \widetilde{R}^{-1} C \widetilde{R}^{-1} \widetilde{V}\| = O_p(1)$. For the third term C,

$$\max_{K_{2T} \leq k \leq K_{3T}} \|T^{1/2} e' \widetilde{R}^{-1}(\widehat{V}_T - \widetilde{V})\| = \max_{K_{2T} \leq k \leq K_{3T}} \|T^{1/2}[x^{-1}, -x^{-1}, \ldots, 0](\widehat{V}_T - \widetilde{V})\|$$

$$= \max_{K_{2T} \leq k \leq K_{3T}} \|\frac{T^{1/2}}{x}[(\widehat{V}_{1,T} - \widetilde{V}_1) - (\widehat{V}_{2,T} - \widetilde{V}_2)]\|$$

$$\leq \frac{T^{1/2}}{x} \frac{c}{T^{1/2}} = O_p(1).$$

For D,

$$T^{1/2} e' \widetilde{R}^{-1} \widetilde{V} = T^{1/2}\left[-x^{-1} + x^{-1}\right] = 0.$$

Combining the results, we have $T^{1/2}\widehat{\beta}_0 = O_p(1)$ uniformly in $K_{2T} \leq k \leq K_{3T}$.

PROOF OF LEMMA 1

LEMMA A.5: *Consider the estimated autoregression* (8) *when the DGP is given by* (14). *Let* $\hat{\sigma}_k^2 = N^{-1}\sum_{t=1}^{T}\hat{e}_{tk}^2$; *then uniformly over* $0 < k \leq K_{3T}$,

$$\ln(\hat{\sigma}_k^2) = \ln(\sigma_e^2) + \frac{(-\theta_T)^{2(k+1)}}{1-(-\theta_T)^{2(k+1)}}(1-\theta_T^2) + O_p(T^{-1/2}).$$

PROOF: Let $E_k = (e_{k_{\max}+1k}, \ldots, e_{Tk})'$, where $e_{tk} = cT^{-1}\sum_{i=1}^{k}(-\theta_T)^i y_{t-i-1} + e_t - (-\theta_T)^{k+1}e_{t-k-1}$. Let $X = (X_1, \ldots, X_T)'$. Then

$$\hat{\sigma}_k^2 = N^{-1}E_k'E_k - N^{-1}E_k'X(X'X)^{-1}X'E_k.$$

By direct calculations,

(A.5) $$N^{-1}E_k'E_k = N^{-1}\sum_{t=k_{\max}+1}^{T} e_t^2 + (-\theta_T)^{2(k+1)}N^{-1}\sum_{t=k_{\max}+1}^{T} e_{t-k-1}^2 + O_p(T^{-1/2})$$

(A.6) $$= \sigma_e^2(1 + (-\theta_T)^{2(k+1)}) + O_p(T^{-1/2}).$$

Consider $N^{-1}E_k'X = (N^{-1}\sum_{t=k_{\max}+1}^{T} y_{t-1}e_{tk}, N^{-1}\sum_{t=k_{\max}+1}^{T} Z_{t-j}'e_{tk})$. Using results established in the proof of Theorem 2,

$$N^{-1}\sum_{t=k_{\max}+1}^{T} y_{t-1}e_{tk}$$

$$= N^{-1}\sum_{t=k_{\max}+1}^{T}(a_T e_{t-1} + b_T X_{t-1})\left(e_t - (-\theta_T)^{k+1}e_{t-k-1} + \frac{c}{T}\sum_{i=1}^{k}(-\theta_T)^i y_{t-i-1}\right)$$

$$= O_p(T^{-1/2})$$

and

$$N^{-1}\sum_{t=k_{\max}+1}^{T} z_{t-j}e_{tk}$$

$$= N^{-1}\sum_{t=k_{\max}+1}^{T}(e_{t-j} - (-\theta_T)^{k+1-j}e_{t-k-1})(e_t - (-\theta_T)^{k+1}e_{t-k-1}) + O_p(T^{-1/2})$$

$$= (-\theta_T)^{2(k+1)-j}N^{-1}\sum_{t=k_{\max}+1}^{T} e_{t-k-1}^2 + O_p(T^{-1/2}),$$

$$= \sigma_e^2(-\theta_T)^{2(k+1)-j} + O_p(T^{-1/2}).$$

Therefore,

$$N^{-1}E_k'X = \sigma_e^2[0, \theta_T^{2(k+1)-1}, \ldots, (-\theta_T)^{2(k+1)-k}] + O_p(T^{-1/2}),$$

$$= \sigma_e^2(\theta_T)^{k+1}[0, \mu'] + O_p(T^{-1/2}),$$

where $\mu' = [(-\theta_T)^k, \ldots, (-\theta_T)]$. From Lemma A.1,

(A.7) $$N^{-1}(X'X) = \sigma_e^2\begin{bmatrix} 1+x & e' \\ e & I_k + \mu\mu' \end{bmatrix} + O_p(T^{-1/2}) \equiv R + O_p(T^{-1/2}).$$

In view of the properties of $N^{-1}E_k'X$,

$$N^{-1}E_k'X(X'X)^{-1}X'E_k = (-\theta_T)^{2(k+1)}\mu'R^{-1}(k)\mu + O_p(T^{-1/2}),$$

where $R^{-1}(k)$ is the lower right $k \times k$ submatrix of R^{-1}. By partitioned matrix inversion,

$$R^{-1}(k) = [I_k + \mu\mu' - e'e/(1+x)]^{-1},$$
$$= [J_k + \mu\mu']^{-1},$$

where $J_k = I_k - ee'/(1+x)$. For a $k \times k$ matrix A and $k \times 1$ vectors u and v, $(A + uv')^{-1} = A^{-1} - A^{-1}u(1 + v'A^{-1}u)^{-1}v'A^{-1}$. Hence, for $\lambda = 1 + \mu'J_k^{-1}\mu$,

$$R^{-1}(k) = J_k^{-1} - J_k^{-1}\mu\mu'J_k^{-1}/\lambda,$$
$$\mu'R^{-1}(k)\mu = \mu'J_k^{-1}\mu - (\mu'J_k^{-1}\mu)^2/\lambda,$$
$$= \frac{\mu'J_k^{-1}\mu}{1 + \mu'J_k^{-1}\mu}.$$

Since $J_k = I_k - ee'/(1+x)$, $J_k^{-1} = I_k + ee'/x$, and

$$\mu'J_k^{-1}\mu = \mu'\mu + \mu'ee'\mu/x = \mu'\mu + (-\theta_T)^{2k}/x.$$

We also have $\mu'\mu = \sum_{i=1}^{k}(-\theta_T)^{2i}$, hence

$$\mu'\mu = \frac{(-\theta_T)^2 - (-\theta_T)^{2(k+1)}}{1 - (-\theta_T)^2}, \qquad 1 + \mu'\mu = \frac{1 - (-\theta_T)^{2(k+1)}}{1 - (-\theta_T)^2}.$$

After some simplications,

$$\frac{\mu'J_k^{-1}\mu}{1 + \mu'J_k^{-1}\mu} = \frac{(-\theta_T)^2[1 - (-\theta_T)^{2k}] + (-\theta_T)^{2k}[1 - (-\theta_T)^2]/x}{[1 - (-\theta_T)^{2(k+1)}] + (-\theta_T)^{2k}[1 - (-\theta_T)^2]/x},$$

$$= (-\theta_T)^2\left[\frac{1 - (-\theta_T)^{2k}}{1 - (-\theta_T)^{2(k+1)}}\right] + O_p(T^{-1/2}).$$

Hence,

$$N^{-1}E_k'X(X'X)^{-1}X'E_k = (-\theta)^{2(k+1)}(-\theta_T)^2\left[\frac{1 - (-\theta_T)^{2k}}{1 - (-\theta_T)^{2(k+1)}}\right] + O_p(T^{-1/2}).$$

It follows that

$$\hat{\sigma}_k^2 = \sigma_e^2\left(1 + (-\theta_T)^{2(k+1)} - (-\theta_T)^{2(k+2)}\left[\frac{1 - (-\theta_T)^{2k}}{1 - (-\theta_T)^{2(k+1)}}\right]\right) + O_p(T^{-1/2}),$$

$$\ln(\hat{\sigma}_k^2) = \ln(\sigma_e^2) + \left[\frac{(-\theta)^{2(k+1)}(1 - (-\theta_T)^2)}{1 - (-\theta_T)^{2(k+1)}}\right] + O_p(T^{-1/2}).$$

Lemma 1 is a consequence of Theorem 2 and Lemma A.5.

PROOF OF THEOREM 3

The strategy of the proof is to first determine the rate of increase of k through a global minimization of the objective function. Once this rate is found, the proportionality factor is deduced. In the following, we use the notation $k \propto f(T)$ for the case where $k/f(T) \to c$ for some finite constant c strictly above 0. We also let $d \geq 0$ be a finite constant not necessarily the same throughout.

For both the IC and MIC, it is readily seen that the objective function is $O_p(C_T/T^{1/2})$ when $k/T^{1/2} \to \kappa(> 0)$, and that it is $O_p(C_Tk/T)$ when $k/T^{1/2} \to \infty$. Hence, irrespective of the choice of C_T, having $k/T^{1/2} \to \kappa$ will always, in large samples, yield a smaller value of the objective function than with $k/T^{1/2} \to \infty$. Hence, we only need to compare the cases $k/T^{1/2} \to \kappa(> 0)$ and $k/T^{1/2} \to 0$.

We start with the analysis for the *IC*. When $k/T^{1/2} \to 0$, there are two cases of interest. The first is when $k^2 C_T/T \to d$ in which case the objective function is $O_p(1/k)$. The minimal value is obtained if k is chosen as the maximal value permissible, which means $k \propto \sqrt{T/C_T}$. The minimized objective function is then $O_p(\sqrt{C_T}/\sqrt{T})$. The second case is when $T/k^2 C_T \to d$ and the objective function is $O_p(kC_T/T)$, which is minimized by taking the minimal permissible rate of increase for k, again $k \propto \sqrt{T/C_T}$, and the minimized objective function is then $O_p(\sqrt{C_T}/\sqrt{T})$. So the case with $k/T^{1/2} \to \kappa(>0)$ yields an objective function of order $O_p(C_T/T^{1/2})$ and the case $k/T^{1/2} \to 0$ an objective function of order $O_p(\sqrt{C_T}/\sqrt{T})$. Hence, $k \propto \sqrt{T/C_T}$. To derive the constant of proportionality, note that if C_T is fixed, $k = \kappa\sqrt{T}$ and the objective function is (ignoring constants)

$$\frac{2\delta}{\sqrt{T}} \vartheta(\kappa) + \frac{C_T}{\sqrt{T}} \kappa.$$

Taking derivatives and equating to zero yields the stated solution. When $C_T \to \infty$ as $T \to \infty$, the objective function (ignoring constants) is

$$\frac{1}{\kappa\sqrt{T/C_T}} + \kappa \frac{C_T\sqrt{T/C_T}}{T}$$

and the value of κ that minimizes this expression is 1. Consider now the $MIC(k)$. The first case is where

$$(A.8) \qquad \frac{C_T}{k} \to d \quad \text{and} \quad \frac{C_T k^2}{T} \to d.$$

Then the objective function when $k/T^{1/2} \to 0$ is $O_p(k^{-1})$. Hence, the highest possible rate is chosen and is given by $k \propto \sqrt{T/C_T}$ and the minimized objective function is of order $O_p(\sqrt{C_T/T})$. Since the objective function when $k/T^{1/2} \to \kappa(>0)$ is $O_p(C_T/T^{1/2})$, we have $k \propto T^{1/2}$ if C_T is fixed and $k \propto \sqrt{T/C_T}$ if $C_T \to \infty$. However, in view of the requirement that $C_T/k \to d$ this implies the need for $C_T^3/T \to d$. To obtain the constant of proportionality when C_T is fixed, we substitute $k = \kappa\sqrt{T}$ in (16) and equate its first derivative to zero and solve for κ. When C_T is increasing, we substitute $k = \kappa\sqrt{T/C_T}$ in (15), and again equate the first derivative to zero and solve for κ.

Consider now the case where $C_T^3/T \to \infty$. There are two cases to cover when either of the conditions in (A.8) is not satisfied. Suppose first that $k^3/T \to d$, in which case $k/C_T \to d$. The objective function in the case $k/T^{1/2} \to 0$ is $O_p(C_T/k^2)$ and the highest rate for k is chosen, namely $k \propto T^{1/3}$. The minimized objective function is then of order $O_p(C_T/T^{2/3})$, which is of a smaller order than the case with $k/T^{1/2} \to \kappa(>0)$. Hence, $k \propto T^{1/3}$ is the resulting rate. If $T/k^3 \to d$, in which case $C_T k^2/T \to \infty$, the objective function is of order $O_p(kC_T/T)$ and the smallest rate for k is optimal. So again, $k \propto T^{1/3}$ and the minimized objective function is also of order $O_p(C_T/T^{2/3})$. To obtain the factor of proportionality, let $k = \kappa T^{1/3}$. The objective function is then (ignoring constants)

$$\frac{1}{\kappa T^{1/3}} + \frac{C_T Z_1}{\kappa^2 T^{2/3}} + \frac{C_T \kappa}{T^{2/3}}.$$

When $C_T^3/T \to \infty$, the first term is negligible compared to the last two. Hence, minimizing the last two terms with respect to κ yields $\kappa = (2Z_1)^{1/3}$.

REFERENCES

BERK, K. N. (1974): "Consistent Autoregressive Spectral Estimates," *The Annals of Statistics*, 2, 489–502.

DEJONG, D. N., J. C. NANKERVIS, N. E. SAVIN, AND C. H. WHITEMAN (1992): "The Power Problem of Unit Root Tests in Time Series with Autoregressive Errors," *Journal of Econometrics*, 53, 323–343.

DICKEY, D. A., AND W. A. FULLER (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.

DUFOUR, J. M., AND M. KING (1991): "Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary and Nonstationary Errors," *Journal of Econometrics*, 47, 115–143.

ELLIOTT, G., T. J. ROTHENBERG, AND J. H. STOCK (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836.

FRANSES, P. H., AND N. HALDRUP (1994): "The Effects of Additive Outliers on Tests of Unit Roots and Cointegration," *Journal of Business and Economic Statistics*, 12, 471–478.

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

GOURIEROUX, C., AND A. MONFORT (1995): *Statistical and Econometric Models*, Vol. 1. Cambridge: Cambridge University Press.

HANNAN, E. J., AND M. DEISTLER (1988): *The Statistical Theory of Linear Systems*. New York: John Wiley.

LOPEZ, J. H. (1997): "The Power of the ADF Test," *Economics Letters*, 57, 5–10.

NABEYA, S., AND P. PERRON (1994): "Local Asymptotic Distribution Related to the AR(1) Model with Dependent Errors," *Journal of Econometrics*, 62, 229–264.

NG, S., AND 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.

——— (2000): "A Note on the Selection of Time Series Models," Manuscript, Department of Economics, Boston College.

PANTULA, S.G. (1991): "Asymptotic Distributions of Unit Root Tests When the Process is Nearly Stationary," *Journal of Business and Economic Statistics*, 9, 63–71.

PERRON, P., AND S. NG (1996): "Useful Modifications to Unit Root Tests with Dependent Errors and their Local Asymptotic Properties," *Review of Economic Studies*, 63, 435–465.

——— (1998): "An Autoregressive Spectral Density Estimator at Frequency Zero for Nonstationarity Tests," *Econometric Theory*, 14, 560–603.

PERRON, P., AND G. RODRÍGUEZ (1998): "GLS Detrending, Efficient Unit Root Tests and Structural Change," manuscript, Boston University.

PHILLIPS, P. C. B. (1987): "Time Series Regression with Unit Roots," *Econometrica*, 55, 277–302.

PHILLIPS, P. C. B., AND P. PERRON (1988): "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.

SAID, S. E., AND 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.