Cowles Foundation Paper 867

Journal of Business & Economic Statistics, April 1994, Vol. 12, No. 2

Approximately Median-Unbiased Estimation of Autoregressive Models

Donald W. K. ANDREWS and Hong-Yuan CHEN
Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520-8281

This article introduces approximately median-unbiased estimators for univariate AR(p) models with time trends. Confidence intervals also are considered. The methods are applied to the Nelson–Plosser macroeconomic data series, the extended Nelson–Plosser macroeconomic data series, and some annual stock-dividend and price series. The results show that most of the series exhibit substantially greater persistence than least squares estimates and some Bayesian estimates suggest. For example, for the extended Nelson–Plosser data set, 8 of the 14 series are estimated to have a unit root, but 6 are estimated to be trend stationary. In contrast, the least squares estimates indicate trend stationarity for all of the series.

KEY WORDS: Bias correction; Macroeconomic time series; Time trend; Trend stationary; Unit root.

1. INTRODUCTION AND SUMMARY

This article focuses on methods for, and applications of, median-unbiased (MU) estimation and confidence-interval (CI) construction in univariate pth order autoregressive [AR(p)] models with time trends. This focus reflects our interest in assessing the degree of persistence exhibited by various economic time series. In particular, the time series we analyze here include the 14 Nelson–Plosser (NP) macroeconomic time series, the 14 extended Nelson–Plosser (ENP) time series, and 6 stock-dividend and price series that have received considerable attention in the literature. Our interest in persistence of economic time series is in common with many recent works in empirical macroeconomics that focus on the question of whether economic time series possess a unit root or are trend stationary.

There are two motivations for this article. The first is the emphasis placed in the unit-root literature on hypothesis testing. A problem that arises with hypothesis tests in the unit-root context is that tests have low power in many scenarios of empirical interest, including those analyzed here. In such cases, point and interval estimators can be used to provide more information than that given by unit-root tests.

The second motivation is the recent growth in Bayesian estimation methods for the models considered here (e.g., see the fourth issue of the 1991 edition of the *Journal of Applied Econometrics*, which is devoted entirely to this subject). This development of Bayesian methods is very useful. We feel, however, that a corresponding development of classical estimation methods also is likely to provide useful tools and applied results. In particular, classical estimation methods that exhibit unbiasedness properties can provide results that exhibit a degree of impartiality that may not be attainable via Bayesian methods.

The problem with using the standard classical estimators [i.e., the least squares (LS) estimator] in the AR(p) model with time trend is that of bias. The LS estimators of key parameters exhibit substantial biases. In particular, for estimating the sum of the AR coefficients, α, the bias tends to be downward and quite large. For estimating the coefficient on the time trend, β, the bias is upward and quite large. In consequence, the LS estimator is a misleading indicator of the true values of α and β.

To deal with the problem of bias, this article introduces a bias correction for the LS estimator. The proposed method is an extension to AR(p) models of the exactly MU estimation method introduced by Andrews (1993) for AR(1) models. The extended method yields only approximately, not exactly, MU estimators. Here the *approximation* occurs both in the usual statistical sense (due to the use of estimators rather than true parameters in one stage) and in a numerical sense (due to the use of pseudorandom numbers).

The long-run persistence properties of time series are exhibited by their impulse response function (IRF). For a series with a unit root, the IRF never dies out. For a trend-stationary series, on the other hand, the IRF does die out. In either case, the magnitude of the IRF across different time horizons indicates how much persistence is present in the series.

It is often useful to focus attention on a scalar measure of persistence rather than consider the whole IRF. In this article, we develop an MU estimator and CI's for such a measure. The measure we focus on is the cumulative impulse response (CIR)—that is, the sum of the IRF over all time horizons. This measure has the attribute that its relation to the persistence of the series is immediate—it is a simple function of the IRF. In addition, the CIR is a monotone transformation of the spectral density function at zero frequency. In AR(p) models, the CIR equals $1/(1-\alpha)$, where α is the sum of the AR coefficients. Thus our analysis can focus on the parameter α (since median unbiasedness and CI coverage probabilities are invariant under monotone transformations).

The essence of our bias-correction method for the LS estimator of α is as follows: If the LS estimate of α equals .8, say, one does not use .8 as the estimate α but rather one uses the value of α that yields the LS estimator to have a median of .8. If the distribution of the LS estimator of α depends only on α and is monotone in α, as in the AR(1) case, then the resultant estimator is exactly median unbiased. If the distribution of the LS estimator of α depends on some nuisance parameters as well as on α, as in the AR(p) case for $p > 1$, then we use an iterative procedure that jointly estimates α and the nuisance parameters and yields only an approximately MU estimate of α. In fact, simulations reported later show that the approximation is very good—the proposed estimator is essentially median unbiased.

Once we obtain the approximately MU estimator of α, we impose this estimate on the model and run an augmented Dickey–Fuller LS regression to obtain estimates of the other parameters in the model. Again, simulations show that the resultant procedure leads to estimators that are essentially median unbiased. We obtain standard-error estimates for all of the parameter estimates via simulation.

The method for obtaining an MU estimator of α can be extended to generate approximate CI's for α. In addition, it leads to an approximately unbiased model-selection procedure for determining whether a data series has a unit root or is trend stationary. This selection procedure chooses the unit-root model if the bias-corrected estimator of α equals 1 and chooses the trend-stationary model otherwise. The term *approximation* is used here and later in the same sense as previously.

An alternative scalar parameter that has been considered in the literature to be a parameter of interest in AR(p) models is the magnitude of the largest root of the model. For example, Stock (1991) developed asymptotic CI's for this parameter and DeJong and Whiteman (1991a,b) considered Bayes estimators of this parameter. We show in the following that the persistence properties of two series with the same dominant root can be very different depending on the values of the other roots. In consequence, the empirical results of Stock (1991) and DeJong and Whiteman (1991a,b) are not as informative as is desirable.

Although Stock (1991) focused on what we consider the wrong parameter, his method requires considerably less computational effort than is required by our method outlined previously. For this reason, we provide a trivial extension of his method to the estimation of the parameter α that we consider to be of main interest. Using his tables and a simple iterative scheme, one can compute asymptotically MU estimates of α and corresponding CI's for it.

The procedures just described are applied here to three different data sets. The first is the Nelson and Plosser (1982) (NP) data set. Nelson and Plosser applied tests for unit roots on these data series and found that they could reject the null hypothesis of a unit root for only 1 of the 14 series, viz., the unemployment rate. Subsequently, however, several authors pointed out that the tests employed by Nelson and Plosser have relatively low power against relevant trend-stationary alternatives given the sample sizes employed (e.g., see De Jong, Nankervis, Savin, and Whiteman 1992).

As an alternative to classical hypothesis tests, Bayesian estimation methods have now been applied to the NP data series (see DeJong and Whiteman 1991a, Phillips 1991a, and Zivot and Phillips 1991). The results are mixed with respect to the degree of persistence found, depending on the priors employed and the parameters considered. DeJong and Whiteman (1991a), for instance, obtained estimates of the magnitude of the largest root that are substantially less than one for most series. Phillips (1991a) found more evidence of unit-root behavior, but still the evidence he found for it is not strong.

The MU estimates of α obtained here show considerably more persistence in the NP data than the LS, DeJong and Whiteman, or Phillips estimates show. For 3 series (out of 14) the estimates equal 1.0, for 7 series the estimates are .96 or larger, and for 13 series the estimates are .89 or larger.

A second data set we consider is an extension of the NP data set, which terminates in 1970, to include observations up to 1988. This extension was compiled by Schotman and van Dijk (1991). The MU estimates for the extended Nelson–Plosser (ENP) data set show very high levels of persistence for many of the series. Eight of the fourteen series, including all of the nominal variables except money stock, have α estimates equal to 1.0. Most of the real variables—including real gross national product (GNP), real per capita GNP, industrial production, and employment—have α estimates in the range .86 to .91, which corresponds to considerable persistence although less than unit-root-like behavior. In fact, for the former two series, as well as for the unemployment rate, the null hypothesis of a unit root can be rejected at the 5% level using the ENP data. Nevertheless, the overall picture obtained from the ENP data set is one of noticeably greater persistence than with the NP data set.

The third data set we analyze consists of annual series for Dow–Jones (DJ) dividends and prices (1928–1978), New York Stock Exchange (NYSE) dividends and prices (1926–1981), and Standard and Poor's (S&P) dividends and prices. We use the same data as DeJong and Whiteman (1991b), some of which were compiled by Shiller (1981). Interest in the unit-root-versus-trend-stationarity question for these data series arises because of their implications for volatility tests of the perfect-markets hypothesis as initiated by Shiller (1981). DeJong and Whiteman (1991b) presented some Bayesian estimates of the magnitude Λ of the largest root of AR(3) models with time trend fitted to the preceding data series. Their estimates for Λ were quite low—.72, .76, .77, .84, .72, and .87, respectively.

In contrast, our MU parameter estimates are considerably larger. The CI's obtained are very wide, however, so a key feature of our results is that for most of the series it is not possible to make definitive statements one way or another regarding the unit-root/trend-stationary question. Our parameter estimates for α for these series are .79, .91, .90, 1.0, .82, and .94, respectively. Our corresponding estimates for Λ for these series are nearly the same—.79, .92, .90, 1.0, .77, and .94. Our estimators are essentially median unbiased, whereas DeJong and Whiteman's (1991b) Bayesian procedure appears to have a substantial downward bias.

Next we mention several additional related articles. First, a method similar to that introduced here has been considered recently by Rudebusch (1992). Rudebusch's procedure differs from that considered here in that he aimed for MU estimators of each of the AR(p) parameters, whereas we focus on the single parameter α. His procedure was subject to the criticism that the existence and uniqueness of his estimator is an open question. In addition, Rudebusch obtained estimates of the AR parameters but did not provide any measure of the variability of these estimates. Other related works include those of Quenouille (1949, 1956), Hurwicz (1950), Marriott and Pope (1954), Kendall (1954), Orcutt and Winokur (1969), Stine and Shaman (1989), and Fair (1992).

The remainder of this article is organized as follows. Section 2 defines the model to be considered and provides a discussion of the parameters of interest. Section 3 introduces approximately MU estimators, CI's, and approximately unbiased model-selection procedures. Section 4 extends the local-to-unity asymptotic results of Stock (1991) to cover the parameter α. Sections 5–7 provide the empirical results.

2. THE MODEL AND PARAMETERS OF INTEREST

2.1 Definition of the Model

The model we consider is an AR(p) model with intercept and time trend. It can be written in an unobserved-components form and in a regression form. In unobserved-components form, it is given by

$$Y_t = \mu^* + \beta^* t + Y_t^* \quad \text{for} \quad t = -p+1, \ldots, T,$$

$$Y_t^* = \alpha Y_{t-1}^* + \psi_1 \Delta Y_{t-1}^* + \cdots + \psi_{p-1} \Delta Y_{t-1}^* + U_t \quad \text{for} \quad t = 1, \ldots, T,$$

$$U_t \sim \text{iid} N(0, \sigma^2) \quad \text{for} \quad t = 1, \ldots, T, \tag{2.1}$$

where $\{Y_t : t = -p+1, \ldots, T\}$ is the observed series. The variable ΔY_t^* denotes $Y_t^* - Y_{t-1}^*$. The parameters $(\mu^*, \beta^*, \sigma^2, \alpha)$ satisfy $\mu^* \in R, \beta^* \in R, \sigma^2 > 0$, and $\alpha \in (-1, 1]$. When $\alpha = 1$ the model is nonstationary. The parameters $(\psi_1, \ldots, \psi_{p-1})$ are such that the AR model for Y_t^* is stationary when $\alpha \in (-1, 1)$ and the AR model for ΔY_t^* is stationary when $\alpha = 1$. The initial values of Y_t^*—that is, $(Y_{-p+1}^*, \ldots, Y_0^*)$—are taken to be such that $\{Y_t^* : t \geq -p+1\}$ is stationary when $\alpha \in (-1, 1)$ and $\{\Delta Y_t^* : t \geq -p+2\}$ is stationary when $\alpha = 1$. The level of the ΔY_t^* series is arbitrary when $\alpha = 1$. [That is, when $\alpha = 1$, the initial random variable (rv) Y_{-p+1}^* can be fixed or can have any distribution provided the subsequent Y_t^* values are such that ΔY_t^* is stationary.]

The regression form of Model (2.1) is given by

$$Y_t = \mu + \beta t + \alpha Y_{t-1} + \psi_1 \Delta Y_{t-1} + \cdots + \psi_{p-1} \Delta Y_{t-p+1} + U_t \quad \text{for} \quad t = 1, \ldots, T$$

$$\mu = \mu^*(1-\alpha) + (\alpha - \psi_1 - \cdots - \psi_{p-1})\beta^*$$
$$\beta = \beta^*(1-\alpha), \tag{2.2}$$

where $(Y_{-p+1}, \ldots, Y_0)$ and $\{U_t : t = 1, \ldots, T\}$ are as defined in (2.1). The AR(p) model for Y_t in (2.2) is written in *augmented Dickey–Fuller* regression form. It can be written in standard AR(p) regression form as

$$Y_t = \mu + \beta t + \gamma_1 Y_{t-1} + \gamma_2 Y_{t-2} + \cdots + \gamma_p Y_{t-p} + U_t. \tag{2.3}$$

The parameter α in the augmented Dickey–Fuller form equals the *sum of the AR coefficients* $(\gamma_1, \ldots, \gamma_p)$. As will be argued, the augmented Dickey–Fuller parameterization is the most useful for the purposes of this article. The parameters $(\psi_1, \ldots, \psi_{p-1})$ and $(\gamma_1, \ldots, \gamma_p)$ are related via $\psi_j = -(\gamma_{j+1} + \cdots + \gamma_p)$ for $j = 1, \ldots, p-1$.

Note that the parameter β on the time trend is necessarily 0 when $\alpha = 1$ in (2.2) and (2.3). This is a desirable feature of the model because it implies that the mean of Y_t is a linear function of t for all $\alpha \in (-1, 1]$. If $\beta \neq 0$ was allowed when $\alpha = 1$, then the mean of Y_t would be a linear function of Y_t when $\alpha \in (-1, 1)$ but a quadratic function of t when $\alpha = 1$. This discontinuity is naturally avoided in the preceding model.

2.2 Scalar Measures of Persistence

In this article, as in many empirical articles, in the macroeconomic and financial literature, we are interested in assessing the persistence of a time series Y_t. In particular, we are interested in the persistence of shocks to the series. The IRF is a suitable measure of such persistence. The IRF traces out the effect of a change in the innovation U_t by a unit quantity on the current and subsequent values of Y_t. In particular, if Y_t is the series based on the innovations $\{U_1, U_2, \ldots\}$ and $\tilde{Y}_t$ is the series based on $\{U_1, \ldots, U_{t-1}, U_t + 1, U_{t+1}, U_{t+2}, \ldots\}$, then

$$\text{IRF}(h) = \tilde{Y}_{t+h} - Y_{t+h} \quad \text{for} \quad h = 0, 1, \ldots. \tag{2.4}$$

By linearity, the IRF does not depend on t, on the values $\{U_1, U_2, \ldots\}$, or on the parameters (μ^*, β^*). It only depends on $(\alpha, \psi_1, \ldots, \psi_{p-1})$. The IRF can be computed by supposing $\mu^* = \beta^* = 0$ and then by calculating the (infinite-order) moving average representation of Y_t. The coefficient on U_{t-h} in this representation is IRF(h). That is, when $\mu^* = \beta^* = 0$, we can write

$$Y_t = (1 - \gamma_1 L - \cdots - \gamma_p L^p)^{-1} U_t = \sum_{h=0}^{\infty} c_h U_{t-h},$$
$$\text{IRF}(h) = c_h, \tag{2.5}$$

for $h = 0, 1, \ldots$, where L is the lag operator.

Being an infinite vector of numbers, the IRF is a rather unwieldy measure of persistence. In consequence, it is often convenient to have a scalar measure of persistence that summarizes the information contained in the IRF. One such measure is the CIR. It is defined by

$$\text{CIR} = \sum_{h=0}^{\infty} \text{IRF}(h). \tag{2.6}$$

The CIR yields an especially useful summary of the IRF if one is dealing with different series for which the IRF's are of the same basic shape. This is the case for the data

Table 1. Comparison of Impulse Response Functions for Two Pairs of Models

Data series mimicked	Order of AR model	Magnitude of largest root	α	Magnitudes of other roots	$\psi_1, \psi_2, \ldots$	Impulse response function										
						1	2	3	4	5	7	10	15	20	25	30
Industrial production—NP	6	.95	.92	.80,.80,.78, .78,.70	.05,−.08,.01, −.08,−.26	1.0	.8	.7	.6	.3	.3	.4	.2	.2	.2	.1
Nominal wages—NP	3	.94	.96	.36,.36	.53,−.12	1.5	1.6	1.5	1.4	1.3	1.2	.9	.7	.5	.4	.3
Unemployment rate—ENP	3	.81	.80	.50,.50	.22,−.20	1.0	.6	.4	.4	.3	.2	.1	.04	.01	.01	.00
Money stock—ENP	2	.81	.96	.81	.65	1.6	1.9	2.1	2.1	2.0	1.7	1.1	.5	.2	.06	.02

series considered here. With few exceptions, the series considered have IRF's that start at 1, increase monotonically and smoothly for several periods, and then decrease monotonically and smoothly to 0. In some cases, there is no increase over the first few periods. In a few cases, the IRF becomes negative for some large values of h, but the magnitudes of such negative values are always small (.03 or less). Also, in a few cases the decrease to 0 is not completely monotone but exhibits some small wiggles.

If one is considering several series whose IRF's are of quite different shapes, then the CIR may not be sufficiently informative about the difference in their IRF's. Consider the following two examples. The first example is the case in which one series has an everywhere-positive IRF and another series has an IRF that oscillates between positive and negative values. The two series could have equal CIR's but quite different IRF's due to the cancellation of positive and negative terms for the second series.

The second example is the case in which one series is given by $Y_t = \alpha Y_{t-1} + U_t$ and another series is given by $Y_t = \alpha Y_{t-k} + U_t$ for some $k > 1$. The IRF functions of these series are (a) IRF(h) = α^h for $h = 0, 1, \ldots$ and (b) IRF(h) = α^b for $h = bk$ for $b = 0, 1, \ldots$ and IRF(h) = 0 otherwise, respectively. The CIR's of these two series are identical, but their IRF's are noticeably different with the latter exhibiting more persistence as k is increased. (We thank Chris Sims for suggesting this example.)

Fortunately, neither of the two examples just mentioned, in which the CIR is noticeably deficient, are of real concern for the economic applications we shall consider. In no cases are there IRF's with substantial positive and negative terms. In no cases are there IRF's with the nonmonotone and nonsmooth behavior of that of the model $Y_t = \alpha Y_{t-k} + U_t$ for $k > 1$. The one feature of the IRF's that appears empirically but is not captured by the CIR is the difference between a relatively large initial increase and a subsequent quick decrease in the IRF and a relatively small initial increase and subsequent slow decrease in the IRF. Two series can have the same CIR but somewhat differently shaped IRF's due to such differences. In the empirical applications, differences of this sort arise, but they are not extreme.

Based on the preceding discussion, we conclude that the CIR yields a fairly good scalar summary of the IRF, at least for the type of data series that are of interest here. In addition, the CIR is a simple function of the parameters of the model

$$\text{CIR} = \frac{1}{1-\alpha}. \tag{2.7}$$

The fact that the CIR is directly related to α in such a simple way means that one can rely on α as a measure of the persistence of a series. Different values of α can be interpreted easily in terms of persistence because they correspond straightforwardly to different values of the CIR. It is for this reason that we use the augmented Dickey–Fuller parameterization of the AR(p) model in (2.2) rather than the standard AR parameterization in (2.3).

The parameter α can be interpreted as a measure of persistence in a second way—via the spectrum of Y_t. This interpretation was discussed by Phillips (1991b). The spectrum at zero frequency is a measure of the low-frequency autocovariance of the series. For the model (2.1)–(2.3), it is given by

$$\text{spectrum at zero} = \frac{\sigma^2}{(1-\alpha)^2}. \tag{2.8}$$

Thus by this measure too persistence of Y_t depends directly on the magnitude of the parameter α.

Before deciding to emphasize the parameter α as an appropriate scalar measure of persistence, we need to consider another possibility—the magnitude of the largest root of the AR(p) model. The latter parameter was relied on by DeJong and Whiteman (1991a,b) and Stock (1991), among others.

The magnitude of the largest root of the AR(p) model turns out to be a very poor summary measure of the IRF. The reason is simply that the shape and height of the IRF depends on more than just the magnitude of the largest root. Depending on the values of the other roots, one can observe a very wide range of different persistence properties from series that have the same magnitude of largest root. This is illustrated by Table 1. Table 1 considers two pairs of models. Each model corresponds to an estimated model (estimated via the approximately MU method described later) using the NP or ENP data. (The reason for considering estimated models is to ensure empirical relevance of the results. We are not considering pathological cases in Table 1.)

In sum, the parameter α is a fairly reliable measure of the persistence of a series because it alone determines both the CIR and the spectrum at 0 of the series. On the other hand, we find that the magnitude of the largest root does not provide an adequate summary measure of the IRF. The other roots have too great an effect on the persistence of the series to rely solely on the magnitude of the largest root. For these reasons we focus attention in the following primarily on the estimation of α and secondarily on the estimation of the other parameters. We note that a graphical presentation provides an alternative to a scalar measure of persistence (e.g., see Gallant, Rossi, and Tauchen 1993).

3. APPROXIMATELY MEDIAN-UNBIASED ESTIMATORS

3.1 Definition of the Approximately Median-Unbiased Estimators

Here we describe a method for obtaining approximately MU estimators of the parameters of the augmented Dickey–Fuller model (2.2). The method is an extension of an exactly MU estimation procedure introduced by Andrews (1993) for the AR(1) version of Model (2.2).

We start by defining median unbiasedness. By definition, a number m is a *median* of an rv X if

$$P(X \geq m) \geq \frac{1}{2} \quad \text{and} \quad P(X \leq m) \geq \frac{1}{2}. \tag{3.1}$$

This definition of a median allows for nonuniqueness, but all of the medians considered here are unique. It also allows for the median of X to be a probability mass point of X. This feature of the definition is used here. If a median m of X is not a probability mass point, then $P(X > m) = P(X < m) = \frac{1}{2}$.

Let $\hat{\alpha}$ be an estimator of the parameter α. By definition, $\hat{\alpha}$ is *median unbiased* for α if the true parameter α is a median of $\hat{\alpha}$ for each α in the parameter space. The condition of median unbiasedness has the intuitive impartiality property that the probability of underestimation equals the probability of overestimation. This holds unless the true parameter value is estimated with positive probability, and in this case the probabilities of underestimation and overestimation are each less than one-half. In scenarios in which the magnitude of a parameter is a contentious issue, such as in the (trend) stationary-versus-unit-root debate, this impartiality property is quite useful. Advocates of one view are not likely to accept estimates that are biased toward a different view. MU estimators are more likely to be acceptable to a broad audience than biased estimators because they do not favor any particular outcome. For a comparison of median and mean unbiasedness, see Andrews (1993).

We note that in the classical Gaussian linear-regression model with fixed regressors the LS estimator is median unbiased. In fact, it is the best MU estimator for a wide variety of loss functions (see Andrews and Phillips 1987). In the AR(p) model (2.2), on the other hand, the LS estimator is not median unbiased and hence does not possess the same optimality properties.

Next, we describes the method used by Andrews (1993) for obtaining exactly MU estimators of α in the AR(1) version of Model (2.2). Suppose that $\hat{\alpha}$ is an estimator whose median function $m(\alpha)(= m_T(\alpha))$ is uniquely defined, depends only on α, and is strictly increasing on the parameter space $(-1,1]$. Then an MU estimator, $\hat{\alpha}_U$, of α is given by

$$\begin{aligned} \hat{\alpha}_U &= 1 && \text{if} \quad \hat{\alpha} > m(1) \\ &= m^{-1}(\hat{\alpha}) && \text{if} \quad m(-1) < \hat{\alpha} \leq m(1) \\ &= -1 && \text{if} \quad \hat{\alpha} \leq m(-1), \end{aligned} \tag{3.2}$$

where $m(-1) = \lim_{\alpha \to -1} m(\alpha)$ and $m^{-1} : (m(-1), m(1)] \to (-1, 1]$ is the inverse function of $m(\cdot)$ that satisfies $m^{-1}(m(\alpha)) = \alpha$ for $\alpha \in (-1, 1]$. Thus, if the observed value of $\hat{\alpha}$ is .8, say, one does not use .8 as the estimate of α, but rather one uses the value of α that yields the estimator $\hat{\alpha}$ to have a median of .8. This method was applied in Andrews (1993) with $\hat{\alpha}$ equal to the LS estimator of α for Model (2.2) with $p = 1$. [The general method is not due to Andrews (1993); e.g., it more or less corresponds to the method discussed by Lehmann (1959, sec. 3.5, p. 83).]

For higher-order versions of Model (2.2) (i.e., $p > 1$), the LS estimator of α has a distribution that depends on more parameters than just α. In consequence, the exact bias-correction method outlined previously cannot be applied. In fact, the LS estimator of α, $\hat{\alpha}_{LS}$, has distribution that depends on $(\alpha, \psi_1, \ldots, \psi_{p-1})$. It does not depend on μ^*, β^*, or σ^2, and when $\alpha = 1$ it does not depend on the value or distribution of the initial rv Y^*_{-p+1}; see the Appendix. [Similar invariance properties in the AR(1) model have been pointed out by several authors. For references, see Andrews (1993).] In consequence, if $(\psi_1, \ldots, \psi_{p-1})$ were known, the bias-correction method of (3.2) could be applied.

Since $(\psi_1, \ldots, \psi_{p-1})$ are unknown in practice, we suggest a simple iterative procedure that yields an approximately MU estimator. First, compute the LS estimator of $(\alpha, \psi_1, \ldots, \psi_{p-1}, \mu, \beta)$ by regressing Y_t on $(Y_{t-1}, \Delta Y_{t-1}, \ldots, \Delta Y_{t-p+1}, 1, t)$, call it $(\hat{\alpha}_{LS1}, \hat{\psi}_{1,LS1}, \ldots, \hat{\psi}_{p-1,LS1}, \hat{\mu}_{LS1}, \hat{\beta}_{LS1})$. Second, treat $(\hat{\psi}_{1,LS1}, \ldots, \hat{\psi}_{p-1,LS1})$ as though they were the true values of $(\psi_1, \ldots, \psi_{p-1})$ and compute the bias-corrected estimator of α, $\hat{\alpha}_{U1}$, using (3.2). Third, treat $\hat{\alpha}_{U1}$ as though it was the true value of α and compute a second round set of LS estimates of $(\psi_1, \ldots, \psi_{p-1})$—call them $(\hat{\psi}_{1,LS2}, \ldots, \hat{\psi}_{p-1,LS2})$—by regressing $Y_t - \hat{\alpha}_{LS1} Y_{t-1}$ on $(\Delta Y_{t-1}, \ldots, \Delta Y_{t-p+1}, 1, t)$. When $\hat{\alpha}_{LS1} = 1$, exclude the regressor t in the latter regression to impose the constraint that $\beta = 0$. Next, treat $(\hat{\psi}_{1,LS2}, \ldots, \hat{\psi}_{p-1,LS2})$ as though they were the true values of $(\psi_1, \ldots, \psi_{p-1})$ to generate a second-round bias-corrected estimator of α, $\hat{\alpha}_{U2}$. Continue this procedure either for a fixed number of iterations or until convergence. For the following empirical results we specify a maximum of 10 iterations. For most of the series, convergence is obtained in 2 iterations, but one series took 4.

If $\hat{\alpha}_{LSj}$ is the final estimate of α, then $(\hat{\psi}_{1,LSj+1}, \ldots, \hat{\psi}_{p-1,LSj+1}, \hat{\mu}_{LSj+1}, \hat{\beta}_{LSj+1})$ are the final estimates of $(\psi_1, \ldots, \psi_{p-1}, \mu, \beta)$. Let

$$\begin{aligned} &(\hat{\alpha}_{MU}, \hat{\psi}_{1,MU}, \ldots, \hat{\psi}_{p-1,MU}, \hat{\mu}_{MU}, \hat{\beta}_{MU}) \\ &\quad = (\hat{\alpha}_{LSj}, \hat{\psi}_{LSj+1}, \ldots, \hat{\psi}_{p-1,LSj+1}, \hat{\mu}_{LSj+1}, \hat{\beta}_{LSj+1}) \end{aligned} \tag{3.3}$$

denote the final round approximately MU estimators. We refer to these estimators as the MU estimators.

Simulation methods can be used to compute the bias-corrected estimator defined in (3.2) given a vector $(\psi_1, \ldots, \psi_{p-1})$. More specifically, for a given value of α and fixed $(\psi_1, \ldots, \psi_{p-1})$, set $\mu^* = \beta^* = 0$ and $\sigma^2 = 1$ in (2.2) and simulate a data set $\{Y_{r,t} : t = -p+1, \ldots, T\}$ according to the model (which has Gaussian errors). Regress the simulated $Y_{r,t}$ on $(Y_{r,t-1}, \Delta Y_{r,t-1}, \ldots, \Delta Y_{r,t-p+1}, 1, t)$ to obtain a single random draw of the LS estimator $\hat{\alpha}_{\text{LS}}$. Repeat this procedure R times; that is, $r = 1, \ldots, R$($R = 1{,}000$ is used in the empirical results following) and take the sample median of the simulated LS estimates of α to be a Monte Carlo estimate of the median of $\hat{\alpha}_{\text{LS}}$ when the true parameters are $(\alpha, \psi_1, \ldots, \psi_{p-1})$. Given the observed value of $\hat{\alpha}_{\text{LS}}$, say .8, iteratively determine the value of α such that the Monte Carlo estimate of the median of the estimator $\hat{\alpha}_{\text{LS}}$ equals .8. This yields the desired estimator $\hat{\alpha}_U$. Monotonicity of the median function of $\hat{\alpha}_{\text{LS}}$ for given $(\psi_1, \ldots, \psi_{p-1})$ generally makes the iterative procedure converge quickly.

In keeping with the computer-intensive methods just employed, one can generate standard errors for all of the MU estimators $(\hat{\alpha}_{\text{MU}}, \hat{\psi}_{1,\text{MU}}, \ldots, \hat{\psi}_{p-1,\text{MU}}, \hat{\mu}_{\text{MU}}, \hat{\beta}_{\text{MU}})$ as follows. Treat the observed estimates, say $(\hat{\alpha}^0_{\text{MU}}, \hat{\psi}^0_{1,\text{MU}}, \ldots, \hat{\beta}^0_{\text{MU}})$, as though they were the true values and perform a simulation study of the estimators $(\hat{\alpha}_{\text{MU}}, \ldots, \hat{\beta}_{\text{MU}})$ with these true values. For each repetition of the simulation, one generates a simulated data set, computes the LS estimates for this data set, and then computes the corresponding approximately median unbiased estimators of $(\alpha, \psi_1, \ldots, \psi_{p-1}, \mu, \beta)$ for this data set. Having completed the desired numbers of repetitions R^*($R^* = 1{,}000$ for the cases to be reported), one has R^* realizations from the distribution of $(\hat{\alpha}_{\text{MU}}, \ldots, \hat{\beta}_{\text{MU}})$ (up to simulation error) when the true parameters are $(\hat{\alpha}^0_{\text{MU}}, \ldots, \hat{\beta}^0_{\text{MU}})$. The sample standard errors from these R^* realizations are used as estimates of the standard errors of $(\hat{\alpha}_{\text{MU}}, \ldots, \hat{\beta}_{\text{MU}})$ for the original data series.

The preceding method of simulating standard errors is straightforward but computer intensive, because each of the R^* repetitions involves computing bias-corrected estimates $(\hat{\alpha}_{\text{MU}}, \ldots, \hat{\beta}_{\text{MU}})$, which by themselves require a simulation procedure. Using a 486 33-megahertz personal computer, it took 50 hours to generate the parameter estimates and corresponding standard errors for each of the data series to be analyzed. Although slow, this performance shows the proposed method to be quite feasible. It is hoped that within a few years the required time will be reduced to a few hours on the fastest PC's.

3.2 Confidence Intervals for α

Approximate confidence intervals (CI's) for α can be obtained in a similar way to that of the approximately median-unbiased estimator of α. Suppose that $\hat{\alpha}$ is an estimator whose p_1 and p_2 quantiles are uniquely defined, depend only on α, and are strictly increasing in α on the parameter space $(-1, 1]$. Let $q_{p_1}(\alpha)$ and $q_{p_2}(\alpha)$ denote these quantile functions. Then, an exact level 100 $(1 - p_1 - p_2)$% CI for α is given by $[\hat{c}_L, \hat{c}_U]$, where

$$\begin{aligned}
\hat{c}_L &> 1 && \text{if} \quad \hat{\alpha} > q_{p_2}(1) \\
&= q_{p_2}^{-1}(\hat{\alpha}) && \text{if} \quad q_{p_2}(-1) < \hat{\alpha} \le q_{p_2}(1) \\
&= -1 && \text{if} \quad \hat{\alpha} \le q_{p_2}(-1) \\
\hat{c}_U &= 1 && \text{if} \quad \hat{\alpha} > q_{p_1}(1) \\
&= q_{p_1}^{-1}(\hat{\alpha}) && \text{if} \quad q_{p_1}(-1) < \hat{\alpha} \le q_{p_1}(1) \\
&= -1 && \text{if} \quad \hat{\alpha} \le q_{p_1}(-1).
\end{aligned} \tag{3.4}$$

In (3.4), for $i = 1, 2$, $q_{p_i}(-1) = \lim_{\alpha \to -1} q_{p_i}(\alpha)$ and $q_{p_i}^{-1} : (q_{p_i}(-1), q_{p_i}(1)] \to (-1, 1]$ is the inverse function of $q_{p_i}(\cdot)$ that satisfies $q_{p_i}^{-1}(q_{p_i}(\alpha)) = \alpha$ for $\alpha \in (-1, 1]$. Andrews (1993) used this method to construct exact CI's for α for the first-order AR version of Model (2.2). (Note that this method of constructing CI's is time honored; only the application of it in the present context is original.)

Letting $\hat{\alpha}$ of (3.4) be the LS estimator of α from the regression in (2.2), one finds that its distribution depends on $(\alpha, \psi_1, \ldots, \psi_{p-1})$ rather than just α. Hence one cannot obtain an exact CI for α using the method of (3.4). One can obtain an approximate one, however, by taking the final bias-corrected estimates of $(\psi_1, \ldots, \psi_{p-1})$ defined previously and treating them as though they were the true values. Given these values, $\hat{c}_L$ and $\hat{c}_U$ can be computed by simulation using an analogous procedure to that described previously for computing $\hat{\alpha}_U$.

3.3 An Unbiased Model-Selection Procedure

The approximately median-unbiased estimator just introduced can be used to construct approximately unbiased model-selection procedures. By definition, a model-selection procedure is *unbiased* if for any correct model the probability of selecting the correct model is at least as large as the probability of selecting each incorrect model. For example, one might want to select between the (trend) stationary model for which $\alpha \in (-1, 1)$ and the unit-root (with drift) model for which $\alpha = 1$. An unbiased selection procedure in this case has the property that if $\alpha = 1$ the probability of selecting the unit-root model is $\ge$ the probability of selecting the (trend) stationary model and if $\alpha \in (-1, 1)$ the P_α probability of selecting the (trend) stationary model is $\ge$ the P_α probability of selecting the unit-root model for each $\alpha \in (-1, 1)$. Unbiased selection procedures exhibit an intuitive impartiality property that may be useful if the selection of one model or another is a contentious issue.

The concept of unbiased selection procedures is a special case of that of risk-unbiased decision rules (see Lehmann 1959, p. 12). For selection procedures, the space of actions is finite—one chooses one model from a finite set of models. If the loss function equals 0 when the correct model is chosen and 1 otherwise, then a risk-unbiased decision rule for this problem is an unbiased selection procedure.

Consider the problem of selecting one of two models defined by $\alpha \in I_a$ and $\alpha \in I_b$, where I_a and I_b are intervals that partition the parameter space $(-1,1]$ for α. For example, one might have $I_a = (-1, 1)$ and $I_b = \{1\}$ or $I_a = (-1, .975)$ and $I_b = [.975, 1]$. [The latter were considered by DeJong and Whiteman (1991a) and Phillips (1991a).]

The selection procedure we consider here is

$$\text{choose } I_k \text{ if } \hat{\alpha}_{\text{MU}} \in I_k \text{ for } k = a, b. \tag{3.5}$$

This procedure is exactly unbiased if $\hat{\alpha}_{\text{MU}}$ is exactly median unbiased. To see this, suppose that I_a lies to the left of I_b and $\hat{\alpha}_{\text{MU}}$ is exactly median unbiased. Then, for all $\alpha \in I_a$,

$$\begin{aligned} P_\alpha(\hat{\alpha}_{\text{MU}} \in I_b) \le P_\alpha(\hat{\alpha}_{\text{MU}} > \alpha) \le \frac{1}{2} \\ \le P_\alpha(\hat{\alpha}_{\text{MU}} \le \alpha) \le P_\alpha(\hat{\alpha}_{\text{MU}} \in I_a), \end{aligned} \tag{3.6}$$

where the second and third inequalities use the median unbiasedness of $\hat{\alpha}_{\text{MU}}$. For $\alpha \in I_b$, the argument is analogous, so the selection procedure of (3.5) is unbiased. We note that the selection procedure of (3.5) is also a valid level .5 (unbiased) test in this case of $H_0 : \alpha \in I_a$ versus $H_1 : \alpha \in I_b$ and of $H_0 : \alpha \in I_b$ versus $H_1 : \alpha \in I_a$.

Since $\hat{\alpha}_{\text{MU}}$ is only approximately median unbiased when $p > 1$, the model-selection procedure of (3.5) is correspondingly only approximately unbiased. In fact, simulations reported in the next section show that $\hat{\alpha}_{\text{MU}}$ is very close to being median unbiased for several scenarios of empirical relevance. In consequence, the model-selection procedure is also very close to being unbiased at least in these scenarios.

3.4 Properties of the Approximately Median-Unbiased Estimators

Consider now the finite-sample properties of the MU estimators $(\hat{\alpha}_{\text{MU}}, \dots, \hat{\beta}_{\text{MU}})$. The main features of these estimators that are of interest are their median-bias properties and their variability relative to the LS estimators. These properties can be assessed using the same simulation procedure as is used to generate the standard error estimates. In particular, given a data series and the corresponding observed MU estimates $(\hat{\alpha}^0_{\text{MU}}, \dots, \hat{\beta}^0_{\text{MU}})$, the simulation procedure generates R^* random draws from the distribution of the LS and MU estimators of $(\alpha, \psi_1, \dots, \psi_{p-1}, \mu, \beta)$. In addition, one can compute estimates of the IRF at different time horizons and of the magnitudes of the roots of the AR(p) model corresponding to both the LS and MU estimates for each repetition.

The difference between the sample median of any of these parameter estimates over the R^* repetitions and the true value gives a Monte Carlo estimate of the median bias of the LS and MU estimators when the true parameters are $(\hat{\alpha}^0_{\text{MU}}, \dots, \hat{\beta}^0_{\text{MU}})$. Corresponding Monte Carlo estimates of the standard deviation, root mean squared error (MSE), and interquartile range of the LS and MU estimators can be computed analogously.

Table 2 provides the results of the preceding simulation procedure when the true parameter values are taken to mimic those of three different series that exhibit varying degrees of persistence. The series mimicked are the NP series for real GNP, GNP deflator, and consumer prices whose α values are .88, .96, and 1.0, respectively (as estimated by $\hat{\alpha}_{\text{MU}}$).

The results for the $\alpha = .88$ (real GNP) and $\alpha = .96$ (GNP deflator) cases with $p = 2$ show the following. The median bias of the MU estimator for all estimands is essentially 0. The median bias of the LS estimator, on the other hand, is substantial for α, μ, β, the magnitudes of the two roots, and the IRF at most time horizons. The standard deviation of the MU estimator is the same or somewhat larger than that of the LS estimator for all estimands except the IRF at long-time horizons, for which it is substantially larger. The root MSE of the MU estimator is noticeably smaller than that of the LS estimator for estimation of α, μ, β, and the IRF at short-time horizons. It is approximately equal for ψ_1, σ^2, and the magnitudes of the roots and is substantially larger for the IRF at long-time horizons. The interquartile range of the LS estimator does not include the true value for the estimands α, μ, β, and IRF$(h)\, \forall\, h \ge 3$. On the other hand, the interquartile range of the MU estimator includes the true value and is symmetrically centered around it for these estimands. For the other estimands, the interquartile range results for the two estimators are more comparable.

Next we describe the results for the $\alpha = 1.0$ (consumer prices) case with $p = 4$. The MU estimators of $(\alpha, \psi_2, \psi_3, \mu, \beta, \sigma^2)$ are essentially median unbiased, while that of ψ_1 has a small downward median bias. In contrast, the LS estimators of $(\alpha, \psi_1, \psi_2, \psi_3, \mu, \beta)$ are all significantly median biased. The MU and LS estimators of the magnitudes of the roots each have median biases. Those of the MU estimator are smaller. The MU estimator of the IRF is downward median biased, especially at long-time horizons. Its downward bias is quite small, however, in comparison to that of the LS estimator, which is huge, especially for long-time horizons. The standard deviations of the MU and LS estimators are approximately equal for all estimands except the IRF at long-time horizons, for which the MU estimator has considerably larger standard deviations. The root MSE of the MU estimator is substantially smaller than that of the LS estimator for the estimands α, μ, β, the magnitudes of the two largest roots, and IRF (h) for all h. For the other estimands, the MU and LS estimators have comparable root MSE's. The length and location of the interquartile ranges of the MU and LS estimators corroborate the results based on the standard deviations and median biases.

Central 90% CI's for α calculated as described in Section 3.2 are found to have simulated confidence levels of 88.9%, 89.7%, and 86.9% for the $\alpha = .88$, .96, and 1.0 case, respectively. These simulated confidence levels have standard errors of approximately .7% each. Thus there appears to be a tendency for the CI's coverage probabilities to be somewhat too low.

In conclusion, we find that the MU estimator achieves a substantial reduction in median bias over the LS estimator for almost all of the estimands considered. The MU estimator is essentially median unbiased for most of the estimands with the greatest exception being the IRF when $\alpha = 1$. The MU estimator pays a negligible-to-small price in terms of increased standard deviation for its improved median-bias properties, except when estimating the IRF at long-time horizons, in which case the price is large. In consequence, the root MSE of the MU estimator is noticeably smaller than that of the LS estimator for many estimands, with the main exception being the IRF at long-time horizons when $\alpha \le .96$.

The robustness of the MU estimator and CI's to nonnor-

Table 2. Properties of the Median-Unbiased (MU) Estimators for Various Parameters

	α	ψ_1	ψ_2	ψ_3	μ	$100 \times \beta$	$100 \times \sigma^2$	Magnitude of roots in descending order 1	2	3	4	Impulse response function 1	2	3	4	5	10	15	20	25	30
								Date series mimicked: real GNP-NP T + p = 62													
True parameter values	.88	.39			.54	.39	.35	.75	.52			1.27	1.22	1.06	.87	.69	.18	.04	.01	.00	.00
Median bias of MU	−.001	−.004			.00	.00	.00	.01	.00			−.01	−.01	.00	.00	−.01	−.02	−.01	.00	.00	.00
Median bias of LS	−.061	.026			.27	.19	−.01	−.05	.12			−.04	−.13	−.21	−.26	−.30	−.18	−.04	−.01	.00	.00
Standard deviation of MU	.079	.12			.35	.26	.06	.12	.19			.12	.22	.29	.36	.41	.45	.42	.41	.41	.41
Standard deviation of LS	.069	.12			.31	.24	.06	.08	.16			.13	.22	.27	.28	.28	.15	.06	.04	.02	.02
Root MSE of MU	.079	.12			.35	.26	.06	.13	.19			.12	.22	.29	.36	.41	.47	.47	.46	.45	.44
Root MSE of LS	.10	.12			.44	.31	.06	.10	.17			.14	.25	.34	.39	.41	.20	.07	.03	.03	.00
Interquartile range of MU	.83,.93	.29,.47			.32,.78	.23,.57	.31,.39	.69,.88	.35,.67			1.18,1.34	1.06,1.36	.86,1.24	.63,1.12	.42,1.0	.01,.51	.00,.26	.00,.14	.00,.08	.00,.04
Interquartile range of LS	.77,.86	.33,.49			.63,1.03	.45,.74	.30,.38	.65,.76	.49,.70			1.14,1.31	.95,1.25	.68,1.02	.42,.78	.19,.58	−.04,.10	.00,.03	.00,.01	.00,.00	.00,.00
								Date series mimicked: GNP deflator—NP T + p = 82													
True parameter values	.96	.44			.12	.11	.21	.92	.47			1.40	1.52	1.51	1.45	1.36	.93	.63	.42	.28	.19
Median bias of MU	.000	−.01			.00	−.01	.00	−.001	−.003			−.01	−.02	−.02	−.01	−.01	.01	.00	−.01	−.01	−.01
Median bias of LS	−.047	.01			.12	.13	.00	−.12	.10			−.04	−.13	−.22	−.31	−.39	−.58	−.52	−.39	−.27	−.19
Standard deviation of MU	.48	.10			.13	.13	.04	.10	.14			.11	.21	.29	.36	.42	.63	.71	.74	.76	.78
Standard deviation of LS	.044	.10			.12	.13	.03	.088	.15			.11	.21	.28	.32	.34	.33	.24	.18	.15	.12
Root MSE of MU	.048	.10			.13	.13	.04	.11	.14			.11	.21	.29	.36	.42	.63	.74	.82	.88	.92
Root MSE of LS	.070	.10			.19	.20	.03	.15	.17			.12	.25	.35	.44	.52	.62	.50	.36	.26	.18
Interquartile range of MU	.93,1.0	.36,.49			.02,.21	.00,.20.	19,.24	.83,1.0	.38,.58			1.31,1.46	1.36,1.63	1.29,1.67	1.18,1.67	1.04,1.65	.47,1.54	.19,1.52	.07,1.52	.03,1.52	.01,1.52
Interquartile range of LS	.88,.94	.38,.51			.17,.33	.17,.32	.19,.23	.73,.87	.45,.70			1.28,1.43	1.26,1.54	1.11,1.48	.92,1.36	.73,1.22	.14,.61	.01,.30	.00,.15	.00,.08	.00,.04
								Date series mimicked: consumer prices — NP T + p = 111													
True parameter values	1.0	.71	−.29	.08	.005	.00	.19	1.0	.45	.42	.42	1.71	1.92	1.94	1.95	1.97	1.98	1.98	1.98	1.98	1.98
Median bias of MU	.000	−.02	.00	.00	.000	.00	.00	.00	.13	.09	−.04	−.04	−.09	−.15	−.21	−.25	−.40	−.52	−.55	−.56	−.56
Median bias of LS	−.047	−.02	.02	.02	.000	.04	−.01	−.11	.19	.09	.00	−.08	−.19	−.32	−.44	−.57	−1.13	−1.53	−1.73	−1.85	−1.91
Standard deviation of MU	.030	.09	.12	.10	.013	.04	.03	.07	.11	.10	.18	.10	.20	.29	.37	.44	.66	.79	.85	.89	.92
Standard deviation of LS	.031	.09	.12	.10	.024	.06	.03	.07	.13	.11	.17	.10	.20	.28	.36	.41	.50	.48	.43	.40	.39
Root MSE of MU	.037	.09	.12	.10	.013	.05	.03	.09	.18	.13	.19	.11	.22	.33	.43	.51	.82	1.00	1.10	1.17	1.21
Root MSE of LS	.061	.10	.12	.10	.024	.08	.03	.14	.24	.13	.18	.12	.28	.43	.56	.69	1.22	1.51	1.66	1.75	1.80
Interquartile range of MU	.96,1.0	.63,.75	−.37,−.21	.01,.14	.000,.010	.00,.03	.17,.20	.92,1.0	.51,.65	.43,.57	.19,.49	1.61,1.74	1.69,1.96	1.59,1.99	1.48,1.99	1.40,2.00	.99,1.97	.66,1.97	.44,1.97	.28,1.96	.18,.96
Interquartile range of LS	.93,.97	.67.75	−.35,−.19	.03,.16	−.01,.02	.01,.08	.16,.20	.83,.93	.54,.76	.43,.57	.24,.51	1.57,1.70	1.59,1.86	1.42,1.82	1.28,1.76	1.13,1.69	.50,1.18	.18,.85	.05,.58	.02,.41	.01,.28

Table 3. Properties of the Approximately Unbiased Model-Selection Procedure

True parameters				Probability of selecting a model	Data series mimicked
α	ψ_1	ψ_2	$T+p$	with $\alpha = 1$	(using NP data)
.81	.21	−.20	81	.04	Unemployment rate
.87	.10	−.05	111	.04	Industrial production
.87	.39	.01	62	.12	Real GNP
.89	.23	−.02	71	.20	Real wages
.92	.39	−.11	81	.19	Employment
.95	.70	−.08	82	.19	Money stock
.96	.42	.05	82	.34	GNP deflator
.97	.27	−.18	100	.44	Common-stock prices
1.0	.74	−.27	111	.56	Consumer prices
1.0	.10	−.05	102	.55	Velocity
1.0	.18	.37	71	.59	Interest rate
1.0	.50	−.14	62	.54	Nominal GNP

mality of the innovations is discussed briefly in Section 5. The robustness results are quite similar to those reported by Andrews (1993) for the AR(1) model.

Last, we note that the standard bootstrap procedure for $\hat{\alpha}_{LS}$ is asymptotically invalid in the unit-root case (see Basawa, Mallik, McCormick, Reeves, and Taylor 1991). The MU estimator is a type of parametric bootstrap estimator. It does not suffer from the same problem in the unit-root case, however, because it does not rely on estimates of the unit-root parameter α in formulating its approximation to the quantiles of $\hat{\alpha}_{LS}$.

3.5 Properties of the Approximately Unbiased Model-Selection Procedure

Here we briefly investigate the properties of the approximately unbiased model-selection procedure introduced in Section 3.3. We consider the two models defined by $I_a = (-1, 1)$ and $I_b = \{1\}$. The selection rule is to choose the unit-root model I_b if $\hat{\alpha}_{MU} = 1$ and otherwise to choose the trend-stationary model.

Table 3 shows how the probability of selecting a unit-root model varies as a function of the true parameter α for several AR(3) models. This probability also depends on the parameters ψ_1 and ψ_2 and on the sample size $T+p$. The α, ψ_1, ψ_2, and $T+p$ combinations considered were chosen to mimic different NP data series. (That is, the true parameters listed correspond to the MU estimates for the data series listed.) The probabilities of selecting a model in which $\alpha = 1$ were calculated by simulation using 1,000 repetitions. The simulation standard errors for these probabilities range from .0062 for the $\alpha = .81$ case to .017 for the $\alpha = 1.0$ case.

The first eight rows of Table 3 show the probabilities of erroneously choosing a unit-root model for different α values less than 1.0. When the value of α is $\leq .95$, the probabilities are small ($\leq .20$) for the sample sizes considered. For α values closer to 1.0, the probabilities are larger. For example, for $\alpha = .97$, the probability is .44 when $T+p = 100$. The last four rows of Table 3 show the probabilities of correctly selecting a unit-root model when $\alpha = 1$ for several different sample sizes. These probabilities are just above .5. They are much lower than the corresponding probabilities for a level .05 test of a unit-root null hypothesis because the unbiasedness condition precludes giving the unit-root model favorable status a priori.

4. AN ASYMPTOTICALLY MEDIAN-UNBIASED ESTIMATOR OF α

A recent article by Stock (1991) used local-to-unity asymptotics to obtain CI's for the magnitude of the largest root in Model (2.2). His work builds on the local-to-unity testing results of Bobkoski (1983), Cavanagh (1985), Phillips (1987, 1988), Chan and Wei (1987), and Chan (1988) and especially on the local-to-unity CI results of Cavanagh (1985). Cavanagh (1985) considered asymptotic CI's for α in an AR(1) model without intercept or time trend. Stock (1991) extended these results to the empirically relevant case of AR(p) models with intercept and time trend.

As argued in Section 2, point or interval estimates for the magnitude of the largest root of the model (2.2) are not very useful summary measures of the persistence of a series as measured by its IRF or its spectrum at 0. In consequence, it seems worthwhile to introduce a trivial extension to Stock's methods that focuses on point and interval estimation of the parameter α, the sum of the AR coefficients, rather than on the magnitude of the largest root. The method is based on local-to-unity asymptotics and yields estimators and CI's that are easy to compute given the tables provided by Stock (1991).

In comparison with the computer-intensive methods described in Section 3, the methods considered here are very quick to compute. On the other hand, they are probably less accurate, especially when the sample size is small or α is not close to 1. In addition, they do not yield estimates and standard-error estimates for the wide range of estimands considered in Table 2 as the method of Section 3 does. As noted in Section 3, the methods there can be given asymptotic justifications even if the errors are nonnormal, just as the methods here can. Thus there is no inherent advantage of either method with respect to robustness against nonnormal errors (with several moments finite).

Our asymptotically median-unbiased (AMU) estimator $\hat{\alpha}_{AMU}$ of α and central CI $[\hat{L}, \hat{U}]$ for α of asymptotic confidence level $100(1-p_0)\%$ are defined by

$$\begin{aligned} \hat{\alpha}_{AMU} &= 1 + c_{med}\hat{b}(1)/T \\ [\hat{L}, \hat{U}] &= [1 + c_0\hat{b}(1)/T, 1 + c_1\hat{b}(1)/T], \end{aligned} \qquad (4.1)$$

where $\hat{b}(1)$ is a consistent estimator (defined later) of $b(1) = 1 - \sum_{j=1}^{p-1} \psi_j$.

The rv's c_{med}, c_0, and c_1 are determined using Stock's (1991) table A.1, part B, as follows. Let $\hat{\tau}^\tau$ denote the t statistic for testing $H_0{:}\alpha = 1$ in the regression of Y_t on $(Y_{t-1}, \Delta Y_{t-1}, \ldots, \Delta Y_{t-p+1}, 1, t)$, where α is the coefficient on Y_{t-1}. (It is often convenient for computing $\hat{\tau}^\tau$ to note that it equals the t statistic for testing whether the coefficient on Y_{t-1} is 0 in the regression of ΔY_t on $(Y_{t-1}, \Delta Y_{t-1}, \ldots, \Delta Y_{t-p+1}, 1, t)$.) In the column labeled

Table 4. Median-Unbiased (MU) and Least Squares (LS) Estimates for the Extended Nelson–Plosser Data Series

										Magnitude of roots in descending order				Impulse response function						NP data
Data series	Estimator	α	90% CI for α	ψ_1	ψ_2	ψ_3	μ	$100 \times \beta$	$100 \times \sigma^2$	1	2	3	4	1	3	5	10	15	25	α [90% CI for α]
Real GNP	MU	.864	[.77,.99]	.39			.64	.44	.27	.67	.58			1.26	1.00	.60	.11	.02	.00	.885
(1909–1988,		(.00,.06)		(−.01,.10)			(−.01,.29)	(−.02,.21)	(.00,.05)	(.07,.10)	(−.05,.17)			(−.01,.10)	(−.01,.24)	(.01,.31)	(.01,.27)	(.01,.21)	(.00,.17)	[.77,1.0]
$T + p = 80$)	LS	.824		.41			.82	.57	.27	.64	.64			1.23	.87	.40	.01	.00	.00	.825
		(−.04,.06)		(.01,.10)			(.19,.26)	(.14,.19)	(−.01,.04)	(.01,.07)	(.04,.14)			(−.04,.11)	(−.16,.22)	(−.21,.24)	(−.10,.11)	(−.02,.04)	(.00,−.01)	
Nominal GNP	MU	1.00	[.93,1.0]	.45			.035	.00	.64	1.00	.45			1.45	1.74	1.79	1.81	1.81	1.81	.958
(1909–1988,		(.00,.04)		(−.02,.10)			(.00,.03)	(.00,.27)	(−.01,.11)	(.00,.09)	(.00,.13)			(−.05,.11)	(−.18,.29)	(−.27,.42)	(−.37,.63)	(−.39,.73)	(−.40,.81)	[.86,1.0]
$T + p = 80$)	LS	.939		.47			.64	.41	.61	.87	.54			1.41	1.46	1.22	.63	.31	.08	.899
		(−.07,.04)		(−.01,.11)			(.00,.05)	(.41,.31)	(−.04,.10)	(−.14,.09)	(.07,.15)			(−.09,.11)	(−.40,.28)	(−.70,.36)	(−1.25,.40)	(−1.54,.36)	(−1.75,.31)	
Real per	MU	.858	[.77,.97]	.38			.99	.28	.28	.67	.58			1.24	.96	.56	.10	.02	.00	.875
capita GNP		(.01,.07)		(−.01,.11)			(−.03,.46)	(−.01,.14)	(.00,.05)	(.06,.10)	(−.01,.17)			(−.01,.11)	(.00,.24)	(.02,.31)	(.01,.26)	(.00,.24)	(.00,.17)	[.76,1.0]
(1909–1988,	LS	.816		.40			1.29	.36	.28	.63	.63			1.22	.83	.36	.00	.00	.00	.818
$T + p = 80$)		(−.04,.06)		(.02,.10)			(.28,.43)	(.08,.13)	(.00,.05)	(.02,.07)	(.05,.14)			(−.03,.11)	(−.14,.23)	(−.19,.24)	(−.10,.11)	(−.01,.04)	(.00,.01)	
Industrial	MU	.910	[.79,1.0]	.06	−.10	.00	.10	.34	.86	.94	.79	.79	.76	.97	.70	.30	.39	.22	.13	.919
production*		(.00,.09)		(.01,.10)	(−.01,.10)	(.00,.10)	(.00,.04)	(.01,.33)	(.00,.12)	(.00,.07)	(.02,.06)	(.01,.06)	(−.01,.07)	(−.01,.09)	(−.03,.16)	(−.01,.20)	(.00,.22)	(−.01,.22)	(−.01,.24)	[.78,1.0]
(1860–1988,	LS	.841		.10	−.06	.03	.11	.63	.85	.88	.78	.78	.75	.94	.62	.19	.22	.08	.03	.835
$T + p = 129$)		(−.08,.08)		(.05,.10)	(.03,.09)	(.04,.09)	(.03,.05)	(.30,.30)	(−.01,.11)	(−.06,.06)	(.00,.07)	(.00,.07)	(−.01,.07)	(−.04,.09)	(−.11,.15)	(−.13,.18)	(−.17,.14)	(−.16,.09)	(−.11,.06)	
Employment	MU	.904	[.82,1.0]	.40	−.11		.98	.15	.11	.86	.36	.36		1.30	.99	.71	.33	.16	.03	.914
		(.00,.06)		(−.01,.10)	(−.01,.10)		(−.04,.64)	(−.01,.10)	(.00,.02)	(.01,.11)	(.09,.13)	(.01,.16)		(−.01,.11)	(−.01,.22)	(.02,.29)	(.02,.34)	(.02,.33)	(.01,.32)	[.81,1.0]
(1890–1988,	LS	.864		.41	−.08		1.38	.21	.10	.78	.32	.32		1.27	.90	.53	.15	.04	.00	.861
$T + p = 99$)		(−.04,.06)		(.00,.10)	(.03,.10)		(.43,.57)	(.07,.09)	(.00,.02)	(−.08,.10)	(.10,.14)	(−.03,.17)		(−.05,.11)	(−.13,.21)	(−.20,.23)	(−.19,.17)	(−.12,.10)	(−.03,.05)	
Unemployment	MU	.756	[.63,.88]	.36	−.23	.22	.40	.04	13.7	.72	.72	.65	.65	1.11	.52	.27	−.03	−.01	.00	.765
rate		(.00,.09)		(.00,.10)	(.00,.10)	(−.01,.10)	(−.01,.16)	(.01,.15)	(−.07,2.08)	(.04,.07)	(.01,.09)	(.00,.09)	(−.02,.17)	(−.01,.11)	(−.01,.18)	(.01,.20)	(.02,.12)	(.01,.06)	(.00,.03)	[.62,.89]
(1890–1988,	LS	.715		.38	−.21	.23	.47	.03	13.7	.74	.74	.66	.66	1.09	.47	.18	−.06	.00	.00	.706
$T + p = 99$)		(−.05,.08)		(.03,.10)	(.03,.10)	(.02,.10)	(.07,.17)	(.02,.17)	(−.17,2.07)	(.03,.07)	(.03,.08)	(.01,.09)	(.00,.14)	(−.03,.11)	(−.07,.18)	(−.10,.17)	(−.01,.08)	(.01,.03)	(.00,.01)	
GNP deflator	MU	1.00	[.97,1.0]	.50			.014	.00	.20	1.0	.50			1.50	1.87	1.96	1.98	1.99	1.99	.960
		(.00,.03)		(−.02,.09)			(.00,.01)	(.00,.10)	(.00,.03)	(.00,.07)	(.00,.10)			(−.04,.10)	(−.15,.28)	(−.26,.40)	(−.41,.62)	(−.47,.74)	(−.51,.85)	[.89,1.0]
(1889–1988,	LS	.968		.47			.092	.11	.19	.94	.50			1.44	1.62	1.51	1.12	.81	.42	.915
$T + p = 100$)		(−.05,.03)		(−.02,.09)			(.00,.02)	(.14,.12)	(.00,.03)	(−.11,.08)	(.05,.13)			(−.07,.10)	(−.33,.27)	(−.61,.36)	(−1.19,.44)	(−1.55,.42)	(−1.85,.35)	

(continued)

Table 4. (continued)

Data series	Estimator	α	90% CI for α	ψ_1	ψ_2	ψ_3	μ	$100 \times \beta$	$100 \times \sigma^2$	Magnitude of roots in descending order: 1	2	3	4	Impulse response function: 1	3	5	10	15	25	NP data: α [90% CI for α]
Consumer	MU	1.00	[1.0,1.0]	.77	−.31	.13	.007	.00	.17	1.0	.60	.46	.46	1.77	2.15	2.30	2.39	2.39	2.40	1.00
prices		(.00,.02)		(−.03,.09)	(.01,.11)	(−.01,.09)	(.00,.01)	(.00,.05)	(.00,.02)	(.00,.06)	(.02,.13)	(.03,.10)	(−.02,.16)	(−.04,.10)	(−.13,.28)	(−.25,.43)	(−.46,.69)	(−.55,.87)	(−61,.103)	[.97,1.0]
(1860–1988,	LS	.987		.71	−.29	.10	.026	.05	.16	.97	.52	.43	.43	1.70	1.90	1.87	1.67	1.47	1.11	.969
$T+p=129$)		(−.03,.02)		(−.03,.09)	(.02,.10)	(.01,.09)	(.00,.02)	(.05,.06)	(.01,.02)	(−.09,.06)	(.09,.14)	(.04,.10)	(.01,.15)	(−.07,.10)	(−.28,.28)	(−.53,.41)	(−1.16,.56)	(−1.61,.59)	(−2.1,.50)	
Nominal	MU	1.00	[.92,1.0]	.54	−.13		.026	.00	.31	1.0	.36	.36		1.54	1.72	1.70	1.70	1.70	1.70	.970
wages		(.00,.04)		(−.02,.11)	(.01,.11)		(.00,.02)	(.00,.20)	(−.01,.05)	(.00,.09)	(.08,.13)	(.00,.03)		(−.05,.12)	(−.19,.30)	(−.30,.42)	(−.41,.60)	(−.46,.69)	(−.48,.76)	[.88,1.0]
(1900–1988,	LS	.939		.53	−.08		.38	.30	.29	.88	.32	.28		1.47	1.47	1.18	.61	.31	.08	.910
$T+p=89$)		(−.07,.05)		(−.03,.10)	(.04,.11)		(.00,.03)	(.30,.22)	(−.02,.05)	(−.13,.09)	(.12,.15)	(−.05,.18)		(−.10,.11)	(−.37,.29)	(−.65,.36)	(−1.18,.40)	(−1.44,.35)	(−1.64,.28)	
Real wages	MU	1.00	[.91,1.0]	.22			.01	.00	.13	1.0	.22			1.22	1.29	1.29	1.29	1.29	1.29	.896
(1900–1988,		(.00,.05)		(.00,.10)			(.00,.01)	(.00,.08)	(.00,.02)	(.00,.08)	(.01,.12)			(−.04,.11)	(−.12,.23)	(−.17,.32)	(−.22,.45)	(−.23,.51)	(−.24,.58)	[.78,1.0]
$T+p=89$)	LS	.929		.27			.22	.12	.12	.90	.30			1.20	1.07	.88	.52	.30	.11	.831
		(−.08,.06)		(.02,.11)			(.00,.02)	(.11,.09)	(−.01,.02)	(−.11,.09)	(.05,.14)			(−.08,.11)	(−.31,.21)	(−.51,.26)	(−.85,.29)	(−1.05,.27)	(−.1.21,.24)	
Money stock	MU	.958	[.92,1.0]	.65			.071	.26	.20	.81	.81			1.61	2.07	1.99	1.15	.50	.06	.942
(1889–1988,		(.00,.03)		(−.01,.08)			(.00,.04)	(−.01,.19)	(.00,.03)	(.05,.08)	(−.05,.12)			(−.01,.08)	(−.04,.28)	(−.03,.46)	(.04,.79)	(.03,.91)	(.03,.91)	[.89,1.0]
$T+p=100$)	LS	.937		.67			.097	.39	.19	.82	.81			1.60	1.98	1.74	.59	.02	−.03	.916
		(−.02,.03)		(.00,.08)			(.02,.04)	(.13,.17)	(.00,.03)	(.02,.05)	(.00,.09)			(−.03,.09)	(−.16,.29)	(−.33,.43)	(−.53,.50)	(−.43,.37)	(−.05,.13)	
Velocity	MU	1.00	[.96,1.0]				−.009	.00	.41	1.0				1.00	1.00	1.00	1.00	1.00	1.00	1.00
(1869–1988,		(.00,.05)					(.00,.02)	(.00,.06)	(−.01,.05)	(.00,.05)				(.00,.05)	(.00,.12)	(.00,.18)	(.00,.28)	(.00,.34)	(.00,.40)	[.93,1.0]
$T+p=120$)	LS	.962					.018	.00	.40	.962				.96	.89	.83	.68	.56	.38	.941
		(−.08,.05)					(.00,.03)	(−.06,.08)	(−.02,.05)	(−.08,.05)				(−.08,.05)	(−.22,.12)	(−.33,.16)	(−.56,.20)	(−.70,.21)	(−.87,.19)	
Interest rate	MU	1.0	[.95,1.0]	.20	−.18		.068	.00	36.3	1.0	.43	.43		1.20	.99	1.02	1.02	1.02	1.02	1.00
(1900–1988		(.00,.07)		(.00,.11)	(.01,.11)		(.00,.20)	(.00,.72)	(−.55,5.97)	(.00,.08)	(.01,.11)	(.00,.13)		(−.04,.12)	(−.11,.22)	(−.12,.26)	(−.16,.35)	(−.17,.40)	(−.19,.45)	[1.0,1.0]
$T+p=89$)	LS	.953		.22	−.16		.078	.50	35.6	.95	.41	.41		1.17	.88	.81	.63	.49	.29	1.03
		(−.10,.07)		(.02,.10)	(.05,.11)		(−.03,.36)	(.65,1.07)	(−1.83,5.69)	(−.11,.09)	(−.01,.12)	(−.03,.14)		(−.09,.12)	(−.28,.02)	(−.43,.22)	(.69,.22)	(−.83,.20)	(−.96,.17)	
Common	MU	1.00	[.91,1.0]	.23	−.19		.03	.00	2.40	1.0	.43	.43		1.23	1.02	1.05	1.05	1.05	1.05	.970
stock prices		(.00,.05)		(.00,.09)	(.01,.10)		(.00,.04)	(.00,.21)	(−.08,.33)	(.00,.06)	(.00,.11)	(.00,.13)		(−.03,.10)	(−.09,.20)	(−.10,.23)	(−.15,.32)	(−.17,.38)	(−.20,.43)	[.88,1.0]
(1871–1988,	LS	.932		.24	−.16		.06	.29	2.30	.93	.41	.41		1.18	.84	.73	.50	.34	.16	.908
$T+p=118$)		(−.07,.05)		(.02,.09)	(.03,.10)		(.00,.08)	(.24,.25)	(−.14,.32)	(−.08,.07)	(−.01,.11)	(−.02,.14)		(−.07,.10)	(−.22,.19)	(−.34,.20)	(−.58,.23)	(−.74,.22)	(−.92,.19)	

NOTE: For the industrial production series, the MU estimates of ψ_4 and ψ_5 are −.07 (.00,.10) and −.23 (.01,.09), the LS estimates of ψ_4 and ψ_5 are −.05 (.03,.09) and −.20 (.04,.09), the MU estimates of the magnitudes of the 5th and 6th largest roots are .76 (−.01,.08) and .68 (−.02,.13), and the LS estimates of the magnitudes of the 5th and 6th largest roots are .75 (−.02,.09) and .67 (−.04,.14).

"Stat" in Stock's table A.1, part B, one finds the row corresponding to the observed value of $\hat{\tau}^{\tau}$. The value c_{med} in (4.1) is the number in the column labeled "Median" that is in the aforementioned row. The values c_0 and c_1 in (4.1) are the numbers in the columns labeled c_0 and c_1 [corresponding to the desired confidence level $100(1-p_0)\%$ being equal to 95%, 90%, 80%, or 70%] that are in the aforementioned row. [If the model (2.2) of interest does not contain a time trend, then one computes c_{med}, c_0, and c_1 from Stock's table A.1, part A, and one omits the time trend in the regressions used to calculate $\hat{\tau}^{\tau}$ and in the regression described later used to calculate $\hat{b}(1)$.]

Our suggested estimator of $b(1)$ is an iterative one. Let $\hat{b}_1(1) = 1 - \sum_{j=1}^{p-1} \hat{\psi}_{j,\text{LS1}}$, where $\hat{\psi}_{j,\text{LS1}}$ is the LS estimator of ψ_j (the coefficient on ΔY_{t-j}) from the regression of Y_t on $(Y_{t-1}, \Delta y_{t-1}, \ldots, \Delta Y_{t-p+1}, 1, t)$. Let $\hat{\alpha}_{\text{AMU1}} = 1 + c_{\text{med}}\hat{b}_1(1)/T$. Let $\hat{b}_2(1) = 1-\sum_{j=1}^{p-1} \hat{\psi}_{j,\text{LS2}}$, where $\hat{\psi}_{j,\text{LS2}}$ is the LS estimator of ψ_j (the coefficient on ΔY_{t-j}) from the regression of $Y_t - \hat{\alpha}_{\text{AMU1}} Y_{t-1}$ on $(\Delta Y_{t-1}, \ldots, \Delta Y_{t-p+1}, 1, t)$. Let $\hat{\alpha}_{\text{AMU2}} = 1 + c_{\text{med}}\hat{b}_2(1)/T$. The estimators $\hat{b}_3(1), \hat{b}_4(1), \ldots$ and $\hat{\alpha}_{\text{AMU3}}, \hat{\alpha}_{\text{AMU4}}, \ldots$ are defined analogously to $\hat{b}_2(1)$ and $\hat{\alpha}_{\text{AMU2}}$. The estimator $\hat{b}(1)$ is then defined to equal either $\hat{b}_k(1)$ for some fixed integer k or the limiting value of $\hat{b}_1(1), \hat{b}_2(1), \ldots$ (provided convergence occurs). In practice, we find that $k = 2$ is sufficient to achieve convergence to within two decimal places for $\hat{b}_k(1)$ and $\hat{\alpha}_{\text{AMU}k}$ for many series, although $k = 6$ is required for one series reported later.

The asymptotic justification for $\hat{\alpha}_{\text{AMU}}$ and $[\hat{L}, \hat{U}]$ is sketched in the Appendix. It is a straightforward extension of Stock's results. Note that the use of $\hat{\alpha}_{\text{AMU}}$ and $[\hat{L}, \hat{U}]$ is appropriate only when the sample size is not "too" small and α is "near" one.

Lastly, we briefly mention a theoretical issue concerning the CI $[\hat{L}, \hat{U}]$ (see Phillips 1991b). The CI $[\hat{L}, \hat{U}]$ for α can be used to obtain a CI for the parameter c, where $\alpha = 1+cb(1)/T$. Because c cannot be estimated consistently, it may seem odd that one can construct a CI for c. In fact, the fact that c cannot be estimated consistently means that the length of the CI for c does not go to 0 (in some probabilistic sense) as $T \to \infty$, but it does not preclude the construction of a CI for c whose coverage probability is correct asymptotically.

5. EMPIRICAL RESULTS FOR THE EXTENDED NELSON–PLOSSER DATA

In this section, we apply the MU estimation method to the ENP data set compiled by Schotman and van Dijk (1991). (All series except the interest rates are logged.) Table 4 presents the results along with LS estimates. In the table, simulated estimates of the biases and standard deviations of the estimators, computed using the MU estimates as the truth, are given in parentheses below each estimate. The lag lengths (p) of the AR(p) models that are estimated are taken to be the same as those of Nelson and Plosser (1982). This choice is made because it facilitates comparison with the results in the literature and because an analysis of the residuals of the estimated models did not provide evidence that the NP lag lengths are inappropriate. (The only exception is some weak evidence that a longer lag length than $p = 1$ may be appropriate for velocity.) Of course, a data-dependent method of choosing p may very well choose different lag lengths.

Eight of the fourteen MU estimates of α equal 1.0. All of the nominal variables have an MU estimate of α equal to 1.0 except money stock, whose estimate is .96. Real wages is the only real variable for which the MU estimate of α is 1.0. The other real variables, except the unemployment rate, have MU estimates of α in the range of .86 to .91. The unemployment rate has the lowest estimate of α; it is .76. The 90% CI's for α for the nominal variables are relatively short with the lower bound ranging from .91 to 1.0. The 90% CI's for α for the real variables are noticeably longer, ranging in length from .18 to .25. The null hypothesis of a unit root ($\alpha = 1$) can be rejected at a 5% level using a one-sided test for three of the series—real GNP, real per capita GNP, and the unemployment rate.

The MU estimates of α are substantially closer to 1.0 than are the LS estimates. The range of differences is .02 to .07. These differences are due to the downward median bias of the LS estimator.

The MU estimates of the time-trend parameter β are fairly small. The LS estimates of β are larger than the MU estimates for every series except the unemployment rate. The bias of the MU estimator of β is essentially 0. In contrast, the LS estimators of β are upward biased by approximately the amount that the LS estimates exceed the MU estimates.

The unbiased model-selection rule introduced in Section 3.3 says to choose a unit-root model if the MU estimator is 1.0 and to choose a trend-stationary model if the MU estimator is less than 1.0. This rule selects eight series as being unit-root models and six as being trend stationary. The unit-root models include all nominal variables except money stock, plus real wages.

Next, we compare the MU and LS estimates for the ENP series with those for the NP series. The last column of Table 4 provides the estimates of α for the NP data. The biggest changes occur with the real-wage series. The LS and MU estimates increase enormously when the new data are added from .83 and .89 to .93 and 1.0, respectively. The 90% CI for α shrinks in length from .22 to .09. The graph of real wages is flat over the period of new data 1971–1988, whereas it increases throughout the period of the NP data 1900–1970. The next largest changes occur for the nominal GNP, GNP deflator, nominal wages, and common-stock price series. The LS and MU estimates for each of these series increased by .03 or .04 with the MU estimates going from .96 or .97 to 1.0 in each case. The lengths of the 90% CI's for α for these series shrink from .14 to .07, .11 to .03, .12 to .08, and .12 to .09. For the first three of these series, the graphs of the series show a steeper slope (presumably due to increased inflation) over the new period of data 1971–1988 than previously.

In addition, the interest-rate series shows a large drop in the LS estimate of α from 1.03 to .95 with addition of the new data, but the MU estimate stays constant at 1.0. The real GNP and real per capita GNP series show little or no change in the LS and MU estimates of α, but the increased precision due to the addition of data allows one to reject the null hypothesis that $\alpha = 1$ with the ENP data, whereas one cannot reject this hypothesis with the NP data.

To analyze the robustness of the preceding results to nonnormality of the innovations, the MU estimates and CI's were recomputed using several alternative distributions for the innovations. The distributions considered were the t_3, chi-squared with 4 df (shifted to have mean 0), Rademacher (± 1 with probability $\frac{1}{2}$ each), and Cauchy. These distributions exhibit thick tails, skewness, discreteness, and extremely thick tails, respectively. In short, the MU estimates and CI's for α are very robust to nonnormality of the innovations. Only for the case of Cauchy innovations did any of the results differ noticeably.

More specifically, the maximum difference between the MU estimates of α for Gaussian, t_3, χ^2_4, and Rademacher distributions is .006 and almost all differences are .003 or less. The MU estimates for the Cauchy differ from those of the normal by .01 or less for all cases but one, in which case the difference was .02. The maximum differences for the CI bounds for α for the normal, t_3, χ^2_4, and Rademacher distributions is .01 or less for all cases but one, in which case the difference is .02. The CI's for α using Cauchy innovations are all contained in those for the normal case. (This is not surprising, because the LS estimator is known to converge at a rate faster than $\sqrt{T}$ in the AR(1) model with Cauchy innovations.) The lower bounds for the Cauchy case are larger by between .00 and .03. The upper bounds for the Cauchy case are smaller by .02 for the cases in which the bounds for Gaussian innovations are less than 1.00 and otherwise are the same as the Gaussian bounds.

6. COMPARISON OF DIFFERENT ESTIMATES USING THE NELSON–PLOSSER DATA

Here we compare the MU and LS estimates for the NP data series with other estimates given in the literature including those of Rudebusch (1992), DeJong and Whiteman (1991a), and Phillips (1991a). In addition, we make comparisons with estimates given by Stock's (1991) AMU estimator of the magnitude of the largest root [although not with results actually reported by Stock (1991)]. We also make comparisons with the AMU of α considered in Section 4. Because different authors use different lag lengths p, comparisons across all methods are not always possible. Rudebusch (1992) used the Nelson and Plosser (1982) choice of p. DeJong and Whiteman (1991a) and Phillips (1991a) used $p = 3$. Moreover, different authors choose to report different estimands. Rudebusch (1992) gave estimates of $\gamma_1, \ldots, \gamma_p$, from which an estimate of $\alpha = \sum_{j=1}^{p} \gamma_j$ can be obtained. DeJong and Whiteman (1991a) reported only estimates of the magnitude of the largest root, Λ, and the time trend parameter β. Phillips (1991a) reported only estimates of α. First we compare the MU estimates with those of Rudebusch for models with the Nelson–Plosser choices of p. Next, we consider all models with $p = 3$ and compare the MU estimates with those of LS, DeJong and Whiteman (1991a) (DW), Phillips (1991a) (Ph), and AMU.

DeJong and Whiteman's (1991a) estimators of Λ and β are Bayesian posterior means in which the prior is chosen to be uniform over the AR coefficients $\gamma_1, \ldots, \gamma_3$ and over the time-trend parameter β subject to the restriction that $\Lambda \in [.55, 1.055]$ and $\beta \in [.000, .016]$. Phillips's (1991a) estimator is defined here to be the posterior median of his posterior distributions for α obtained using the Jefferies prior and some analytic approximations. [Phillips does not report posterior medians. The Ph estimates reported in Table 5 are obtained by eyeballing Phillips's posterior distributions given in his fig. 4. In consequence, these estimates are subject to (our own) computational error.]

Summing Rudebusch's (1992) estimates of $\gamma_1, \ldots, \gamma_p$ yields the following estimates of α. We give the Rudebusch estimate first and the MU estimates second for the series as ordered in Table 4 (with the NP choice of p): (.898, .885), (.946, .958), (.882, .875), (.919, .919), (.900, .914), (.773, .765), (.968, .960), (.985, 1.00), (.974, .970), (.913, .896), (.947, .942), (.995, 1.00), (.984, 1.00), and (984, .970). (Rudebusch did not provide any measure of the variability of his estimates, so none can be given here.) Overall, the differences are small. They vary from .000 for industrial production to .017 for real wages. Thus the Rudebusch estimates are much closer to the MU estimates than to the LS estimates. This is to be expected because the MU and Rudebusch methods are quite similar.

Next, we turn to comparisons of MU, LS, DW, Ph, and AMU estimates for AR(3) models (see Table 5). Results are reported for estimates of α, β, and the magnitude of the largest root Λ. Bias and standard-deviation estimates are provided in parentheses beside each of the MU and LS estimates. These were obtained by the simulation method outlined in Section 3.1. The standard deviation of the posterior distribution of Λ is provided in parentheses beside the DW estimates. Asymptotic 90% central CI's for α and Λ are provided in brackets beside each estimate for the AMU estimator. The CI for Λ is as defined by Stock (1991); that for α is as defined in Section 4.

First, we summarize the results for the main parameter of interest α. The comparison between the MU and LS estimates is quite similar to that given in Table 4. The MU estimates are uniformly closer to 1 than the LS estimates. The differences between the two estimates range from .02 to .09. These differences correspond to MU estimates of the CIR that are from 38% to ∞% larger than those of the LS estimates.

The Ph estimates are slightly larger (i.e., larger by .01 or .02) than the LS estimates for all series except the industrial-production, velocity, and interest-rate series. For the latter two, the Ph estimates are much larger. The latter two are the series with the largest LS estimates. For these series, the Ph estimates are much larger than the LS estimates because of the large weight that the Jefferies prior puts on $\alpha > 1$. Because the Ph estimates of α are just slightly larger than the LS estimates for most series, the MU estimates are noticeably larger than the Ph estimates for most series.

Table 5. A Comparison of Different Estimates for the Nelson–Plosser Data Series Using AR(3) Models

Data series	Estimator	α	90% CI for α	$100 \times \beta$	Magnitude of largest root	Data series	Estimator	α	90% CI for α	$100 \times \beta$	Magnitude of largest root
Real GNP	MU	.87 (.00,.09)	[.75,1.0]	.45 (.00,.31)	.69 (.12,.12)	Consumer	MU	1.0 (.00,.03)	[.97,1.0]	.00 (.00,.05)	1.0 (.00,.07)
(1909–1970,	LS	.81 (−.06,.08)		.61 (.21,.27)	.70 (.07,.09)	prices	LS	.98 (−.05,.03)		.05(.05,.06)	.95 (−.10,.08)
$T+p=62$)	DW			.6 (.2)	.76 (.09)	(1860–1970,	DW			.0 (.0)	.95 (.04)
	Ph	~.82				$T+p=111$)	Ph	~.98			
	AMU	.88	[.76,1.03]		.81 [.61,1.04]		AMU	1.01	[.97,1.02]		1.02[.95,1.04]
Nominal GNP	MU	1.0 (.00,.07)	[.88,1.0]	.00(.00,.45)	.96 (.00,.11)	Nominal	MU	.97 (.01,.06)	[.88,1.0]	.14 (−.06, .28)	.95 (.02,.10)
(1909–1970,	LS	.91 (−.11,.07)		.52 (.59,.52)	.78 (−.19,.11)	wages	LS	.91 (−.06,.06)		.38 (.28,.28)	.79 (−.13,.09)
$T+p=62$)	DW			.5 (.2)	.82 (.09)	(1900–1970,	DW			.4 (.1)	.82 (.09)
	Ph	~.92				$T+p=71$)	Ph	~.93			
	AMU	~1.02	[.87,1.04]		1.03 [.81,1.04]		AMU	.99	[.88,1.03]		.98 [.80,1.05]
Real per	MU	.86 (.00,.09)	[.74,1.0]	.28(.00,.19)	.65 (.09,.12)	Real	MU	.89(.01,.09)	[.76,1.0]	.24 (−.02,.21)	.85 (.01,.13)
capita GNP	LS	.80 (−.07,.08)		.38 (.13,.17)	.70 (.04,.09)	wages	LS	.82 (−.07,.08)		.38 (.15,.18)	.71 (−.11,.10)
(1909–1970,	DW			.4 (.1)	.75 (.09)	(1900–1970,	DW			.4 (.1)	.76 (.09)
$T+p=62$)	Ph	~.82				$T+p=71$)	Ph	~.83			
	AMU	.87	[.75,1.03]		.79 [.60,1.04]		AMU	.87	[.73,1.03]		.83 [.66,1.04]
Industrial	MU	.87 (.00,.07)	[.76,1.0]	.53 (.01,.30)	.86 (.00,.10)	Money	MU	.95 (.00,.04)	[.89,1.0]	.29 (−.01,.22)	.79 (.07,.09)
production	LS	.82 (−.05,.07)		.75 (.22,.28)	.80 (−.07,.10)	stock	LS	.92 (−.03,.04)		.47(.17,.20)	.80 (.02,.06)
(1860–1970,	DW			.7 (.2)	.78 (.09)	(1889–1970,	DW			.5 (.1)	.83 (.07)
$T+p=111$)	Ph	~.86				$T+p=82$)	Ph	~.93			
	AMU	.87	[.76,1.02]		.86 [.75,1.02]		AMU	.95	[.90,1.01]		.88 [.74,1.04]
Employment	MU	.91 (.00,.07)	[.81,1.0]	.13(.00,.11)	.88 (−.01,.12)	Velocity	MU	1.0(.00,.05)	[.93,1.0]	.00 (.00,.07)	1.0 (.00,.07)
(1890–1970,	LS	.86 (−.06,.06)		.21 (.09,.10)	.76 (−.12,.11)	1869–1970,	LS	.94 (−.09,.06)		−.04 (−.08,.10)	.94 (−.10,.08)
$T+p=81$)	DW			.2 (.0)	.78 (.10)	$T+p=102$)	DW			.0 (.0)	.99 (.02)
	Ph	~.88					Ph	~1.01			
	AMU	.93	[.82,1.03]		.90 [.76,1.04]		AMU	1.02	[.95,1.04]		1.02 [.95,1.04]
Unemployment	MU	.81 (.00,.11)	[.65,.97]	−.17(.03,.27)	.82 (−.01,.13)	Interest	MU	1.0(.00,.05)	[1.0,1.0]	.00 (.00,.55)	1.0 (.00,.08)
rate*	LS	.73 (−.08,.10)		−.20 (−.05,.33)	.72(−.10,.12)	rate	LS	1.03 (−.07,.05)		.32(.53,.86)	1.05 (−.14,.06)
(1890–1970,	DW			.1 (.1)	.75 (.10)	(1900–1970;	DW			.3 (.1)	.98 (.06)
$T+p=81$)	Ph	~.75				$T+p=71$)	Ph	~1.15			
	AMU	.82	[.67,1.02]		.82 [.67,1.03]		AMU	1.03	[1.02,1.04]		1.05 [1.03,1.08]
GNP deflator	MU	.95 (.00,.05)	[.88,1.0]	.13 (.00,.13)	.89 (.00,.10)	Common-	MU	.97 (.00,.06)	[.88,1.0]	.15 (− .02,.31)	.97 (.00,.08)
	LS	.91 (−.05,.05)		.22 (.12,.13)	.75(−.10,.07)	stock	LS	.91 (−.07,.06)		.32 (.34,.32)	.89 (−.08,.09)
(1889–1970,	DW			.2 (.0)	.82 (.07)	prices	DW			.3 (.1)	.88 (.07)
$T+p=82$)	Ph	~.92				(1871–1970,	Ph	~.93			
	AMU	.96	[.88,1.02]		.92 [.77,1.04]	$T+p=100$)	AMU	1.01	[.88,1.04]		1.01 [.87,1.04]

NOTE: The Philips estimate for this series is based on an AR (4) model.

The MU and AMU estimates of α are quite similar. The differences are between .00 and .02. The differences in the lower bounds of the MU and AMU CI's for α also are fairly small in most cases, although they differ by .03 for real wages.

Next, we compare estimates of the magnitude of the largest root Λ. Although Λ is not a parameter of great interest by itself, as argued in Section 2, these comparisons indicate whether the differences between the MU estimates and other estimates in the literature, such as those of DeJong and Whiteman (1991a), are due to differences in the methods employed or just to the choice of estimand considered. The differences between the MU and DW estimates of Λ are very large. They range from $-.10$ to .14, with most being in the .07 to .10 range. The MU estimates are usually significantly closer to 1 than the DW estimates, but not always. For many cases the bias of the MU estimator is small, though for a few cases it is large. In each case in which it is large, the DW estimate is in the direction of the bias relative to the MU estimate, which suggests that the DW estimator is more biased than the MU estimator. The LS and DW estimates of Λ are closer together than the MU and DW estimates are, but there still are noticeable differences. Unlike the estimates of α, the MU and AMU estimates of Λ differ noticeably for a few series.

Last, we compare estimates of the time-trend parameter β. The LS and DW estimates of β are almost the same. The MU estimates are noticeably closer to 0 than the LS and DW estimates. The difference between the MU and LS estimates of β are approximately the same as the upward bias of the LS estimator. The MU estimator of β is essentially unbiased. One might conjecture that the DW estimates of β have an upward bias roughly equal to that of the LS estimates. Bias correction of the LS and DW estimates, then, would yield estimates approximately equal to the MU estimates.

Overall, the results of Table 5 lead to the following conclusions. There are noticeable differences between the MU and AMU estimates on one hand and the LS, DW, and Ph estimates on the other. The former show considerably greater persistence for most of the series than the latter. The differences can be attributed to the fact that the MU and AMU estimators of α and Λ are not biased toward 0 and those of β are not biased away from 0.

7. EMPIRICAL RESULTS FOR THE STOCK–DIVIDEND AND PRICE DATA

In this section, we present empirical results for the stock-dividend and price data referred to in Section 1. (All series are logged.) We use an AR(3) model for each series, as in the work of DeJong and Whiteman (1991b). This choice is made for comparative purposes and because residual analysis did not indicate that this choice is inappropriate.

Table 6 presents MU and LS estimates of a variety of estimands for the stock-market data series. In addition, the DW posterior mean estimates of the magnitude Λ of the largest root and the coefficient β on the time trend are provided. Bias and standard-deviation estimates for the MU and LS estimators (computed using the simulation method outlined in Sec. 3.1 and taking the MU estimates as the truth) are given in parentheses below each estimate. The standard deviations of the posterior distributions of Λ and β are given in parentheses below the DW estimates of these parameters.

We now discuss the results of Table 6. Four of the six series show considerable persistence; two show noticeably less persistence. In particular, DJ prices, NYSE dividends and prices, and S&P prices all have MU estimates of α equal to .90 or greater, whereas DJ dividends and S&P dividends have MU estimates of α equal to .79 and .82, respectively. Only NYSE prices have an MU estimate of α equal to 1.0. Thus the unbiased model-selection procedure of Section 3.3 chooses a unit-root model for NYSE prices and trend-stationary models for all other series.

The 90% central CI's for α are extremely wide for the DJ dividend and price and S&P dividend series with widths of .45, .31, and .28, respectively. The CI's for α for the NYSE price and S&P price and dividend series are also wide, but much less so, with widths of .19, .21, and .16. The principal explanation for the excessively wide CI's is the small number of observations ($T+p$) for the DJ and NYSE series—51 and 55, respectively.

The LS estimates of α and of the IRF are much smaller than the MU estimates, especially for the DJ and NYSE series. The differences in LS and MU estimates of α for these series range from .10 to .21, which are very large. In all cases, the LS estimates are closer to 0 than the MU estimates. This is due to the downward bias of the LS estimators, which is particularly large for small-sample sizes. Given these biases, we do not believe that the LS estimates give impartial estimates of the amount of persistence in the series, as measured by α or by the IRF. The MU estimates of α and the IRF, on the other hand, are essentially median unbiased in most cases. Hence they provide a more objective estimate of the amount of persistence.

The MU estimates of α and of the magnitude Λ of the largest root are approximately the same for each series except S&P dividends. The same is true of the LS estimates. In consequence, for five of these series, the magnitude of the largest root can be given an interpretation related to the persistence of the series.

Comparing the DW estimates of Λ with those of LS, we find the DW and LS estimates are approximately equal for all series except DJ dividends and S&P dividends. Comparing the DW estimates of Λ with the MU estimates, we find that the DW estimates are uniformly smaller than the MU estimates. The differences for the six series are .07, .16, .13, .16, .05, and .07, which are substantial. Thus the MU estimates indicate considerably greater persistence in the series than the DW estimates do. The explanation for the differences is the difference in the bias properties of the MU and DW estimators.

We conclude that the MU estimates differ noticeably from the LS and DW estimates. Of the point estimates given, we believe that the MU estimates of α and the IRF to be the most informative regarding persistence because they are approximately median unbiased. The interval estimates for α also are quite informative because they make clear that the

Table 6. Median-Unbiased (MU), Least Squares (LS), and Dejong–Whiteman (DW) Estimates for Stock Dividend and Price Series

Data series	Estimator	α	90% CI for α	ψ_1	ψ_2	μ	$100 \times \beta$	$100 \times \sigma^2$	Magnitude of roots in descending order: 1	2	3	Impulse response function: 1	2	3	5	10	15	25
DJ dividends	MU	.786	[.55,1.0]	.30	−.25	.42	.41	1.8	.79	.56	.56	1.09	.63	.34	.29	.08	.03	.00
(1928–1978,		(−.01,.17)		(.00,.14)	(.00,.16)	(.00,.31)	(−.01,.37)	(−.01,.41)	(−02,.14)	(.04,.12)	(−.01,.18)	(−.02,.16)	(−.01,.25)	(.02,.29)	(.00,.31)	(−.01,.29)	(−.01,.29)	(.00,.28)
$T + p = 51$)	LS	.675		.34	−.18	.65	.61	1.8	.64	.52	.52	1.02	.52	.18	.09	.01	.00	.00
		(−.13,.14)		(.05,.13)	(.08,.14)	(.22,.27)	(.24,.37)	(−.06,.39)	(−.11,.11)	(.04,.12)	(−.07,.20)	(−.09,.15)	(−.15,.24)	(−.20,.24)	(−.19,.18)	(−.07,.06	(−.02,.03)	(.00,.01)
	DW						.6		.72									
							(.03)		(.10)									
DJ prices	MU	.912	[.69,1.0]	.00	−.10	.44	.22	5.3	.92	.33	.33	.91	.73	.68	.58	.39	.26	.11
(1928–1978,		(.00,.15)		(.01,.15)	(.00,.14)	(−.02,.69)	(−.15,.49)	(−.04,1.2)	(.00,.14)	(.07,.15)	(.03,.16)	(.03,.16)	(.05,.23)	(−.02,.26)	(.01,.32)	(.02,.37)	(.01,.39)	(.01,.41)
$T + p = 51$)	LS	.767		.07	−.03	1.15	.56	5.1	.76	.19	.19	.84	.60	.45	.26	.06	.02	.00
		(−.16,.14)		(.08,.14)	(.06,.14)	(.74,.66)	(.38,.64)	(−.29,1.1)	(−.18,.14)	(.10,.16)	(.00,.15)	(−.11,.16)	(−.17,.22)	(−.27,.22)	(−.35,.20)	(−.34,.14)	(−.25,.12)	(−.11,.14)
	DW						.6		.76									
							(.03)		(.11)									
NYSE dividends	MU	.900	[.72,1.0]	.30	−.24	1.52	.39	1.4	.90	.51	.51	1.20	.91	.68	.58	.34	.19	.06
(1926–1981,		(−.01,.13)		(.01,.13)	(.00,.14)	(.13,1.93)	(.03,.53)	(.02,.32)	(−.01,.14)	(.04,.12)	(.01,.17)	(−.02,.15)	(−.05,.25)	(−.06,.31)	(−.02,.36)	(−.04,.40)	(−.03,.41)	(−.01,.43)
$T + p = 55$)	LS	.795		.34	−.17	3.08	.79	1.4	.76	.47	.47	1.13	.77	.47	.27	.07	.02	.00
		(−.13,.12)		(.04,.12)	(.07,.13)	(1.94,1.73)	(.50,.51)	(−.06,.30)	(−.15,.12)	(.04,.02)	(−.04,.19)	(−.10,.15)	(−.19,.24)	(−.28,.26)	(−.34,.24)	(−.29,.13)	(−.18,.09)	(−.06,.07)
	DW						.8		.77									
							(.03)		(.10)									
NYSE prices	MU	1.00	[.81,1.0]	.06	−.22	.04	.00	4.7	1.0	.47	.47	1.06	.85	.82	.87	.86	.86	.86
(1926–1981,		(−.02,.13)		(.03,.14)	(.02,.14)	(−.01,.10)	(.00,.53)	(−.12,.98)	(−.01,.13)	(.02,.14)	(.00,.02)	(−.06,.15)	(−.11,.24)	(−.16,.26)	(−.15,.31)	(−.19,.37)	(−.23,.39)	(−.24,.42)
$T + p = 55$)	LS	.851		.14	−.14	2.68	.68	4.5	.85	.40	.40	.99	.71	.56	.43	.20	.09	.02
		(−.20,.13)		(.08,.13)	(.07,.14)	(−.03,.17)	(.56,.75)	(−.35,.90)	(−.19.13)	(.01,.14)	(−.03,.17)	(−.14,.15)	(−.26,.23)	(−.39,.23)	(−.54,.23)	(−.76,.20)	(−.83,.22)	(−.86,.56)
	DW						.7		.84									
							(.03)		(.10)									
S & P dividends	MU	.820	[.71,.92]	.25	−.07	−.05	.21	1.4	.77	.31	.31	1.07	.82	.61	.36	.10	.03	.00
(1871–1985,		(.00,.07)		(.00,.08)	(−.01,.10)	(.00,.04)	(.00,.09)	(.00,.19)	(−.00,−.01)	(.09,.13)	(.02,.15)	(−.01,.10)	(−.02,.14)	(−.02,.16)	(.01,.18)	(.00,.14)	(.00,.09)	(.00,.06)
$T + p = 115$)	LS	.777		.27	−.05	−.07	.26	1.4	.68	.27	.27	1.05	.78	.53	.24	.04	.01	.00
		(−.04,.07)		(.02,.09)	(.02,.09)	(−.01,.04)	(.05,.09)	(−.01,.19)	(−.07,.11)	(.12,.14)	(−.02,.15)	(−.04,.10)	(−.06,.14)	(−.09,.15)	(−.10,.15)	(.01,.08)	(−.02,.03)	(.00,.01)
	DW						.3		.72									
							(.01)		(.09)									
S &P prices	MU	.936	[.84,1.0]	.10	−.14	.19	.07	3.1	.94	.39	.39	1.04	.84	.76	.69	.51	.37	.20
(1871–1985,		(−.01,07)		(.00,.10)	(.00,.10)	(.00,.19)	(−.01,.11)	(.00,.44)	(.00,.08)	(.01,.12)	(.00,.13)	(−.02,.10)	(−.02,.15)	(−.03,.17)	(−.02,.21)	(−.03,.29)	(−.03,.32)	(−.02,.35)
$T + p = 115$)	LS	.881		.13	−.11	.33	.15	3.0	.88	.36	.36	1.01	.78	.65	.52	.28	.15	.04
		(−.07,.06)		(.03,.09)	(.03,.09)	(.17,.18)	(07,.13)	(−.06,.43)	(−.07,.09)	(−.01,.12)	(−0.3,.14)	(−.05,.10)	(−.09,.15)	(−.14,. 15)	(−.02,.16)	(−.26,.15)	(−.25,.13)	(−.17,.08)
	DW						.2		.87									
							(.01)		(.07)									

level of uncertainty about the "true" values of α is quite high.

The MU estimates of α indicate a high degree of persistence for four of the six series and a lesser degree for two series. One of the six series is estimated to have a unit root and five are estimated to be trend stationary.

ACKNOWLEDGMENT

We thank Charles Nelson, Herman van Dijk, and David DeJong for making available to us, respectively, the Nelson–Plosser, extended Nelson–Plosser, and stock-dividend and price series referred to in the text. We thank John McDermott for his help with some of the estimation results. We also thank David DeJong, Peter Phillips, Glenn Rudebusch, Chris Sims, Jim Stock, a referee, an associate editor, and the editor for helpful comments. Andrews acknowledges financial support from the National Science Foundation via Grants SES-8821021 and SES-9121914.

APPENDIX: PROOFS

A.1 Invariance of $\hat{\alpha}_{\mathrm{LS}}$

First we establish the claim made in Section 3.1 that $\hat{\alpha}_{\mathrm{LS}}$ has distribution that does not depend on $(\mu^*, \beta^*, \sigma^2)$ and when $\alpha = 1$ on Y^*_{-p+1}. In fact, we will show that these invariance properties hold for the distribution of $(\hat{\alpha}_{\mathrm{LS}}, \hat{\psi}_{1,\mathrm{LS}}, \ldots, \hat{\psi}_{p-1,\mathrm{LS}}$.

Consider successive regressions of Y_t, Y_{t-1}, $\Delta Y_{t-1}, \ldots, \Delta Y_{t-p+1}$ on $(1, t)$ for $t = 1, \ldots, T$. Then $(\hat{\alpha}_{\mathrm{LS}}, \hat{\psi}_{1,\mathrm{LS}}, \ldots, \hat{\psi}_{p-1,\mathrm{LS}})$ equals the LS estimator from the regression of the residuals from the regression with Y_t as dependent variable on the vector of residuals from the regressions with $Y_{t-1}, \Delta Y_{t-1}, \ldots, \Delta Y_{t-p+1}$ as dependent variables. Because $Y_t = \mu^* + \beta^* t + Y_t^*$ by (2.1), all of the preceding residuals are invariant with respect to (μ^*, β^*). In consequence, the distribution of $(\hat{\alpha}_{\mathrm{LS}}, \ldots, \hat{\psi}_{p-1,\mathrm{LS}})$ is invariant with respect to (μ^*, β^*).

Given the invariance, we can suppose that $\mu^* = \beta^* = 0$ and $Y_t = Y_t^*$ in the remainder of the proof. Multiplication of σ^2 by a positive constant c in (2.1) causes Y_t^* and Y_t to be multiplied by the same constant c for $t = -p + 1, \ldots, T$ when $\alpha \in (-1, 1)$ (using the fact that stationarity of $\{Y_t : t \geq -p + 1\}$ requires that the initial rv's $Y^*_{-p+1}, \ldots, Y^*_0$ are scaled by the same constant c). In consequence, the residuals from the regressions of $Y_t, \ldots, \Delta Y_{t-p+1}$ on $(1, t)$ are multiplied by the same constant. This constant cancels out in the expression for the LS estimator $(\hat{\alpha}_{\mathrm{LS}}, \ldots, \hat{\psi}_{p-1,\mathrm{LS}})$ given by the regression of the residuals from Y_t on those from $(Y_{t-1}, \ldots, \Delta Y_{t-p+1})$. Thus the distribution of $\hat{\alpha}_{\mathrm{LS}}$ is invariant with respect to σ^2 when $\alpha \in (-1, 1)$.

Now suppose that $\alpha = 1$. We can always write $Y_t^* = Y^*_{-p+1} + \sum_{s=-p+2}^{t} \Delta Y_s^*$. By assumption, when $\alpha = 1$, $\{\Delta Y_t^* : t \geq -p + 2\}$ is stationary with level that is arbitrary. That is, a change in Y^*_{-p+1} has no effect on $\{\Delta Y_t^* : t \geq -p + 2\}$. In consequence, because $Y_t = Y_t^*$, the residuals from the regressions of $Y_t, \ldots, \Delta Y_{t-p+1}$ on $(1, t)$ are invariant with respect to the value of Y^*_{-p+1}, and $(\hat{\alpha}_{\mathrm{LS}}, \ldots, \hat{\psi}_{p-1,\mathrm{LS}})$ is likewise. Given this invariance, suppose that $Y^*_{-p+1} = 0$. Then, the multiplication of σ^2 by a constant c causes $\Delta Y_t^*, Y_t^*, \Delta Y_t$, and Y_t to be scaled by the same constant. As previously, this leaves $(\hat{\alpha}_{\mathrm{LS}}, \ldots, \hat{\psi}_{p-1,\mathrm{LS}})$ unchanged. The proof is now complete.

A.2 Asymptotic Properties of $\hat{\alpha}_{\mathrm{AMU}}$ and $[\hat{\mathrm{L}}, \hat{\mathrm{U}}]$

Next we consider the asymptotic justification for $\hat{\alpha}_{\mathrm{AMU}}$ and $[\hat{L}, \hat{U}]$. We use the same model and assumptions as Stock (1991). The parameter α in his notation is $\alpha(1) = 1 + cb(1)/T$, where c is a constant and $b(1) = 1 - \sum_{j=1}^{p-1} \psi_j$. Equation (5) of Stock (1991) gives the local-to-unity asymptotic distribution of the statistic $\hat{\tau}^\tau$. This distribution depends only on c. Let $f_{\ell,p}(c)$ and $f_{u,p}(c)$ denote the lower and upper p quantiles of this distribution.

Consider the following CI for α:

$$\hat{\mathrm{CI}} = \{\alpha : \alpha = 1 + \tilde{c}\,\hat{b}(1)/T \text{ and } f_{\ell,p_\ell}(\tilde{c}) \leq f_{u,p_u}(\tilde{c})\}. \quad \text{(A.1)}$$

This CI has asymptotic confidence level $100(1 - p_\ell - p_u)\%$:

$$\begin{aligned} P_{\alpha_T}(a_T \in \hat{\mathrm{CI}}) &= P_{\alpha_T}[f_{\ell,p_\ell}(\tilde{c}) \leq \hat{\tau}^\tau \leq f_{u,p_u}(\tilde{c}) \\ &\quad \text{for } \tilde{c} \text{ defined by } a_T = 1 + \tilde{c}\,\hat{b}(1)/T] \\ &= P_{\alpha_T}[f_{\ell,p_\ell}(cb(1)/\hat{b}(1)) \leq \hat{\tau}^\tau \leq f_{u,p_u}(cb(1)/\hat{b}(1))] \\ &\to 1 - p_\ell - p_u \text{ as } T \to \infty, \end{aligned} \quad \text{(A.2)}$$

where $\alpha_T = 1 + cb(1)/T$ and $P_{\alpha_T}(\cdot)$ denotes the probability measure when α_T is the true value of α. The preceding convergence to $1 - p_\ell - p_u$ uses the fact that $f_{\ell,p}(c)$ and $f_{u,p}(c)$ are continuous functions of c and $\hat{\tau}^\tau$ has absolutely continuous limit distribution.

Let $f_{\ell,p}^{-1}(y) = \sup\{c : f_{\ell,p}(c) \leq y\}$ and $f_{u,p}^{-1}(y) = \inf\{c : f_{u,p}(c) \geq y\}$. If $f_{\ell,p_\ell}(c)$ and $f_{u,p_u}(c)$ are monotone increasing functions of c, then $f_{\ell,p_\ell}(\tilde{c}) \leq \hat{\tau}^\tau \leq f_{u,p_u}(\tilde{c})$ iff $1 + f_{u,p_u}^{-1}(\hat{\tau}^\tau)\hat{b}(1)/T \leq 1 + \tilde{c}\,\hat{b}(1)/T \leq 1 + f_{\ell,p_\ell}(\hat{\tau}^\tau)\hat{b}(1)/T$. In this case, $\hat{\mathrm{CI}} = [\hat{L}, \hat{U}]$ with $c_0 = f_{u,p_u}^{-1}(\hat{\tau}^\tau)$ and $c_1 = f_{\ell,p_\ell}^{-1}(\hat{\tau}^\tau)$ and $[\hat{L}, \hat{U}]$ is an asymptotically valid $100(1 - p_\ell - p_u)\%$ CI for α. If $f_{\ell,p_\ell}(c)$ and $f_{u,p_u}(c)$ is not everywhere monotone increasing in c, then $\hat{\mathrm{CI}} \subset [\hat{L}, \hat{U}]$ and $[\hat{L}, \hat{U}]$ is an asymptotically valid CI for α with confidence level $\geq 100(1 - p_\ell - p_u)\%$. In fact, $f_{\ell,p_\ell}(c)$ and $f_{u,p_u}(c)$ are almost, but not quite, monotone in c (see Stock 1991, fig. 2) In consequence, $[\hat{L}, \hat{U}]$ has asymptotic significance level just slightly above $100(1 - p_\ell - p_u)\%$. (To obtain a CI with asymptotic confidence level *exactly* $100(1 - p_\ell - p_u)\%$, if this precision is deemed necessary for some reason, one can use $\hat{\mathrm{CI}}$ defined previously in conjunction with Stock's fig. 2).

Furthermore, if $f_{l,1/2}(c)$ is monotone increasing in c, then for (p_l, p_u) equal to $(0, \frac{1}{2})$ and $(\frac{1}{2}, 1)$ the two corresponding $\hat{\mathrm{CI}}$ CI's are of the form $[\hat{\alpha}_{\mathrm{AMU}}, \infty)$ and $(-\infty, \hat{\alpha}_{\mathrm{AMU}}]$, respectively. These CI's have the property that their probabilities of covering the true α are both 1/2 asymptotically. In consequence, $\hat{\alpha}_{\mathrm{AMU}}$ is asymptotically median unbiased. In fact, $f_{l,1/2}(c)$ is not quite monotone increasing in c (see Stock 1991, fig. 2). The extent of nonmonotonicity is sufficiently small that $\hat{\alpha}_{\mathrm{AMU}}$ is very close to being asymptotically median unbiased (close enough for practical purposes), although it is not exactly so. Furthermore, the small region where nonmonotonicity occurs is just above $\alpha = 1$, so if one restricts the parameter space to be $(-1, 1]$, then this problem disappears.

[Received September 1992. Revised August 1993.]

REFERENCES

Andrews, D. W. K. (1993), "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models," *Econometrica*, 61, 139–165.

Andrews, D. W. K., and Phillips, P. C. B. (1987), "Best Median-Unbiased Estimation in Linear Regression With Bounded Asymmetric Loss Functions," *Journal of the American Statistical Association*, 82, 886–893.

Basawa, I. V., Mallik, A. K., McCormick, W. P, Reeves, J. H., and Taylor, R. L. (1991), "Bootstrapping Unstable First-Order Autoregressive Processes, *The Annals of Statistics*, 19, 1098–1101.

Bobkoski, M. J. (1983), "Hypothesis Testing in Nonstationary Times Series," unpublished Ph.D. dissertation, University of Wisconsin, Dept. of Statistics.

Cavanagh, C. L. (1985), "Roots Local to Unity," unpublished manuscript, Harvard University, Dept. of Economics.

Chan, N. H. (1988), "The Parameter Inference for Nearly Nonstationary Time Series," *Journal of the American Statistical Association*, 83, 857–862.

Chan, N. H., and Wei, C. Z. (1987), "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *The Annals of Statistics*, 15, 1050–1063.

DeJong, D., Nankervis, J. C., Savin, N. E., and Whiteman, C. H. (1992), "The Power Problems of Unit Root Tests in Time Series With Autoregressive Errors," *Journal of Econometrics*, 53, 323–343.

DeJong, D., and Whiteman, C. H. (1991a), "Reconsidering 'Trends and Random Walks in Macroeconomic Times Series,'" *Journal of Monetary Economics*, 28, 221–254.

——— (1991b), "The Temporal Stability of Dividends and Stock Prices: Evidence From the Likelihood Function," *American Economic Review*, 81, 600–617.

Fair, R. C. (1992), "Estimates of the Bias of Lagged Dependent Variable Coefficient Estimates in Macroeconomic Equations," Discussion Paper 1005, Yale University, Cowles Foundation.

Gallant, A. R., Rossi, P. E., and Tauchen, G. (1993), "Nonlinear Dynamic Structures," *Econometrica*, 61, 871–907.

Hurwicz, L. (1950), "Least-Squares Bias in Time Series," in *Statistical Inference in Dynamic Economic Models*, ed. T. C. Koopmans, New York: John Wiley, pp. 365–383.

Kendall, M. G. (1954), "Note on Bias in the Estimation of Autocorrelation," *Biometrika*, 41, 403–404.

Lehmann, E. L. (1959), *Testing Statistical Hypotheses*, New York: John Wiley.

Marriott, F. H. C., and Pope, J. A. (1954), "Bias in the Estimation of Autocorrelations," *Biometrika*, 41, 390–402.

Nelson, C. R., and Plosser, C. I. (1982), "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 10, 129–162.

Orcutt, G. H., and Winokur, H. S., Jr. (1969), "First Order Autoregression: Inference, Estimation , and Prediction," *Econometrica*, 37, 1–14.

Phillips, P. C. B. (1987), "Toward a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.

——— (1988), "Regression Theory for Near-Integrated Times Series," *Econometrica*, 56, 1021–1043.

——— (1991a), "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends," *Journal of Applied Econometrics*, 66, 333–364.

——— (1991b), "Bayesian Routes and Unit Roots: De Rebus Prioribus Semper Est Disputandum," *Journal of Applied Econometrics*, 6, 436–473.

Quenouille, M. H. (1949), "Approximate Tests of Correlation in Time-Series," *Journal of the Royal Statistical Society*, Ser. B, 11, 68–84.

——— (1956), "Notes on Bias in Estimation," *Biometrika*, 43, 353–360.

Rudebusch, G. D. (1992), "Trends and Random Walks in Macroeconomic Time Series: A Reexamination," *International Economic Review*, 33, 661–680.

Schotman, P., and van Dijk, H. K. (1991), "On Bayesian Routes to Unit Roots," *Journal of Applied Econometrics*, 6, 387–401.

Shiller, R. J. (1981), "Do Stock Prices Move Too Much To Be Justified by Subsequent Changes in Dividends?" *American Economic Review*, 71, 421–436.

Stine, R. A., and Shaman, P. (1989), "A Fixed Point Characterization for Bias of Autoregressive Estimators," *The Annals of Statistics*, 17, 1275–1284.

Stock, J. H. (1991), "Confidence Intervals for the Largest Autoregressive Root in U. S. Macroeconomic Time Series," *Journal of Monetary Economics*, 28, 435–459.

Zivot, E., and Phillips, P. C. B. (1991), "A Bayesian Analysis of Trend Determination in Economic Time Series," Discussion Paper 1002, Yale University, Cowles Foundation.