Trends versus Random Walks in Time Series Analysis

Steven N. Durlauf, Peter C. B. Phillips

Econometrica, Volume 56, Issue 6 (Nov., 1988), 1333-1354.

Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198811%2956%3A6%3C1333%3ATVRWIT%3E2.0.CO%3B2-5

Econometrica, Vol. 56, No. 6 (November, 1988), 1333–1354

TRENDS VERSUS RANDOM WALKS IN TIME SERIES ANALYSIS

BY STEVEN N. DURLAUF AND PETER C. B. PHILLIPS[1]

This paper studies the effects of spurious detrending in regression. The asymptotic behavior of traditional least squares estimators and tests is examined in the context of models where the generating mechanism is systematically misspecified by the presence of deterministic time trends. Most previous work on the subject has relied upon Monte Carlo studies to understand the issues involved in detrending data that are generated by integrated processes and our analytical results help to shed light on many of the simulation findings. Standard F tests and Hausman tests are shown to inadequately discriminate between the competing hypotheses. Durbin-Watson statistics, on the other hand, are shown to be valuable measures of series stationarity. The asymptotic properties of regressions and excess volatility tests with detrended integrated time series are also explored.

KEYWORDS: Excess volatility tests, integrated processes, misspecification, specification tests, spurious detrending.

1. INTRODUCTION

TRADITIONAL ANALYSES of economic time series frequently rely on the assumption that the time series in question are stationary, ergodic processes. Stationarity and ergodicity together with a few other technical conditions ensure that the first and second sample moments of such series satisfy a strong law of large numbers (SLLN), and that suitably standardized sums of elements of the series obey a central limit theorem (CLT). However, the assumptions of the traditional theory do not provide much solace to the empirical worker. Even casual examination of such time series as GNP reveals that these series do not possess constant means. Similarly, the embedding of such disparate economic events as the great depression and OPEC price shocks in a single data realization renders the stationarity assumption dubious at best.

Time series research has not been insensitive to the needs of empirical workers. In fact, time series methodology has extensively examined the question of modelling processes which are stationary about a deterministic trend; and deterministic trends are capable of dealing with nonstationary means. The methodology of stationary time series analysis then extends in a straightforward fashion to such trending series. The approach is well exposited in the work of Grenander and Rosenblatt (1957) and Anderson (1971). The nonstationarity of second moments has received less attention. However, recent work by White (1980) and White and Domowitz (1984) in econometrics has provided important results on time series modelling with heterogeneously distributed errors. As a result of these generalizations, empirical workers have been generally satisfied with the approach of modelling economic time series as processes with determin-

[1] We are grateful to Joon Park, two referees, Lars Hansen, and Gary Chamberlain for helpful comments on earlier versions of this paper. We thank seminar participants at UC San Diego, UCLA, USC, UC Berkeley, Stanford, Chicago, Penn, Minnesota, and MIT for useful feedback. Our thanks also go to Glena Ames for her skill and effort in preparing this manuscript. This paper was written while the first author was an Alfred P. Sloan Doctoral Dissertation Fellow at Yale University. The research reported here was supported by the NSF under Grant No. SES 8519595.

istic trends. Recently, however, some growing dissatisfaction has been expressed with the assumption that economic time series typically decompose into deterministic trends and stationary (or ergodic) components. In particular, Nelson and Plosser (1982) argue that a large number of macroeconomic aggregates, such as GNP, are better modelled as random walks or integrated processes of order one ($I(1)$ processes), rather than stationary (or ergodic) processes with a trend. An integrated process specification for time series has several important statistical implications. One of these is that the potential cost of misspecification of the generating mechanism is often substantial. A number of papers in the literature have already examined the effects of spuriously detrending integrated series. Thus, Nelson and Kang (1981, 1983) argue that the regression of a driftless random walk against a time trend will result in the inappropriate inference that the trend is significant. Further, detrended random walks will exhibit spurious correlation. Similar results for a different class of models have been generated by Hoffman, Low, and Schlagenhauf (1984) and Mankiw and Shapiro (1985, 1986). These studies have all obtained results by Monte Carlo simulations.

Integrated processes also pose problems for the empirical worker because of the probabilistic properties of the series. In particular, conventional strong laws and central limit theory do not apply to standardized sums of the realizations of an integrated process. These probabilistic properties and their statistical implications have been extensively analyzed in recent work by Phillips (1986, 1987a, 1987b) and Phillips and Durlauf (1986).

The aim of the present paper is to analyze the effects of misspecification of the generating mechanism of a nonstationary time series in terms of deterministic trends. Our approach is based on the recent study of spurious regressions by Phillips (1986). The techniques developed there and in the other papers cited in the previous paragraph may be directly applied to study the effects of spurious detrending in regression. In this paper we provide an explicit analytical solution to the asymptotic behavior of spuriously detrended regressions and thereby help to unify and explain many of the disparate Monte Carlo results that presently exist in the literature.

The paper is organized as follows: Section 2 derives the statistical properties of time trend regressions when the time series is in reality an integrated process. Section 3 extends these results to the case of integrated processes with drift. Section 4 provides several useful theorems on hypothesis testing in nonstationary models. We examine tests of the random walk versus deterministic time trend model. Section 5 explores the effect of spurious detrending on time series regression and Section 6 applies the theory to study the impact of spurious detrending on excess volatility tests. Section 7 provides a summary and some conclusions. A technical appendix contains explicit formulae for the limiting distribution theory discussed in the text of the paper.

2. TIME TRENDS AND INTEGRATED PROCESSES: REGRESSION PROPERTIES

We initially concern ourselves with analysis of the least squares regression:

$$y_t = \hat{\alpha} + \hat{\beta}t + \hat{e}_t \qquad (t = 1, \ldots, T), \tag{1}$$

where y_t is assumed to follow a deterministic trend about white noise, rendering (1) a correctly specified model, when in fact the true generating mechanism for $\{y_t\}$ is:

$$y_t = y_{t-1} + u_t. \tag{2}$$

In order to derive the behavior of (1), we need to place some restrictions on the u_t process in (2). We require only that the partial sum process $S_t = \sum_1^t u_j$ constructed from u_t satisfies a functional CLT of the type discussed and applied in Phillips (1987a). Thus, if $r \in [0,1]$ and we define $X_T(r) = T^{-1/2} S_{[Tr]}$ then we require

$$X_T(r) \Rightarrow B(r), \quad \text{as} \quad T \to \infty. \tag{3}$$

Here, the symbol " $\Rightarrow$ " signifies weak convergence of the associated probability measures and $B(r)$ denotes Brownian motion with variance given by

$$\sigma^2 = \lim_{T\to\infty} T^{-1} E\left(S_T^2\right).$$

Functional CLT's such as (3) are known to apply for a rather general class of innovation sequences u_t, which allow for weak dependence and some heterogeneity over time. The reader is referred to Phillips (1987a) for discussion, references, and a range of applications.

When $\sigma^2 = 1$ we call the process $B(r)$ standard Brownian motion and we use the notation $W(r)$. Thus, in general we may write $B(r) \equiv \sigma W(r)$ where the symbol " $\equiv$ " signifies equality in distribution. Frequently it is convenient to write these and other stochastic processes on $[0,1]$ without the argument as simply B and W.

Using the methods in Phillips (1986, 1987a) it is now easy to find the relevant asymptotic theory for the regression (1).

THEOREM 2.1:

$$(a) \quad T^{-1/2}\hat{\alpha} \Rightarrow 4\int_0^1 B - 6\int_0^1 rB \equiv N(0, 2\sigma^2/15),$$

$$(b) \quad T^{1/2}\hat{\beta} \Rightarrow 12\left[\int_0^1 rB - (1/2)\int_0^1 B\right] \equiv N(0, 6\sigma^2/5).$$

Thus, the estimated time trend coefficient in (1) is consistent and converges to the (true) structural coefficient of zero. However, the constant term $\hat{\alpha}$ in the regression is not consistent and its distribution actually diverges as $T \uparrow \infty$. Here we have an example where the nonstationarity of the true process affects the large sample properties of the regression coefficients differently. This result may be usefully related to the recent spurious regression theory developed in Phillips (1986). In this case we have the regression

$$y_t = \hat{\alpha} + \hat{\beta} x_t + \hat{\varepsilon}_t \tag{4}$$

where x_t and y_t are independent random walks or integrated processes such as (2). However, as shown in Phillips' paper and in contrast to (1) above, the constant $\hat{\alpha}$ in (4) has a divergent asymptotic distribution, whereas the coefficient $\hat{\beta}$ in (4) possesses a nondegenerate limiting distribution.

Both sets of results emphasize the importance of orders of magnitude of sampling variability in determining coefficient consistency. For the variable y_t, the sample moment $T^{-1}\sum_{t=1}^{T} y_t^2$ is $O_p(T)$. For the intercept, the corresponding moment is $O(1)$. For the time trend, the moment $T^{-1}\sum_{t=1}^{T} t^2$ is $O(T^2)$. When the regressor sample variance is of the same order of magnitude as that of the dependent variable, the limiting distribution is usually nondivergent. When the regressor sample variance is of a higher order of magnitude than that of the dependent variable, the regression coefficient usually has a degenerate limiting distribution, despite the nonstationarity of the regressand and the misspecification of the equation. In the latter case, the higher order of magnitude provides leverage in discriminating between the time paths of regressor and regressand. This discriminatory power is clearly seen in the case of the time trend in equation (1). For an integrated y_t the sample variability of y_t is $O_p(T)$. For the regressor t, the sample variability is $O(T^2)$. The probability that a sample path for y_t achieves the same order of magnitude of sample variability approaches zero as $T \to \infty$.

The consistency of $\hat{\beta}$ does not translate into desirable properties for conventional significance tests that $\beta = 0$ in (2). Our next theorem characterizes the asymptotic behavior of the main regression diagnostics for (1). We use $F_{\gamma=0}$ to denote the regression F statistic for testing the hypothesis $\gamma = 0$, DW to denote the Durbin-Watson statistic, and R^2 to denote the coefficient of determination. We say that a statistic diverges when it is asymptotically unbounded with probability one.

THEOREM 2.2: *For the regression equation* (1) *under the generating mechanism* (2) *as* $T \uparrow \infty$: (*a*) $F_{\beta=0}$ *diverges*; (*b*) $F_{\alpha=0}$ *diverges*; (*c*) $DW \xrightarrow[p]{} 0$; (*d*) R^2 *has a nondegenerate limiting distribution.*

Explicit distributional results for (a)–(d) in Theorem 2.2 are provided in the Appendix. The particularly interesting results, for hypothesis testing, are (a) and (b). The distributions of both traditional F tests diverge. The divergence of the test that $\alpha = 0$ is not surprising since the coefficient estimate $\hat{\alpha}$ does not possess an asymptotic distribution. The divergence of $F_{\beta=0}$ is more surprising. Here we have a case where a regression coefficient converges to zero, yet the standard statistical test that the coefficient equals zero diverges to infinity. This latter result mirrors the asymptotic behavior of the F statistic for $\hat{\beta}$ in the spurious regression model (4). In this case Phillips (1986) shows that $\hat{\beta}$ also possesses a divergent F statistic despite the fact that $\hat{\beta}$ itself has a nondegenerate limiting distribution.

Our results for the $F_{\beta=0}$ test corroborate the Monte Carlo findings of Nelson and Kang (1981, 1983). Nelson and Kang provided evidence that finite sample F tests of the significance of the time trend are severely biased. Theorem 2.2 indicates why this bias occurs and shows that it is exacerbated as the number of observations increases.

The results for the Durbin-Watson statistic appear quite promising for the empirical worker. This diagnostic will, with probability one, reject the hypothesis

of correct model specification, as $T \to \infty$. The fact that $\hat{\beta} \xrightarrow[p]{} 0$ in (1) ensures that the estimated regression residuals will evidence greater temporal dependence as $T \to \infty$. The dangers of incorrectly identifying an integrated process as a trend stationary process are naturally diminished to the extent that standard regression diagnostics will detect the specification error. The asymptotic behavior of the Durbin-Watson statistic suggests that the probability of mistaking a nonstationary series for a stationary series about trend is not particularly great for reasonably large data sets. These results strongly reinforce the recommendations made recently by Sargan and Bhargava (1983) concerning the use of the Durbin-Watson statistic as a discriminatory device for unit roots. However, even though global misspecification generates a low Durbin-Watson, a low Durbin-Watson does not necessarily imply that y_t is integrated.

Finally, the R^2 statistic converges weakly to a nondegenerate random variable in the limit as $T \uparrow \infty$. The Monte Carlo simulations of Nelson and Kang indicate that the expected value of this random variable is approximately .44.

We may therefore conclude that conventional hypothesis tests will generate an apparently statistically significant relationship between time and a zero mean, integrated dependent variable. The Durbin-Watson statistic, on the other hand, will provide an asymptotically powerful method of exposing the spurious regression. The results of this section strongly support the importance of combining hypothesis testing with specification analysis. A significant time trend may be the result of global misspecification as well as the presence of a "structural" trend.

The asymptotic results for this section hold for a wide class of error processes. The functional CLT approach that we employ permits a great diversity of potential innovation sequences, in contrast to Monte Carlo studies which have relied upon iid normal errors in the simulations. Our results and the analytic formulae given in the Appendix, therefore, provide a substantial generalization of this literature.

Analogously, the asymptotics verify that Monte Carlo results obtained for the iid normal case may be expected to hold in a much more general setting. The robustness of the Monte Carlo results may then be seen as a manifestation of the invariance principle (3) that underlies our theory and which obtains for a wide class of different innovation sequences.

3. RANDOM WALKS WITH DRIFT

Time series such as GNP clearly are not random walks about a zero mean. It is therefore important to consider data generating processes other than (2) which allow for some secular drift over time. We shall consider as an alternative to equation (2) an $I(1)$ process with drift:

$$y_t = \mu + y_{t-1} + u_t. \tag{5}$$

The regression theory for this data generating process is identical to that of Section 2. In fact, we have the following theorem.

THEOREM 3.1: *If* (5) *constitutes the true data generating process, then for the least squares regression* (1) *we have*:

(*a*) $T^{-1/2}\hat{\alpha} \Rightarrow N[0, 2\sigma^2/15]$;

(*b*) $T^{1/2}(\hat{\beta} - \mu) \Rightarrow N[0, 6\sigma^2/5]$, *exactly as in the zero drift case*;

(*c*) *the asymptotic properties of* $F_{\beta=0}$, $F_{\alpha=0}$, *DW, and* R^2 *are identical to the zero drift case.*

The intuition behind these results is straightforward. An $I(1)$ process with drift may be converted to a series with a time trend in the sense that

$$y_t = \mu + y_{t-1} + u_t = \mu t + y_t^* \tag{6}$$

where y_t^* is a trendless random walk. Regressing this sum against a time trend will generate results identical to those of Section 2 except for the μt term which is now captured by the trend coefficient in the regression.

4. REGRESSION DIAGNOSTICS

In this section, we discuss regression diagnostics which will permit discrimination between stationary and nonstationary time series models. A classical F test to discriminate between models (1) and (2) would analyze a hybrid regression of the form:

$$y_t = \alpha + \beta t + \gamma y_{t-1} + u_t. \tag{7}$$

Model specification may be tested directly via F tests applied to the various regression coefficients. Acceptance of the hypothesis that $\beta = 0$, $\gamma = 1$ corresponds to acceptance of the $I(1)$ process model and the presence of a unit root in the generating mechanism of y_t. Conversely, acceptance of the hypothesis that $\gamma = 0$ corresponds to a complete rejection of the random walk model. Note that the hypothesis $\gamma = 0$ is a polar case to the alternatives of an integrated process. Frequently we will be concerned with alternatives that include stationary autoregressive coefficients γ with $|\gamma| < 1$. Tests of $\gamma = 1$ and joint tests of $\beta = 0, \gamma = 1$ in (7) have been studied recently at the present level of generality by Phillips and Perron (1988). This paper provides extensions to models with trend and drift of the test procedures developed in Phillips (1987a) for detecting the presence of a unit root in models with general time series innovations. In particular, Phillips and Perron provide modifications to the conventional test statistics which eliminate nuisance parameter dependencies through a nonparametric serial correlation correction. Since these procedures are explored in detail in the papers cited they will not be pursued further here.

Instead, our attention will focus on some additional regression diagnostics associated with the misspecified equation (1). The misspecification generated by treating equation (1) as the correct model is subject to detection by the empirical researcher in several ways. For example, specification tests in the Hausman class provide potential test statistics when the transformation of equation (1) generates a stationary regression.

Specifically, we first consider the differencing test discussed by Plosser, Schwert, and White (1982) to examine whether equation (1) is correctly specified. We are

interested in how the test performs when equation (2) is the true specification and when the model estimated by (1) is assumed to be correct. The differencing test consists of examining the regression

$$\Delta y_t = \hat{c} + \hat{\eta}_t, \tag{8}$$

where, under the null, η_t is a difference of white noise, and computing the test statistic:

$$(\hat{\beta} - \hat{c})\big(\mathrm{Var}(\hat{c}) - \mathrm{Var}(\hat{\beta})\big)^{-1}(\hat{\beta} - \hat{c}), \tag{9}$$

where $\hat{\beta}$ and $\hat{c}$ are the regression coefficients from equations (1) and (8), and $\mathrm{Var}(\hat{c})$ and $\mathrm{Var}(\hat{\beta})$ represent the estimated variances of the regression coefficients under the assumption that (1) is the true model. When (2) is the true model, consistency of (9) as a test statistic requires that (9) diverge asymptotically.

To understand the properties of (9) note that under the erroneous assumption that (1) is the correct model, the econometrician would employ the formulae $\mathrm{Var}(\hat{c}) = 2T^{-2}s^2$ and $\mathrm{Var}(\hat{\beta}) = (\Sigma_{t=1}^{T}(t-\bar{t})^2)^{-1}s^2$, where s^2 is the estimate of the error variance. Suppose further, that the estimate s^2 is obtained from regression (1). The test statistic (9) can be rewritten

$$\frac{\big(T^{1/2}(\hat{\beta} - \hat{c})\big)^2\left[2 - T^2\left(\sum_{t=1}^{T}(t-\bar{t})^2\right)^{-1}\right]^{-1}}{T^{-1}s^2}.$$

When (2) is the true model, our previous results indicate that the statistic will possess a nondegenerate asymptotic distribution. Thus, the statistic is not asymptotically divergent and the differencing test is inconsistent.

On the other hand, the Durbin-Watson statistic provides a promising second diagnostic against misspecification. Recent work by Engle and Granger (1987) and Phillips and Ouliaris (1986) on cointegration has provided evidence that the Durbin-Watson statistic can be a powerful diagnostic against nonstationarity. Section 2 verified that the Durbin-Watson statistic will converge to zero at the rate T^{-1}. Large data sets should therefore generate very low Durbin-Watson statistics when nonstationarity is present and ignored, in time series regressions. However, low Durbin-Watson statistics traditionally signal the implementation of rather standardized corrective procedures and these do not always lead to more appropriate specifications. We conclude this section, therefore, with an example of how mechanical corrections for autocorrelation can still leave difficulties in inference.

The convergence of the Durbin-Watson statistic to zero does not imply that two stage generalized least squares (GLS) procedures will produce a correction that directly eliminates nonstationarity. In particular, these procedures will not lead asymptotically to the same coefficient estimates and tests as direct differencing of the data to eliminate nonstationarity. For example, suppose that (1) is estimated with an autoregressive correction using the first order serial correlation

coefficient $\hat{\rho}_T$ of the residuals from the OLS regression (1). GLS applied to (1) is not asymptotically equivalent to direct estimation of (8).

To see that the GLS estimates of β fail to converge to the estimate of β from the differenced equation (8) consider the sequence of GLS estimators:

$$\hat{\beta}_T = \frac{\sum_{t=2}^{T} \left(t - \bar{t} - \hat{\rho}_T\left(t - 1 - \overline{t-1}\right)\right)\left(y_t - \bar{y} - \hat{\rho}_T\left(y_{t-1} - \bar{y}_{-1}\right)\right)}{\sum_{t=2}^{T} \left(t - \bar{t} - \hat{\rho}_T\left(t - 1 - \overline{t-1}\right)\right)^2}. \tag{10}$$

This GLS estimator is the usual Cochrane-Orcutt estimator. We shall assume that the estimates for $\hat{\rho}_T$ are constructed as $\hat{\rho}_T = 1 - \frac{1}{2}DW_T$, where DW_T is the estimated Durbin-Watson statistic based upon the OLS residuals from (1). Finally, notice that the estimate of $\hat{\beta}_T$ may be rewritten as

$$\begin{aligned}\hat{\beta}_T = &\left[\sum_{t=2}^{T} \Delta y_t + \xi_T\left[\sum_{t=2}^{T} (t-1-\bar{t})\Delta y_t + \sum_{t=2}^{T} (y_{t-1} - \bar{y})\right.\right. \\ &\left.\left. + \xi_T \sum_{t=2}^{T} (y_{t-1} - \bar{y})(t-1-\bar{t})\right]\right] \\ &\div \left[(T-1) + \xi_T\left[2\sum_{t=2}^{T}(t-1-\bar{t}) + \xi_T \sum_{t=2}^{T}(t-1-\bar{t})^2\right]\right] \\ &+ O_p(T^{-1})\end{aligned} \tag{11}$$

where $\xi_T = 1 - \hat{\rho}_T$. Note that ξ_T is $O_p(T^{-1})$ since the Durbin-Watson statistic is $O_p(T^{-1})$.

The differenced equation estimate of the time trend coefficient is, of course, equal to $\Sigma_{t=2}^{T}\Delta y_t/(T-1)$. The relationship between (11) and the differenced regression coefficient in (8) will therefore depend upon the terms involving ξ_T. Note that the term $\Sigma_{t=2}^{T}\Delta y_t$ is $O_p(T^{1/2})$. However, the term $\Sigma_{t=2}^{T}(t-1-\bar{t})\Delta y_t$ is $O_p(T^{3/2})$, $\Sigma_{t=2}^{T}(y_{t-1}-\bar{y})(t-1-\bar{t})$ is $O_p(T^{5/2})$, and the term ξ_T is $O_p(T^{-1})$. As a result, the numerator is not asymptotically dominated by the leading term. The first, second, and fourth terms will contribute asymptotically in this expression. Similar reasoning indicates that first and third terms in the denominator will both contribute to the asymptotic distribution of the estimator. As a result, the two stage GLS procedure fails to converge to the differenced regression.

Similar reasoning implies that the F statistic from the GLS regression will also fail to possess a limiting χ^2 distribution. In particular, we have the following theorem.

THEOREM 4.1: *Estimation of equation* (1) *by a two stage* GLS *procedure, employing a Cochrane-Orcutt correction for first order autocorrelation with the estimate* $\hat{\rho}_T = 1 - \frac{1}{2}DW_T$, *will yield coefficient estimates* $\hat{\beta}_{\text{GLS}}$ *and* $\hat{\alpha}_{\text{GLS}}$ *such that*

(a) $T^{1/2}\hat{\beta}_{GLS}$ *converges weakly to a nonnormal random variable*; *(b)* $T^{1/2}\hat{\alpha}_{GLS}$ *converges weakly to a nonnormal random variable*; *(c)* $F_{\beta=0}$ *converges weakly to a non-*χ_1^2 *random variable*; *(d)* $DW \xrightarrow[p]{} 0$. *Explicit characterizations of the limiting distributions are given in the Appendix.*

The failure of the GLS procedure to converge in distribution to a differenced regression has important ramifications for the applied worker. For the case of stationary errors, Amemiya (1973) has verified that across a large class of models, coefficient estimates derived from a GLS procedure with an estimated covariance matrix possess the same asymptotic distribution as coefficient estimates generated from the same GLS procedure where the covariance matrix is known. For nonstationary, nonergodic errors, this equivalence does not hold.

Further, the GLS regression coefficients and test statistics will, in the case of a misspecified integrated time series, diverge from the conventional asymptotics. The limiting distribution is not χ_1^2 under the null. Thus, proposals to automatically prewhiten data in order to vitiate the impact of error autocorrelation on the nominal asymptotic size of standard test statistics will fail to generate the desired χ^2 criteria in the nonstationary case. The GLS procedure will produce a test statistic with a well defined limiting distribution. However, if the GLS test statistic is erroneously treated as χ_1^2, the nominal and actual asymptotic test sizes will remain unequal. Phillips and Durlauf (1986) discuss some new methods for transforming test statistics with unconventional limit distributions into χ^2 form.

The failure of the GLS procedure to generate χ^2 asymptotics reinforces the importance of the Granger and Newbold (1974) discussion of spurious regressions. Even if standard regression diagnostics are employed to detect departures of the regression errors from white noise, spurious inference is still a danger. The diagnosis of misspecification via the Durbin-Watson statistic does not automatically lead to a correct asymptotic nominal size for a mechanical GLS based test. Thus, the failure to model nonstationarity will not be corrected by automatic GLS correction procedures.

5. HYPOTHESIS TESTING WITH DETRENDED DATA

The presence of integrated processes in time series data poses problems to the empirical worker due to the impact of nonstationarity on statistical inference. Regression with integrated processes generates nonnormal coefficient estimates and non-χ^2 test statistics. The failure to account for these deviations from standard theory will lead to improper inferences. The Granger-Newbold work on spurious regressions provided some simulation evidence that the regression of one integrated process on another independent integrated process leads to conventional coefficient tests that are seriously biased towards the rejection of the hypothesis of independence.

Further work on the problem of nonstationary regressors has generated evidence that the inappropriate detrending of integrated series will exacerbate the phenomenon of spurious regression. Monte Carlo evidence on this problem has

been accumulated in a series of papers by Mankiw and Shapiro (1985, 1986). Further simulation evidence may be found in Nelson and Kang (1981, 1983). Our approach in this paper helps to provide an explicit asymptotic answer to the issue of the impact of detrending on spurious regressions. We have the following theorem.

THEOREM 5.1: *Let x_t and y_t be generated as independent $I(1)$ processes*

$$y_t = y_{t-1} + u_t, \tag{12}$$

$$x_t = x_{t-1} + \eta_t. \tag{13}$$

The coefficients in the least squares regression

$$y_t = \hat{\alpha} + \hat{\beta} t + \hat{\gamma} x_t + \hat{u}_t \tag{14}$$

have the following asymptotic behavior: (*a*) $\hat{\gamma}$ *converges weakly to a nondegenerate random variable*; (*b*) $\hat{\beta} \xrightarrow[p]{} 0$; (*c*) $\hat{\alpha}$ *diverges*; (*d*) s^2 *diverges*; (*e*) $F_{\gamma=0}$ *diverges*; (*f*) $DW \xrightarrow[p]{} 0$. *Explicit characterizations of the limiting distributions for standardized versions of* (*a*)–(*f*) *are given in the Appendix.*

In (14) the time trend is the only consistently estimated regression coefficient. The coefficient $\hat{\gamma}$ relating the two time series possesses a nondegenerate limiting distribution. This mirrors the earlier asymptotic results obtained in Phillips (1986) for the original Granger-Newbold spurious regressions model.

The impact of detrending on the asymptotic distributions may be seen by considering the asymptotic distribution of $\hat{\gamma}$ when the time trend is omitted. The asymptotic behavior of $\hat{\gamma}$ in the latter case is derived in Phillips (1986). Both formulae are given in the Appendix (see (A9) and (A15)) and they may be seen to differ by the presence of terms which express the interaction between the time trend and the nonstationary series. Thus the Monte Carlo results of Dickey-Fuller (1979, 1981), Mankiw-Shapiro (1985, 1986), and Nelson-Kang (1981, 1983) are corroborated by analytical evidence and are generalized to a much wider class of processes. Note that the time trend affects the asymptotic behavior of $\hat{\gamma}$ in (14) despite the fact that $\hat{\beta}$ converges to zero.

The F test, as indicated in (e), diverges as in the nondetrended case. The suitably normalized asymptotic distribution of the F test (see (A13) in the Appendix) is affected by the presence of the time trend, but the qualitative result is not. Regardless of the inclusion of the time trend, the F test will diverge. Any suggestion that (in the nonstationary model) the time trend biases hypothesis testing is, thus, at best a small sample result. The nonstationarity of the underlying series is the critical issue, rather than inappropriate detrending.

An important and useful diagnostic in the detrended spurious regressions model is again the Durbin-Watson statistic, which from (f) we see converges in probability to zero. Thus, the Durbin-Watson continues to deliver sharp evidence of underlying nonstationarity in large samples even in detrended regressions. The inappropriateness of the conventional F test should thus become apparent to

the investigator from the inspection of residual DW diagnostics. Once again the results indicate that the employment of conventional significance tests must be suspect until the stationarity of the dependent variable is resolved.

We conclude this section by considering the impact of detrending when the regressand is a stationary series. Suppose y_t follows the process

$$y_t = u_t \tag{15}$$

where u_t is independent of η_t. The estimation of (14) will generate consistent estimates $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$. Further, these coefficients will all converge to zero. Spurious correlation will still occur, however, in the sense that the asymptotic distribution of the $F_{\gamma=0}$ test will not converge to a χ_1^2 distribution. The nominal test size is exceeded by the actual test size. In addition, the presence of the time trend in the regression will affect the asymptotic distributions of the various coefficients. These results are formalized in the next Theorem.

THEOREM 5.2: *Let y_t be generated by* (15) *and x_t be generated by* (13). *Define $z_t' = (u_t, \eta_t)$ and the vector partial sum $S_t = \sum_1^t z_j$. Suppose the elements of S_t satisfy the invariance principle* ($A21$). *Then* (*a*) $T\hat{\gamma}$ *converges to a nonnormal random variable*; (*b*) $T^{3/2}\hat{\beta}$ *converges to a nonnormal random variable*; (*c*) $T^{1/2}\hat{\alpha}$ *converges to a nonnormal random variable*; (*d*) $F_{\gamma=0}$ *converges to a non-χ_1^2 random variable. Again, explicit characterizations of the limiting distributions are provided in the Appendix.*

It is easy to see from the explicit formulae given in the Appendix (see (A16) and (A17)) that the presence of a time trend in (14) does affect the asymptotic distribution of the regression coefficients. However, just as in the Granger-Newbold case, the inclusion of the time trend has only a qualitative effect on the asymptotic distributions. The convergence rates for the coefficients are unaffected (see (A16) and (A20) in the Appendix).

The impact to detrending on the asymptotic properties of regressions such as (14) is particularly significant when the regressor is an integrated process with drift. Suppose that x_t is generated by

$$x_t = \mu + x_{t-1} + \eta_t = \mu t + x_t^*, \quad \mu \neq 0, \tag{16}$$

where x_t^* is a driftless random walk. Consider the OLS regression

$$y_t = \hat{\alpha} + \hat{\gamma} x_t + \hat{u}_t \tag{17}$$

where y_t is generated by (15). The standardized regression coefficient $\hat{\gamma}$ and associated test statistics now follow the standard asymptotic theory. In fact, it is easy to see that

$$T^{3/2}\hat{\gamma} \Rightarrow N(0, 12\sigma^2/\mu^2).$$

The asymptotic behavior of $\hat{\gamma}$ is therefore equivalent to the asymptotic behavior

of $\tilde{\gamma}$ in the OLS regression

$$(18) \qquad y_t = \tilde{\alpha} + \tilde{\gamma}(\mu t) + \tilde{u}_t.$$

The time trend component of x_t in (16) is $O(t)$, a higher order of magnitude than the driftless $I(1)$ component which is $O_p(\sqrt{t})$. It therefore dominates the asymptotics, rendering them conventional in this scalar case.

On the other hand, if we estimate (14) rather than (17), then the asymptotics associated with $\hat{\gamma}$ will replicate the results of Theorem 5.2, as the arguments in Section 3 indicate. Removing the time trend from x_t naturally renders the x_t^* component asymptotically significant. The standard asymptotics will no longer apply. Suitably normalized moments of x_t^* converge to random variables rather than constants, unlike the original x_t series.

We therefore conclude that spurious detrending does affect the asymptotic behavior of regressions with integrated regressors. However, this impact is secondary to the statistical properties of the time series being analyzed. Spurious detrending may exacerbate, but is not the source of unconventional asymptotics, when the regressors possess zero drift. If the underlying regressors possess nonzero drift, then detrending may, in fact, induce departures from conventional asymptotics as we have seen in the final example above.

6. SPURIOUS DETRENDING IN EXCESS VOLATILITY ANALYSIS

In this section, we extend the analysis of spurious detrending in regressions to spurious detrending in excess volatility testing. Specifically, we examine the statistical properties of excess volatility tests when the underlying time series are integrated processes, possibly with drift, rather than ergodic processes with deterministic trends.

Excess volatility tests represent a method developed by Shiller (1979, 1981a, 1981b) to analyze the rationality of asset market prices and returns. Consider a sequence of forecasts $\{P_t\}$ and a sequence of realizations $\{P_t^*\}$ such that each element of P_t represents a prediction of the corresponding element of P_t^*. The difference $u_t = P_t^* - P_t$ will therefore equal a set of observations of forecast errors. If these forecast errors are generated by "rational" forecasts, then they must be orthogonal to information available at the time the forecasts are made. Since the forecasts themselves are part of this information set, this implies that

$$(19) \qquad \mathrm{Var}(P_t^*) = \mathrm{Var}(P_t) + \mathrm{Var}(u_t)$$

or

$$(20) \qquad \mathrm{Var}(P_t^*) - \mathrm{Var}(P_t) > 0.$$

Equation (20) represents a testable restriction on the time series. Shiller (1981a) applied (20) to an examination of the relation between stock prices and

dividends. In particular, he has examined whether stock prices are equal to the expected present discounted value of future dividends, i.e.

$$(21) \qquad H_0: P_t = \sum_{i=0}^{\infty} \delta^i E_t(D_{t+i})$$

where P_t is stock price, δ is discount rate, and D_{t+i} is dividend payment at $t+i$. The realizations P_t^* are constructed by replacing the expected values of dividends with their historical realizations. Shiller assumed that in both cases, P_t and P_t^* were ergodic with deterministic trends. He thus detrended the series and calculated the sample variances finding dramatic violations of the inequality bound (20).

A number of objections have been raised to the asymptotic properties of excess volatility tests. A number of authors, most notably Marsh and Merton (1983) and Kleidon (1986), have argued that the stationarity assumption is false, as the short rate and dividend series are integrated processes. In particular, Kleidon (1986) has provided Monte Carlo evidence that when the forecast series is an integrated or near integrated process, then the calculation of variance inequality statistics such as (20) will lead to a large number of negative realizations and hence rejections of the null hypothesis even when the null hypothesis is true.[2] Kleidon further contends that the failure of the sample variances to converge in probability to constants when P_t and P_t^* are integrated asymptotically invalidates the use of variance bounds tests altogether. In addition, Kleidon has found results similar to Mankiw and Shapiro (1986) suggesting that the detrending of forecast series and *ex post* rational series exacerbates the rate of rejection of the Shiller excess volatility test under the null hypothesis.

Marsh and Merton (1983) have brought the critique of the excess volatility tests in the case of stock prices a step further. These authors argue first that the dividend process is not exogenous but rather a choice variable determined by stock prices. They then demonstrate that if the stock price series is integrated, the inequality (20) is reversed for all data realizations, when stock prices Granger cause dividends. Within this framework, the Shiller inequality violations constitute confirmation of the efficient market hypothesis.

The Monte Carlo and analytical results cited in these studies may be given a precise formulation by employing the asymptotic techniques of Section 5. We shall therefore consider the asymptotic properties of excess volatility tests when the forecast series P_t obeys

$$(22) \qquad P_t = P_{t-1} + \eta_t.$$

Further, we assume that u_t and η_t possess the statistical properties needed for Theorem 5.2. In particular, we assume that the process (u_t, η_t) has partial sums which satisfy the multivariate functional CLT of Phillips and Durlauf (1986).

[2] Test bias refers to the discrepancy between actual and nominal asymptotic test size. For excess volatility tests which accept or reject based upon the nonnegativity of a sample variance difference, the nominal test size is zero.

Finally, let P_{dt}, P_{dt}^*, and u_{dt} denote that detrended series corresponding to P_t, P_t^*, and u_t respectively.

In order to calculate the asymptotic properties of (20) we calculate the sample moment differential

$$\text{(23)} \qquad \text{Vol} = T^{-1}\sum_{t=1}^{T}\left(P_{dt}^* - \bar{P}_d^*\right)^2 - T^{-1}\sum_{t=1}^{T}\left(P_{dt} - \bar{P}_d\right)^2$$

$$= T^{-1}\sum_{t=1}^{T}\left(u_{dt} - \bar{u}_d\right)^2 + 2T^{-1}\sum_{t=1}^{T}\left(P_{dt} - \bar{P}_d\right)\left(u_{dt} - \bar{u}_d\right).$$

This expression will be greater than zero iff

$$\text{(24)} \qquad \text{Vol}' = \frac{\sum_{t=1}^{T}\left(P_{dt} - \bar{P}_d\right)\left(u_{dt} - \bar{u}_d\right)}{\sum_{t=1}^{T}\left(u_{dt} - \bar{u}_d\right)^2} > -\frac{1}{2}.$$

The expression (24) is nothing more, however, than the regression coefficient $\hat{\gamma}$ in

$$\text{(25)} \qquad P_t = \hat{\alpha} + \hat{\beta}t + \hat{\gamma}u_t + \xi_t.$$

Thus, we see that an excess volatility test is equivalent to an examination of the value of a regression coefficient when forecasts are regressed against a constant, time trend, and forecast errors. This result permits an easy linkage of the Monte Carlo results in Flavin (1983) and Kleidon (1986) to the asymptotic theory presented in Section 5.

Theorem 6.1: *The Shiller excess volatility test statistic Vol′ will possess nonnegligible asymptotic size when the forecast series is an integrated process. The limiting distribution of Vol′ is nondegenerate and is explicitly characterized by (A22) in the Appendix.*

The excess volatility test possesses nonnegligible asymptotic size because of the inconsistency of the regression coefficient $\hat{\gamma}$ in (25). As a result, the sample variance differential fails to converge to a constant greater than zero, as would occur in the standard, ergodic case. The fact that the excess volatility test statistic converges to a random variable implies that there is no guarantee that the test will possess zero asymptotic size under the null, even though we are testing against a nonlocal alternative and are employing a zero-one decision rule based upon the value of the statistic. In the ergodic case, $\hat{\gamma}$ converges to zero under the null; thus we accept the hypothesis with asymptotic probability 1 since we are

examining whether $\hat{\gamma}$ is greater than $-\frac{1}{2}$. The Kleidon and Flavin results thus detect the failure of the variance differential to fulfill the requirements of the strong law of large numbers. The limiting distribution of the Vol′ statistic possesses nonzero mass for values less than $-\frac{1}{2}$.

The asymptotic results do suggest that two arguments in the empirical literature on excess volatility testing are not correct. First, there is Kleidon's assertion that with an integrated forecast series, excess volatility tests are not interpretable. Kleidon assumed that the failure of the sample variances of P_t and P_t^* to converge implies that the differential failed to possess a limiting distribution. As the theorem indicates, this conjecture is false. The $\text{Var}(P_t)$ component of the two sample variances cancels out in the calculation of the Vol′ statistic, which generates a well defined limiting distribution for the statistic. One could therefore perform an asymptotically valid excess volatility test by comparing the realization of Vol′ to the critical value of its limiting distribution.

Second, the oral tradition that has become associated with Marsh and Merton's work on excess volatility testing, that nonergodicity of the forecast series somehow reverses the Shiller inequality, is also incorrect. There is no guarantee under the null that the Vol′ test statistic is always less than $-\frac{1}{2}$. The Marsh and Merton results stem from a reversal of the causality between dividends and stock prices, rather than from nonergodicity per se. The Shiller inequality is reversed only when stock prices drive dividends.

Finally, we conclude with an interpretation of a violation of the excess volatility bound when the forecast series is integrated. Consider the ratio $\text{Var}(P_{dt}^*)/\text{Var}(P_{dt})$. The excess volatility bound requires that this ratio be greater than 1. This ratio is linked to the regression

$$P_{dt}^* = \alpha + \gamma P_{dt} + \xi_t \tag{26}$$

in that $\text{Var}(P_{dt}^*)/\text{Var}(P_{dt}) > 1$ is equivalent to

$$\hat{\gamma} - 1 > -\sum_{t=1}^{T} (u_{dt} - \bar{u}_d)^2 / 2 \sum_{t=1}^{T} (P_{dt} - \bar{P}_d)^2$$

where $\hat{\gamma}$ is the OLS coefficient. When the forecast series is integrated and $\gamma = 1$, the right hand side of this inequality converges to zero. Thus, a test of the variance inequality bound for integrated processes is a test of whether P_{dt}^* and P_{dt} are cointegrated with cointegration vector $(1, -1)$. Again, a consideration of this ratio without confidence bounds will lead to poor size properties for the test, since $\hat{\gamma}$ is skewed to the left, as verified by Stock (1987). However, direct tests of cointegration in this case, using the methods of Phillips and Ouliaris (1986), are now available.

7. SUMMARY AND CONCLUSIONS

This paper develops a framework for understanding the behavior of integrated time series which are misspecified as trend stationary time series. We have

provided an asymptotic theory for the behavior of regression coefficients in models which attempt to estimate time trends when the dependent variables are actually $I(1)$ processes. In addition, the asymptotic properties of test statistics associated with the misspecified regression have been explored. In particular, the F statistic examining the significance of the time trend coefficient will diverge when the dependent variable is a zero drift random walk. This divergence occurs in spite of the fact that the coefficient estimate converges in probability to zero.

Formal asymptotic results have also been developed for testing whether a time series is an integrated process or a stationary process about a deterministic trend. This includes classical F and Hausman type procedures. These test statistics possess the feature that they do not possess limiting χ^2 distributions when the data generating process is nonstationary. This implies that when the null hypothesis is that the series is integrated, excess rejection will normally occur if the limiting distribution is incorrectly treated as χ^2.

Further, we have investigated a number of issues concerning statistical inference with spuriously detrended data. We have investigated the behavior of spurious regressions among inappropriately detrended, nonstationary series. The detrending of the series does affect the limiting distributions of the regression coefficients. Test statistics are also affected. However, the impact of detrending on hypothesis testing is a second order effect. The nonstationarity of the series ensures that F statistics will erroneously indicate a statistical relationship with or without detrending.

Our analysis next provided some univariate results on the impact of detrending on regression analysis with cointegrated time series. Our results confirm Monte Carlo findings which indicate that the detrending of cointegrated series has an important effect on the asymptotic properties of the regression coefficient estimates and associated test statistics. In particular, detrending can increase hypothesis test bias.

Finally, we have provided a set of asymptotic results for excess volatility tests with detrended integrated series. We demonstrate that excess volatility tests with integrated processes possess nonnegligible asymptotic size. In addition, our formulae provide a framework for conducting excess volatility tests with integrated processes.

Our results constitute an extension of the ongoing literature in three senses. First, a number of issues, such as the asymptotic behavior of the differencing test for model specification, do not appear to have been addressed previously. Second, our results provide an analytic asymptotic theory whereas previous work on the issues we have addressed has been based upon Monte Carlo studies. Third, our results are robust with respect to a large variety of error processes. The underlying errors need not be normal, nor independent, nor identically distributed.

The Durbin-Watson statistic has been shown to possess promising asymptotic properties as a regression diagnostic in this context. This confirms earlier work by Engle-Granger and Sargan-Bhargava on related topics. The Durbin-Watson statistic converges to zero when an integrated process is erroneously treated as

stationary. The unconventional asymptotics for the test statistics considered in this paper underline the importance of correctly identifying the behavior of error processes prior to engaging in significance testing.

Potentially valuable extensions include the generalization of the results of this paper on regression asymptotics and diagnostics to more complicated systems of equations with integrated regressors. General methods for dealing with multiple systems of equations with integrated regressors have been developed by Phillips and Durlauf (1986), and these may be generalized to the case of detrended and spuriously detrended series.

Department of Economics, Stanford University, Stanford, CA, U.S.A.
and
Cowles Foundation for Research in Economics, Yale University, New Haven, CT, U.S.A.

Manuscript received April, 1986; final revision received October, 1987.

APPENDIX

Explicit Formulae for the Limiting Distribution Theory

In this Appendix we provide explicit characterizations of the limiting distributions of the various estimators and test statistics considered in the text. Most of these involve simple functionals of Brownian motion. They are obtained using the approach explored in detail in Phillips (1986a, 1987a) and Phillips and Durlauf (1986) which makes use of functional CLT's such as (3) and the continuous mapping theorem. In some cases, particularly Theorem 6.1, the derivations require a theory of weak convergence to stochastic integrals with respect to Brownian motion. In such cases the results draw on the relevant theory given in Phillips (1987a) (in the scalar case) and Phillips (1987c) (in the matrix case). Since the derivations in all cases are relatively straightforward we state only the final results here. Details are given in an earlier version of this paper (Durlauf and Phillips (1986)) and Durlauf (1986) which can be obtained from the authors on request. The results below are stated with reference to the respective theorems in the text.

1. Formulae for Theorem 2.2:

$$\text{(A1)} \qquad T^{-1}F_{\beta=0} \Rightarrow \frac{12\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]^2}{\int_0^1 W(t)^2\,dt - \left[\int_0^1 W(t)\,dt\right]^2 - 12\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]^2};$$

$$\text{(A2)} \qquad T^{-1}F_{\alpha=0} \Rightarrow \frac{\left[2\int_0^1 W(t)\,dt - 3\int_0^1 tW(t)\,dt\right]^2}{\int_0^1 W(t)^2\,dt - \left[\int_0^1 W(t)\,dt\right]^2 - 12\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]^2};$$

$$\text{(A3)} \qquad TDW \Rightarrow \frac{\sigma_u^2}{\sigma^2\left[\int_0^1 W(t)^2 - \left[\int_0^1 W(t)\,dt\right]^2 - 12\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]\right]^2};$$

where

$$\sigma_u^2 = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E\left(u_t^2\right),$$

$$\sigma^2 = \lim_{T\to\infty} T^{-1} E\left(\sum_{t=1}^{T} u_t\right)^2;$$

$$\text{(A4)} \qquad R^2 = 1 - \left[\left[\int_0^1 W(t)^2\,dt - \left[\int_0^1 W(t)\,dt\right]^2 - 12\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]^2\right] \div \left[\int_0^1 W(t)^2\,dt - \left[\int_0^1 W(t)\,dt\right]^2\right]\right].$$

2. *Formulae for Theorem* 4.1: Set $T\xi_T = T(1-\hat{\rho}_T) = (T/2)DW$. Then

$$\text{(A5)} \qquad T\xi_T \Rightarrow \sigma_V^2 \div \left[2\sigma^2\left[\int_0^1 W(t)^2\,dt - \left[\int_0^1 W(t)\,dt\right]^2 - 12\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]^2\right]\right] \Rightarrow \xi, \quad \text{say;}$$

$$\text{(A6)} \qquad T^{1/2}\hat{\beta}_{\mathrm{GLS}} \Rightarrow \frac{\sigma W(1) + \left[\frac{\sigma W(1)}{2} - \sigma\int_0^1 W(t)\,dt + \xi\sigma\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]\right]}{1+\xi^2/12};$$

$$\text{(A7)} \qquad T^{1/2}\hat{\alpha}_{\mathrm{GLS}} \Rightarrow \sigma W(1) - \xi\sigma\int_0^1 W(t)\,dt - \left[\left\{\sigma W(1) + \xi\left[\frac{\sigma W(1)}{2} - \sigma\int_0^1 W(t)\,dt + \xi\sigma\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]\right]\right\}(1+\xi/2)\right] \div 1+\xi^2/12;$$

$$\text{(A8)} \qquad F_{\beta=0} \Rightarrow \frac{\left[\sigma W(1) + \xi\left[\frac{\sigma W(1)}{2} - \sigma\int_0^1 W(t)\,dt + \xi\sigma\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]\right]\right]^2}{\sigma_u^2\left(1+\xi^2/12\right)}.$$

3. *Formulae for Theorem* 5.1:

$$\text{(A9)} \qquad \hat{\gamma} \Rightarrow \left[\frac{1}{12}\sigma_V\sigma_W\left[\int_0^1 V(t)W(t)\,dt - \int_0^1 V(t)\,dt\int_0^1 W(t)\,dt\right] - \sigma_V\sigma_W\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]\left[\int_0^1 tV(t)\,dt - \frac{1}{2}\int_0^1 V(t)\,dt\right]\right] \div \left[\frac{1}{12}\sigma_W^2\left[\int_0^1 W(t)^2\,dt - \left[\int_0^1 W(t)\,dt\right]^2\right] - \sigma_W^2\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]^2\right] = \zeta_1, \quad \text{say;}$$

$$
\begin{aligned}
\text{(A10)}\qquad T^{1/2}\hat{\beta} \Rightarrow & \left[\sigma_W^2\left[\int_0^1 W(t)^2\,dt - \left[\int_0^1 W(t)\,dt\right]^2\right]\right] \\
& \times\left[\sigma_V\left[\int_0^1 tV(t)\,dt - \frac{1}{2}\int_0^1 V(t)\,dt\right]\right] \\
& -\sigma_V\sigma_W^2\left[\int_0^1 V(t)W(t)\,dt - \int_0^1 V(t)\,dt\int_0^1 W(t)\,dt\right] \\
& \times\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]\Bigg] \\
& \div\left[\frac{\sigma_W^2}{12}\left[\int_0^1 W(t)^2\,dt - \left[\int_0^1 W(t)\,dt\right]^2 - \sigma_W^2\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]^2\right]\right] \\
= & \ \zeta_2, \quad \text{say;}
\end{aligned}
$$

$$
\text{(A11)}\qquad T^{-1/2}\hat{\alpha} \Rightarrow \sigma_V\int_0^1 V(t)\,dt - \zeta_2/2 - \zeta_1\sigma_W\int_0^1 W(t)\,dt;
$$

$$
\begin{aligned}
\text{(A12)}\qquad T^{-1}s^2 \Rightarrow & \ \sigma_V^2\int_0^1 V(t)^2\,dt - \sigma_V^2\left[\int_0^1 V(t)\,dt\right]^2 + \zeta_2/12 \\
& +\zeta_1^2\left[\sigma_W^2\int_0^1 W(t)^2\,dt - \sigma_W^2\left[\int_0^1 W(t)\,dt\right]^2\right] \\
& -2\zeta_1\left[\sigma_V\sigma_W\left[\int_0^1 V(t)W(t)\,dt - \int_0^1 V(t)\,dt\int_0^1 W(t)\,dt\right]\right] \\
& -2\zeta_2\sigma_V\left[\int_0^1 tV(t)\,dt - \frac{1}{2}\int_0^1 V(t)\,dt\right] \\
& +2\zeta_1\zeta_2\sigma_W\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right] \\
= & \ \omega^2, \quad \text{say;}
\end{aligned}
$$

$$
\text{(A13)}\qquad T^{-1}F_{\gamma=0} \Rightarrow \frac{\zeta_1^2\left[\frac{\sigma_W^2}{12}\left[\int_0^1 W(t)^2\,dt - \left[\int_0^1 W(t)\,dt\right]^2\right] - \sigma_W^2\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]^2\right]}{\omega^2/12};
$$

$$
\text{(A14)}\qquad TDW \Rightarrow \left(\sigma_u^2 + \zeta_1^2\sigma_\eta^2\right)/\omega^2.
$$

In (A9)–(A14) $V(t)$ and $W(t)$ are independent standard Brownian motions and

$$
\sigma_V^2 = \lim_{T\to\infty} T^{-1}E\left(\sum_{t=1}^{T} u_t\right)^2,
$$

$$
\sigma_W^2 = \lim_{T\to\infty} T^{-1}E\left(\sum_{t=1}^{T} \eta_t\right)^2,
$$

$$
\sigma_u^2 = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E\left(u_t^2\right),
$$

$$
\sigma_\eta^2 = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E\left(\eta_t^2\right).
$$

When the time trend is omitted from the regression (14) we obtain, in place of (A9),

$$\text{(A15)} \qquad \hat{\gamma} \Rightarrow \frac{\sigma_V\left[\int_0^1 V(t)W(t)\,dt - \int_0^1 V(t)\,dt\int_0^1 W(t)\,dt\right]}{\sigma_W\left[\int_0^1 W(t)^2\,dt - \left[\int_0^1 W(t)\,dt\right]^2\right]},$$

a formula earlier derived in Phillips (1986).

4. Formulae for Theorem 5.2:

$$\begin{aligned}\text{(A16)} \qquad T\hat{\gamma} \Rightarrow &\left[\left[\frac{\sigma_V\sigma_W}{12}\int_0^1 W(t)\,dV(t) - \frac{\sigma_V\sigma_W}{12}V(1)\int_0^1 W(t)\,dt\right]\right.\\ &\left.-\sigma_V\sigma_W\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]\left[\int_0^1 V'(t)\,dt - \frac{V(1)}{2}\right]\right]\\ &\div\left[\frac{\sigma_W^2}{12}\left[\int_0^1 W(t)^2\,dt - \left[\int_0^1 W(t)\,dt\right]^2\right] - \sigma_W^2\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]^2\right]\\ &= \zeta_3, \quad \text{say;}\end{aligned}$$

$$\begin{aligned}\text{(A17)} \qquad T^{3/2}\hat{\beta} \Rightarrow &\left[\sigma_W^2\left[\int_0^1 W(t)^2\,dt - \left[\int_0^1 W(t)\,dt\right]^2\right]\sigma_V\left[\int_0^1 V'(t)\,dt - \frac{V(1)}{2}\right]\right.\\ &\left.-\sigma_W\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]\sigma_V\sigma_W\left[\int_0^1 W(t)\,dV(t) - V(1)\int_0^1 W(t)\,dt\right]\right]\\ &\div\left[\frac{\sigma_W^2}{12}\left[\int_0^1 W(t)^2\,dt - \left[\int_0^1 W(t)\,dt\right]^2\right] - \sigma_W^2\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]^2\right]\\ &= \zeta_4, \quad \text{say;}\end{aligned}$$

$$\text{(A18)} \qquad T^{1/2}\hat{\alpha} \Rightarrow \sigma_V V(1) - \zeta_4/2 - \zeta_3\sigma_W\int_0^1 W(t)\,dt,$$

$$\text{(A19)} \qquad F_{\gamma=0} \Rightarrow \frac{\zeta_3^2\left[\frac{\sigma_W^2}{12}\left[\int_0^1 W(t)^2\,dt - \left[\int_0^1 W(t)\,dt\right]^2\right] - \sigma_W^2\left[\int_0^1 tW(t)\,dt - \frac{1}{2}\int_0^1 W(t)\,dt\right]^2\right]}{\sigma_u^2/12},$$

where $V'(t) = V(1) - V(t)$ and the remaining notation is as defined above in paragraph 3. When the time trend is omitted from (14) we obtain, in place of (A16),

$$\begin{aligned}\text{(A20)} \qquad T\hat{\gamma} \Rightarrow &\left[\sigma_W^2\left[\int_0^1 W(t)^2\,dt - \left[\int_0^1 W(t)\,dt\right]^2\right]\right]^{-1}\\ &\times\sigma_V\sigma_W\left[\int_0^1 W(t)\,dV(t) - V(1)\int_0^1 W(t)\,dt\right].\end{aligned}$$

5. *Formulae for Theorem* 6.1: Define $w_t' = (u_t, \eta_t)$, the vector partial sum process $S_t = \sum_1^t w_j$ and the limiting covariance matrix

$$\Omega = \lim_{T\to\infty} T^{-1}E(S_T S_T'), \quad r \in [0,1].$$

According to the multivariate functional CLT of Phillips and Durlauf (1986) we have

$$\text{(A21)} \qquad T^{-1/2}S_{[Tr]} \Rightarrow B(r)$$

where $B(r)$ is 2-dimensional vector Brownian motion with covariance matrix Ω. We also define

$$\Lambda = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \sum_{j=1}^{t} E(w_j w_t') = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix}.$$

Then, using the weak convergence theory in Phillips (1986) and Phillips and Durlauf (1986) we obtain:

$$\text{(A22)} \qquad \text{Vol}' \Rightarrow \left\{ \left(\int_0^1 B_1 dB_2 + \lambda_{12} \right) - B_2(1) \int_0^1 B_1 \right\} \Big/ 12\sigma_u^2 - \left\{ \int_0^1 rB_1 - (1/2) \int_0^1 B_1 \right\} \left\{ (1/2) B_2(1) - \int_0^1 B_2 \right\} \Big/ \sigma_u^2.$$

The differences in (A22) and (A20) stem from the fact that forecast errors can Granger cause forecasts. (A20) treated the innovations as independent.

REFERENCES

AMEMIYA, T. (1973): "Generalized Least Squares with an Estimated Autocovariance Matrix," *Econometrica*, 41, 723–732.

ANDERSON, T. W. (1971): *The Statistical Analysis of Time Series*. New York: Wiley.

BILLINGSLEY, P. (1968): *Convergence of Probability Measures*. New York: Wiley.

DICKEY, D. A., AND W. A. FULLER (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 47, 427–431.

——— (1981): "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, 49, 1057–1072.

DURLAUF, S. N. (1986): "Essays in Econometrics and Macroeconomics," Doctoral Dissertation, Yale University.

DURLAUF, S. N., AND P. C. B. PHILLIPS (1986): "Trends versus Random Walks in Time Series Analysis," Cowles Foundation Discussion Paper No. 743.

ENGLE, R. F., AND C. W. J. GRANGER (1987): "Dynamic Model Specification with Equilibrium Constraints: Co-Integration and Error-Correction," *Econometrica*, 55, 251–276.

FLAVIN, M. (1983): "Excess Volatility in the Financial Markets: A Reassessment of the Empirical Evidence," *Journal of Political Economy*, 91, 929–956.

GRANGER, C. W. J., AND P. NEWBOLD (1974): "Spurious Regressions in Econometrics," *Journal of Econometrics*, 2, 111–120.

GRENANDER, U., AND M. ROSENBLATT (1957): *Statistical Analysis of Stationary Time Series*. New York: Wiley.

HOFFMAN, DENNIS L., STUART A. LOW, AND DON E. SCHLAGENHAUF (1984): "Tests of Rationality, Neutrality, and Market Efficiency: A Monte Carlo Analysis of Alternative Test Statistics," *Journal of Monetary Economics*, 14, 339–364.

KLEIDON, A. (1983): "Variance Bounds Tests and Stock Price Valuation Models," *Journal of Political Economy*, 94, 953–1001.

MANKIW, N. G., AND M. SHAPIRO (1985): "Trends, Random Walks and Tests of the Permanent Income Hypothesis," *Journal of Monetary Economics*, 16, 165–174.

——— (1986): "Do We Reject Too Often? Small Sample Bias in Tests of Rational Expectations Models," Cowles Foundation Discussion Paper No. 743, Yale University.

MARSH, T., AND R. C. MERTON (1983): "Aggregate Dividend Behavior and Its Implications for Stock Market Rationality," MIT Working Paper 1475–83.

NELSON C. R., AND H. KANG (1981): "Spurious Periodicity in Inappropriately Detrended Time Series," *Econometrica*, 49, 741–751.

——— (1983): "Pitfalls in the Use of Time as an Explanatory Variable in Regression," *Journal of Business and Economic Statistics*, 2, 73–82.

NELSON, C. R., AND C. PLOSSER (1982): "Trends and Random Walks in Macro-economic Time Series: Some Evidence and Implications," *Journal of Monetary Economics*, 10, 139–162.

PHILLIPS, P. C. B. (1986): "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33, 311–340.

——— (1987a): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

——— (1987b): "Asymptotic Expansions in Nonstationary Vector Autoregressions," *Econometric Theory*, 3, 45–68.

——— (1987c): "Towards a Unified Asymptotic Theory of Autoregression," *Biometrika*, 74, 535–547.

——— (1988): "Weak Convergence to the Matrix Stochastic Integral $\int_0^1 BdB'$," *Journal of Multivariate Analysis* (forthcoming).

PHILLIPS, P. C. B., AND S. N. DURLAUF (1986): "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473–495.

PHILLIPS, P. C. B., AND S. OULIARIS (1986): "Testing for Cointegration Using Principal Components Methods," Cowles Foundation Discussion Paper No. 809R.

——— (1987): "Asymptotic Properties of Residual Based Tests for Cointegration," Cowles Foundation Discussion Paper #847.

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

PLOSSER, C. I., G. W. SCHWERT, AND H. WHITE (1982): "Differencing as a Test of Specification," *International Economic Review*, 23, 535–552.

SARGAN, J. D., AND A. BHARGAVA (1983): "Testing Residuals from Least Squares Regression for Being Generated by a Gaussian Random Walk," *Econometrica*, 51, 153–174.

SHILLER, R. (1979): "The Volatility of Long Term Interest Rates and Expectations Models of the Term Structure," *Journal of Political Economy*, 87, 1190–1219.

——— (1981a): "Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?" *American Economic Review*, 7, 421–436.

——— (1981b): "The Use of Volatility Measures in Assessing Market Efficiency," *Journal of Finance*, 291–304.

STOCK, J. (1987): "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors," *Econometrica*, 55, 1035–1056.

WHITE, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 50, 817–838.

WHITE, H., AND I. DOMOWITZ (1984): "Nonlinear Regression with Dependent Observations," *Econometrica*, 52, 143–162.