Econometrica, Vol. 56, No. 6 (November, 1988), 1397–1417

ASYMPTOTIC NORMALITY, WHEN REGRESSORS HAVE A UNIT ROOT

BY KENNETH D. WEST[1]

Under fairly general conditions, ordinary least squares and linear instrumental variables estimators are asymptotically normal when a regression equation has nonstationary right-hand side variables. Standard formulas may be used to calculate a consistent estimate of the asymptotic variance covariance matrix of the estimated parameter vector, even if the disturbances are conditionally heteroskedastic and autocorrelated. So inference may proceed in the usual way. The key requirements are that the nonstationary variables share a common unit root and that the unconditional mean of their first differences is nonzero.

KEYWORDS: Unit root, nonstationary, random walk.

1. INTRODUCTION

LET y_t BE A SCALAR TIME SERIES whose first difference is stationary with a nonzero unconditional mean. This paper is concerned with ordinary least squares and linear instrumental variables estimation of models that can be written

$$w_t = X_{1t}'\alpha + \gamma y_t + e_t. \tag{1.1}$$

In (1.1), X_{1t} and e_t are stationary, with X_{1t} a $(k \times 1)$ vector of observable regressors and e_t an unobservable disturbance; the $(k \times 1)$ vector α and the scalar γ are unknown coefficients. The dependent variable w_t is stationary if $\gamma = 0$, nonstationary if not. In most applications of interest $\gamma \neq 0$ and w_t is nonstationary.

Equations of the form (1.1) occur first of all in atheoretical or reduced form time series regressions. One example is estimation of a unit autoregressive root (e.g., Dickey and Fuller (1979, 1981), Evans and Savin (1981, 1984), Fuller (1976), Said and Dickey (1984)). In this case, $y_t = w_{t-1}$, X_{1t} consists of a constant and, possibly, lags of Δw_t (but *not* a trend term, since X_{1t} is stationary), and e_t is the serially uncorrelated innovation in Δw_t. A second example occurs in estimation of the nonzero "cointegrating parameter" γ when the ARIMA process for (w_t, y_t) is assumed to be such that $w_t - \gamma y_t$ is stationary (e.g., Campbell (1987), Campbell and Shiller (1986), Engle and Granger (1987), Stock (1987)). Here, X_{1t} is either identically zero or is simply the constant term, and e_t is in general serially correlated.

Potentially of equal or greater importance than such reduced form regressions are regressions when an equation of the form (1.1) is implied by a structural macroeconomic model. With a little rearrangement, many, many macroeconomic

[1] I thank Christian Leger, Whitney Newey, an anonymous referee, and participants in seminars at U. C. Berkeley, the Federal Reserve Bank of Minneapolis, Princeton University, and Stanford University for helpful comments and discussion, and the National Science Foundation for financial support. Responsibility for remaining errors is mine. The initial version of this paper was written while I was a National Fellow at the Hoover Institution.

models can be written in the form (1.1). Consider, for example,

$$(1.2) \qquad w_t = X_{1t}^{*\prime}\alpha^* + \gamma_1 w_{t-1} + \sum_{i=1}^{n} \gamma_{i+1} y_{t-i} + \gamma_{n+2} y_t + e_t,$$

$$|\gamma_1| < 1, \text{ roots of } \gamma_{n+2} + \sum_{i=1}^{n} \gamma_{i+1} z^i \text{ outside unit circle,}$$

X_{1t}^* a vector of observed stationary regressors,

e_t an unobserved stationary disturbance.

Examples include: flexible accelerator models of investment, e.g., w_t is inventories, y_t is sales (Rowley and Trevedi (1975)); present value models of asset prices, e.g., w_t is stock price, y_t is dividend (West (1987)); dynamic linear rational expectations models, e.g., w_t is labor demand, y_t is real wage (Sargent (1978)). It is easy to see that a linear transformation of the regressors in (1.2) puts the equation in the form (1.1). One such transformation, for example, rewrites (1.2) as

$$(1.3) \qquad w_t = X_{1t}^{*\prime}\alpha^* + \gamma_1(w_{t-1} - \gamma y_t) + \sum_{i=1}^{n} \gamma_{i+1}(y_{t-i} - y_t) + \gamma y_t + e_t,$$

$$\equiv X_{1t}'\alpha + \gamma y_t + e_t, \qquad \gamma \equiv (1-\gamma_1)^{-1} \sum_{i=1}^{n+1} \gamma_{i+1}.$$

(Stationarity of $w_t - \gamma y_t$, and therefore of $w_{t-1} - \gamma y_t$, follows from subtracting $\gamma_1 w_t$ from each side of (1.3) and dividing by $(1-\gamma_1)$.) Since I will be concerned only with linear estimators, analysis of the properties of equation (1.1) will suffice to establish the properties of estimators of equation (1.2).

Now, macroeconomists have recently recognized that unit autoregressive roots may exist in many U.S. macroeconomic time series, including among others GNP (Nelson and Plosser (1982)), stock prices (Kleidon (1985)), real wages (Altonji and Ashenfelter (1980)), and personal disposable income (Mankiw and Shapiro (1985)). Standard theorems in time series, however, rule out unit roots (e.g., Hansen (1982)). This leaves open the question of how to interpret regressions with variables with unit roots.

Perhaps the most familiar environment in which this question has been answered is in estimation and test for cointegration and for a unit autoregressive root. With X_{1t} and e_t described as in the second paragraph in the text, this work has established the properties of the least squares estimate of γ, emphasizing the case when $0 = E\Delta y$. In this case, this estimate is such that $T^{1-\delta}(\hat{\gamma} - \gamma)$ converges in probability to zero for any $\delta > 0$. The limiting distribution of $T(\hat{\gamma} - \gamma)$ is not normal (Dickey and Fuller (1979, 1981), Engle and Granger (1987), Evans and Savin (1981), Fuller (1976), Said and Dickey (1984), Stock (1987)).

These results presumably carry over to structural models that imply an equation of the form (1.1). But they leave open the question of what happens if $0 \neq E\Delta y$, for either a reduced form or structural equation. Since most macroeconomic time series tend to drift (upwards) over time, it will often be reasonable to model the first differences of such series as having nonzero means.

This paper considers ordinary least squares and linear instrumental variables estimates of (1.1) when $0 \neq E\Delta y$. It does so under assumptions about X_{1t} and e_t more general than those of the papers cited above, including allowing e_t to be serially correlated and conditionally heteroskedastic, with the form of the correlation and heteroskedasticity unknown. It is shown that the estimate of γ is such that convergence is even more rapid than when $0 = E\Delta y$: $T^{(3/2)-\delta}(\hat{\gamma} - \gamma)$ converges in probability to zero for any $\delta > 0$. In addition, and more importantly, $T^{3/2}(\hat{\gamma} - \gamma)$ and $T^{1/2}(\hat{\alpha} - \alpha)$ are asymptotically normal. In models where e_t is iid, the usual computer output standard errors can be used for inference. And even in more general models, standard formulas may be used to calculate the variance covariance matrix of the estimated parameter vector. So inference about γ and α may proceed in a fashion that is robust to possible nonstationarity in y_t. A small Monte Carlo experiment suggests that when y_t is nonstationary, the asymptotic distribution can provide a useful approximation to the actual small sample distribution.

It may seem surprising that the asymptotic distributions of estimates of γ are so sensitive to whether the means of the first differences are zero. In a stationary environment, means in general do not affect the asymptotic distribution of estimates of parameters of economic interest. In theoretical econometric analyses it is therefore harmless to assume that means are zero if it is convenient to do so. A basic implication of this paper is that this is not at all true in the nonstationary environment considered in this paper.[2]

Section 2 of the paper analyzes ordinary least squares estimates of (1.1) when X_{1t} is simply the constant term. Section 3 is a compact generalization of Section 2 to allow for stochastic X_{1t}, which may be correlated with e_t. I have treated the case of X_{1t} being the constant term separately for two reasons. The first is to illustrate the basic ideas involved, in an environment where the algebra is relatively uncluttered. The second is because this special case is an important one, appearing in the empirical estimates in several of the papers cited above. Section 4 presents the results of a small Monte Carlo experiment. Section 5 has conclusions.

2. ESTIMATION, WHEN X_{1t} IS THE CONSTANT TERM

Along with the scalar α, the constant (possibly zero) unconditional mean of $w_t - \gamma y_t$, the scalar γ is in this section assumed to be estimated from the equation

$$w_t = \alpha + \gamma y_t + e_t. \tag{2.1}$$

[2] In least squares estimators of a unit autoregressive root, with iid disturbances, asymptotic normality when $E\Delta y \neq 0$ was noted in Dicker and Fuller (1979) and Evans and Savin (1981), and was emphasized in some papers that I was not aware of when this paper was first written (Fuller, Hasza, and Goebel (1981), Fuller (1985)).

This is precisely the environment in the "cointegration" papers of Campbell (1987), Campbell and Shiller (1986), and with the constant term suppressed, in the "cointegration" papers of Engle and Granger (1987) and Stock (1987). This equation is also appropriate if an investigator wants to check whether w_t has a unit root, but does not wish to model the Δw_t process parametrically (Said and Dickey (1984)). In this case, $y_t = w_{t-1}$, and, under the null hypothesis that $\gamma = 1$, $\alpha = E\Delta w$, $e_t = \Delta w_t - E\Delta w$.

ASSUMPTION 1: (1.a) *$\Delta y_t - E\Delta y_t$ is a regular, zero mean, strictly stationary stochastic process.* (1.b) $0 \neq \mu = E\Delta y_t$. (1.c) *For $t > 0$, $y_t = \sum_{s=1}^{t} \Delta y_s$, where the moving average representation of Δy is $\Delta y_s = \mu + \sum_{n=0}^{\infty} \theta_n \nu_{s-n}$, ν_s the innovation in Δy_s, $\theta_0 = 1$.*

ASSUMPTION 2: (a) *e_t is a regular, zero mean, strictly stationary stochastic process.* (b) *The spectral density of e_t is strictly positive at frequency zero.* (c) *Let I_t be the σ-algebra determined by e_s for $s \leqslant t$, v_j^2 the unconditional variance of $E(e_t | I_{t-j}) - E(e_t | I_{t-j-1})$. Then $\sum_{j=0}^{\infty} v_j < \infty$.*

ASSUMPTION 3: *Δy_t and e_t have finite fourth moments. In addition, second central moments and fourth cumulants of the $(\Delta y_t, e_t)$ process are absolutely summable. For the Δy_t process, this means:* (a) $\sum_{i=-\infty}^{\infty} |E(\Delta y_t - \mu)(\Delta y_{t+i} - \mu)| < \infty$; (b) $\sum_{ijk=-\infty}^{\infty} |\lambda(i, j, k)| < \infty$, *where* (*Hannan* (1970, *p.* 23))

$$\begin{aligned}\lambda(i, j, k) = {} & E(\Delta y_t - \mu)(\Delta y_{t+i} - \mu)(\Delta y_{t+j} - \mu)(\Delta y_{t+k} - \mu) \\ & - E(\Delta y_t - \mu)(\Delta y_{t+i} - \mu) E(\Delta y_{t+j} - \mu)(\Delta y_{t+k} - \mu) \\ & - E(\Delta y_t - \mu)(\Delta y_{t+j} - \mu) E(\Delta y_{t+i} - \mu)(\Delta y_{t+k} - \mu) \\ & - E(\Delta y_t - \mu)(\Delta y_{t+k} - \mu) E(\Delta y_{t+j} - \mu)(\Delta y_{t+i} - \mu).\end{aligned}$$

The infinite sums in (*a*) *and* (*b*) *are finite as well when one or more of the $\Delta y_s - \mu$ are replaced by e_s.*

As noted in the introduction, Assumption (1.b) is key to insuring asymptotic normality. This appears to be reasonable for many of the series that have been used when (2.1) was estimated, including, for example, consumption/income (Campbell (1987), Engle and Granger (1987), money/income (Engle and Granger (1987)) and stock prices/dividends (Campbell and Shiller (1986)). For other series, the assumption may not be appropriate. If not, the analyses in, e.g., Dickey and Fuller (1979, 1981), Engle and Granger (1987), Fuller (1976), Said and Dickey (1984), and Stock (1987) are applicable (although this paper provides a generalization that may be useful in some contexts). Assumption (1.c) simultaneously states that for $t > 0$ the Δy_t are drawn from the stationary distribution for Δy_t, and that the initial value of y_0 can be ignored in the analysis (e.g., because y_0 is nonstochastic). This allows inference to be unconditional. The assumption is maintained in muchof the work on estimation of unit autoregressive roots (e.g., Dickey and Fuller (1979, 1981), Fuller (1976), Said and Dickey (1984)).

Assumptions 2(a) and 2(c) are maintained because these are necessary to exploit a central limit theorem of Hannan (1973). The statement about the spectral density in Assumption (2.b) is made to guarantee that the asymptotic variance-covariance matrix is not identically zero (see below). If e_t follows a finite parameter ARMA process, the assumption rules out, for example, a factor of $(1-L)$ in the MA polynomial, but leaves the AR polynomial unconstrained. This restriction is of course commonly maintained in time series analysis.

Assumption 3 bounds the second and fourth moments of Δy_t and e_t. It is used in establishing not only asymptotic normality but even consistency of the estimator. Note that it does *not* constrain y_t and e_t, or Δy_t and e_t, to be uncorrelated. It is to be stressed that Assumption 3 is implied whenever the $(\Delta y_t, e_t)$ process has an absolutely summable moving average representation and iid innovations (Hannan (1970, p. 211)); this summability and independence assumption is maintained in Dickey and Fuller (1979, 1981), Engle and Granger (1987), Fuller (1976), Said and Dickey (1984), and Stock (1987). I maintain the more general Assumption 3 because it allows heteroskedasticity of the innovations in Δy_t and e_t conditional on their past values; independence does not. Such heteroskedasticity is often found in empirical work (e.g., Engle (1982)).

The (2×1) parameter vector $\beta = [\alpha\gamma]'$ is assumed to be estimated by OLS. Let $X_t = [1 y_t]'$, $X = [X_t']$, $W = [w_t]$. X is $(T\times 2)$, W is $(T\times 1)$, where T is the sample size. Then $\hat\beta$ is estimated as $\hat\beta = (X'X)^{-1}X'W$.

Let $C_T = \operatorname{diag}[T^{1/2}, (\Sigma t^2)^{1/2}]$ be a 2×2 diagonal matrix. The summations in $(\Sigma t^2)^{1/2}$ and throughout run over t, from 1 to T, unless otherwise noted. Then

$$C_T(\hat\beta-\beta) = \left[C_T^{-1}X'XC_T^{-1}\right]^{-1}C_T^{-1}X'e$$

$$= \begin{vmatrix} 1 & \sum y_t\Big/\left[T\sum t^2\right]^{1/2} \\ \sum y_t\Big/\left[T\sum t^2\right]^{1/2} & \sum y_t^2\Big/\left(\sum t^2\right) \end{vmatrix}^{-1} \begin{vmatrix} \sum e_t\Big/T^{1/2} \\ \sum y_t e_t\Big/\left(\sum t^2\right)^{1/2} \end{vmatrix},$$

where $e = [e_t]$ is $T\times 1$.

To establish asymptotic normality of $C_T(\hat\beta-\beta)$, it suffices to show that $C_T^{-1}X'e$ is asymptotically normal and that $C_T^{-1}X'XC_T^{-1}$ converges in probability to a nonsingular matrix of constants. Let

$$(2.2)\qquad R = \begin{vmatrix} 1 & \lim \sum t\Big/\left[T\sum t^2\right]^{1/2} \\ \lim \sum t\Big/\left[T\sum t^2\right]^{1/2} & 1 \end{vmatrix} = \begin{vmatrix} 1 & \sqrt{3}/2 \\ \sqrt{3}/2 & 1 \end{vmatrix},$$

$$M = \begin{vmatrix} 1 & 0 \\ 0 & \mu \end{vmatrix}, \qquad B = MRM = \begin{vmatrix} 1 & \mu(\sqrt{3}/2) \\ \mu(\sqrt{3}/2) & \mu^2 \end{vmatrix},$$

$$\sigma(h) = Ee_t e_{t-h}, \qquad s = \sum_{h=-\infty}^{\infty} \sigma(h).$$

(See Anderson (1971, p. 83) for $\lim \Sigma t/(T\Sigma t^2)^{1/2} = \sqrt{3}/2$.) s is of course the spectral density of the e_t process at frequency zero. So $s > 0$ by Assumption 2.

LEMMA 2.1: *Under Assumptions* 1 *to* 3, $C_T^{-1}X'e$ *converges in distribution to a* $N(0, sB)$ *random variable.*

To illustrate the role played by Assumption (1.b) in the analysis, I will sketch the proof of Lemma 2.1 here in the text, with some details relegated to the Appendix. Proofs of all other propositions are entirely in the Appendix.

Let ΔY_t be the indeterministic portion of the Δy_t process, $\Delta y_t = \mu + \Delta Y_t$. Since $y_t = \Delta y_1 + \Delta y_2 + \cdots + \Delta y_t$ by Assumption (1.c), $y_t = \mu t + (\Sigma_{s=1}^{t} \Delta Y_s)$. So

$$C_T^{-1}X'e = \left[T^{-1/2}\sum e_t, \left(\sum t^2\right)^{-1/2}\sum \mu t e_t + \left(\sum t^2\right)^{-1/2}\sum_{t=1}^{T}\left(\sum_{s=1}^{t}\Delta Y_s\right)e_t\right]'.$$

The Appendix proves that $\text{plim}[T^{-1-\delta}\Sigma_{t=1}^{T}(\Sigma_{s=1}^{t}\Delta Y_s)e_t] = 0$ for any $\delta > 0$, generalizing Stock's (1987) proof of this fact for the iid case. Since $(\Sigma t^2)^{-1/2}$ is of order $T^{-3/2}$, it follows that $\text{plim}[(\Sigma t^2)^{-1/2}\Sigma_{t=1}^{T}(\Sigma_{s=1}^{t}\Delta Y_s)e_t] = 0$. Therefore

$$\text{(2.3)} \qquad \text{plim}\, C_T^{-1}X'e = \text{plim}\left[T^{-1/2}\sum e_t, \left(\sum t^2\right)^{-1/2}\sum \mu t e_t\right]'.$$

It follows from the proof of Theorem 2(a) in Hannan (1973) that $[T^{-1/2}\Sigma e_t, (\Sigma t^2)^{-1/2}\Sigma t e_t]'$ converges in distribution to a $N(0, sR)$ random variable. So (2.3) implies quite trivially that $C_T^{-1}X'e$ converges in distribution to a $N(0, sB)$ random variable. *Q.E.D.*

Any correlation between y_t and e_t does not matter asymptotically, because y_t is dominated by its deterministic component: y_t figures into $\text{plim}\, C_T^{-1}X'e$ just as would a trend term of the form μt. It figures into $\text{plim}[C_T^{-1}X'XC_T^{-1}]$ the same way:

LEMMA 2.2: *Under Assumptions* 1 *to* 3, $[C_T^{-1}X'XC_T^{-1}]$ *converges in probability to* B.

Lemmas 2.1 and 2.2 combined imply the following theorem.

THEOREM 2.1: *Let* $\hat{\beta} = (X'X)^{-1}X'W$. *Then under Assumptions* 1 *to* 3, $C_T(\hat{\beta} - \beta)$ *converges in distribution to a* $N(0, sB^{-1})$ *random variable*, $B = \text{plim}[C_T^{-1}X'XC_T^{-1}]$.

Note that Lemmas 2.1 and 2.2, and therefore Theorem 2.1 as well, hinge crucially on $\mu = 0$. Note also that $\hat{\gamma} - \gamma$ is approaching its asymptotic distribution at the unusually rapid rate of $(\Sigma t^2)^{1/2}$, i.e., at rate $T^{3/2}$. So if, as in Campbell and Shiller (1986), the investigator wishes to use $w_t - \hat{\gamma} y_t$ in subsequent regressions in which all estimates are bounded in probability at the usual rate of $\sqrt{T}$, the uncertainty in $\hat{\gamma}$ may be ignored in conducting inference on the results of the subsequent regressions. Note, finally, that $\hat{\alpha}-\alpha$, by contrast, approaches its asymptotic distribution at the usual rate of $\sqrt{T}$.

For hypothesis testing and confidence intervals, a consistent estimate of sB^{-1} is required. B^{-1} is consistently estimated by $[C_T^{-1}X'XC_T^{-1}]^{-1}$. Estimation of s is also straightforward if the number of nonzero autocorrelations of s is known a priori; see Section 3. In practice, however, the number of nonzero autocorrelations of s may not be known a priori. This is in general the case when the investigator is attempting to estimate the cointegrating parameter for the (w_t, y_t) process. In such a case, the basic idea is to estimate $\sigma(h)$ from the moments of the OLS residuals for $h=-m,\dots,m$ and let $m\to\infty$ at a suitable rate. Let $\hat{e}_t$ be the OLS residual, $\hat{e}_t=w_t-X_t'\hat{\beta}$. Let $\hat{\sigma}(h)=T^{-1}\Sigma_{t=|h|+1}^{T}\hat{e}_t\hat{e}_{t+|h|}$, $k(h,m)=1-[|h|/(m+1)]$ for m a positive integer, $T>m\geqslant|h|$. As in Newey and West (1987), let

$$(2.4)\qquad \hat{s}=\sum_{h=-m}^{m}k(h,m)\hat{\sigma}(h)=\hat{\sigma}(0)+2\sum_{h=1}^{m}k(h,m)\hat{\sigma}(h).$$

The dependence of $\hat{s}$ on m is suppressed for notational simplicity. The weights $k(h,m)$ are present to insure that $\hat{s}\geqslant 0$. Other choices of weights also insure $\hat{s}\geqslant 0$; see Newey and West (1987).

LEMMA 2.3: *Let $\hat{s}$ be calculated as in* (2.4). *Under Assumptions* 1 *to* 3, *if $m\to\infty$ as $T\to\infty$ in such a way that m is $o(T^{1/2})$, then* $\operatorname{plim}\hat{s}=s$.

Lemma 2.3 gives a convenient way to calculate the usual finite sample approximation to the asymptotic variance-covariance matrix. A standard computer program calculates $\hat{\sigma}(0)[X'X]^{-1}$. Simply scale this by $\hat{s}/\hat{\sigma}(0)$. So the "t statistic" on H_0: $\gamma=\gamma_0$, for example, is calculated by dividing the computer output t statistic for this test by $[\hat{s}/\hat{\sigma}(0)]^{1/2}$.

3. ESTIMATION WITH GENERAL X_{1t}

The results from the previous section may be extended to linear models with stochastic X_{1t}. Consider equation (1.1), repeated below for convenience:

$$(3.1)\qquad w_t=X_{1t}'\alpha+\gamma y_t+e_t.$$

The model of the previous section is a special case of (3.1). Many structural macroeconomics models also imply (3.1); see the introduction for examples.

It is possible that X_{1t} will be correlated with e_t. So instrumental variables estimates may be required. To make the exposition concise, I will always assume that X_{1t} is instrumented by a $r\times 1$ $(r\geqslant k)$ vector Z_{1t} of variables uncorrelated with e_t, with $X_{1t}=Z_{1t}$ allowed.

ASSUMPTION 4: (4.a) $\{X_{1t}, Z_{1t}, \Delta y_t, e_t\}$ *is strictly stationary and ergodic, with absolutely summable second central moments.* (4.*b*) $EZ_{1t}Z_{1t}'$ *and the matrix* A *defined in Lemma* 3.2 *are of full row rank.*

ASSUMPTION 5: (5.a) *The* $\{Z_{1t}'e_t, e_t\}'$ *process is a regular, zero mean process.* (5.b) *The spectral density of* $\{Z_{it}'e_t, e_t\}'$ *evaluated at frequency zero is positive semidefinite; those of* $\{Z_{1t}'e_t\}'$ *and* e_t *are positive definite.* (5.c) *Let* I_t *be the* σ*-algebra determined by* $\{(Z_{1s}'e_s, e_s)' \mid s \leqslant t\}$. *Let* v_j *be the unconditional variance-covariance matrix of* $E[(Z_{1t}'e_t, e_t)' \mid I_{t-j}] - E[(Z_{1t}'e_t, e_t)' \mid I_{t-j-1}]$, *with* $\|v_j\|^2$ *its largest eigenvalue. Then* $\sum_{j=0}^{\infty}\|v_j\| < \infty$.

ASSUMPTION 6: (6.a) *Let* z_{it} *be an element of* Z_{1t}. *Then* $E(z_{it}e_t)^4 < \infty$, $i = 1, \ldots, r$. *In addition, the* $\{Z_{1t}'e_t, e_t\}'$ *process has absolutely summable second moments and fourth cumulants.* (6.b) *Let* z_{it} *be an element of* Z_{1t}, x_{nt} *an element of* X_{1t}. *Then* $Ez_{it}^4 < \infty$, $i = 1, \ldots, r$, *and* $Ex_{nt}^4 < \infty$, $n = 1, \ldots, k$.

Assumption 4 is new. It was automatically satisfied under last section's assumption that $X_{1t} = Z_{1t} =$ constant term. Assumption 5 replaces Assumption 2. Assumption 6 is also new. It is used only in establishing consistency of estimators of the variance covariance matrix, and not in establishing asymptotic normality of the estimators of the regression parameters. It may be worth remarking again that Assumption (6.a) is implied whenever $(Z_{1t}'e_t, e_t)'$ has an absolutely summable moving average representation and iid innovations with finite fourth moments. If the constant term is one of the Z_{1t}, variables, reference can be made to just $Z_{1t}'e_t$ rather than to $(Z_{1t}'e_t, e_t)'$ in Assumptions 5 and 6.

Let $\beta = (\alpha', \gamma)'$ be the $(k+1) \times 1$ parameter vector, X be the $T \times (k+1)$ matrix of regressors, with $[X_{1t}', y_t]$ in its tth row, Z be the $T \times (r+1)$ matrix of instruments, with $[Z_{1t}', y_t]$ in its tth row, $C_T = \operatorname{diag}[T^{1/2}, \ldots, T^{1/2}, (\Sigma t^2)^{1/2}]$ be a $(k+1) \times (k+1)$ diagonal matrix, $D_T = \operatorname{diag}[T^{1/2}, \ldots, T^{1/2}, (\Sigma t^2)^{1/2}]$ be a $(r+1) \times (r+1)$ diagonal matrix. Let $\sigma_{11}(h) = EZ_{1t}e_t(Z_{1t-h}e_{t-h})'$, $\sigma_{12}(h) = EZ_{1t}e_te_{t-h}$, $\sigma(h) = Ee_te_{t-h}$. Let

$$S_{11} = \sum_{h=-\infty}^{\infty} \sigma_{11}(h), \quad S_{12} = \sum_{h=-\infty}^{\infty} \sigma_{12}(h), \quad s = \sum_{h=-\infty}^{\infty} \sigma(h).$$

So S_{11} is the spectral density matrix of the $\{Z_{1t}'e_t\}'$ process, evaluated at frequency zero. S_{12} and s are the same quantity for, respectively, the off diagonal block of the $\{Z_{1t}'e_t, e_t\}'$ process and the e_t process.

LEMMA 3.1: *Under Assumptions* 1, 3, 4, *and* 5, $D_T^{-1}Z'e$ *converges in distribution to a* $N(0, S)$ *random variable, where* S *is the positive definite matrix*

$$S = \begin{vmatrix} S_{11} & (\sqrt{3}/2)\mu S_{12} \\ (\sqrt{3}/2)\mu S_{12}' & \mu^2 s \end{vmatrix}.$$

It may be helpful to write out S in the special case where $(Z_{1t}'e_t\, e_t)'$ is serially uncorrelated. This holds if, as in the stock price model analyzed below, e_t is an expectational error and e_{s-1} and Z_{1s} for $s \leqslant t$ are in the information set used in forming expectations. In such a case,

$$S = \begin{vmatrix} EZ_{1t}Z_{1t}'e_t^2 & (\sqrt{3}/2)\mu EZ_{1t}e_t^2 \\ (\sqrt{3}/2)\mu EZ_{1t}'e_t^2 & \mu^2\sigma(0) \end{vmatrix}. \tag{3.2}$$

The formula for this specific case was written out mainly to make clear something that holds in general: heteroskedasticity of e_t conditional on Z_{1t} is manifested in the usual way in the upper left hand block of S. But any possible dependence of e_t on past Δy_t's is not reflected in S:[3] y_t looks like a trend term, asymptotically, and the variance of e_t does not depend on t.

LEMMA 3.2: *Under Assumptions* 1, 3, 4, *and* 5,

$$A = \operatorname{plim}\left[C_T^{-1}X'ZD_T^{-1}\right] = \begin{vmatrix} EX_{1t}Z_{1t}' & (\sqrt{3}/2)\mu EX_{1t} \\ (\sqrt{3}/2)\mu EZ_{1t}' & \mu^2 \end{vmatrix},$$

$$B = \operatorname{plim}\left[D_T^{-1}Z'ZD_T^{-1}\right] = \begin{vmatrix} EZ_{1t}Z_{1t}' & (\sqrt{3}/2)\mu EZ_{1t} \\ (\sqrt{3}/2)\mu EZ_{1t'} & \mu^2 \end{vmatrix}.$$

B is positive definite.

THEOREM 3.1: *Let $\hat{\beta}$ be estimated by two stage least squares*

$$\hat{\beta} = \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1}X'Z(Z'Z)^{-1}Z'W.$$

Then under Assumptions 1, 3, 4, *and* 5, $C_T(\hat{\beta}-\beta)$ *converges in distribution to a*

$$N\left\{0, (AB^{-1}A')^{-1}AB^{-1}SB^{-1}A'(AB^{-1}A')^{-1}\right\}$$

random variable.

The formula for the asymptotic covariance matrix is especially simple if the common regression assumption of iid disturbances is made. Suppose that e_t is distributed independently of e_{s-1} and Z_{1s} for all $s \leqslant t$. Then the equation (3.2) formula for S reduces to $S = \sigma(0)B$, B defined in Lemma 3.2. The asymptotic variance covariance matrix of $C_T(\hat{\beta}-\beta)$ is $\sigma(0)[AB^{-1}A']^{-1}$, i.e., the asymptotic variance covariance matrix of the normalized estimator is as usual the variance of the disturbance times the plim of the normalized $[X'Z(Z'Z)^{-1}Z'X]^{-1}$ matrix. If $Z_{1t} = X_{1t}$ and OLS is run, $A = B$ and the asymptotic covariance is $\sigma(0)B^{-1}$, i.e., the variance of the disturbance times the plim of the normalized $(X'X)^{-1}$ matrix.

[3] Unless, of course, some past Δy_t's are elements of Z_{lt}.

To conduct inference, an estimate of S is required. As in the previous section, define $k(h, m) = 1 - [|h|/(m+1)]$ for $T > m \geqslant |h|$.

LEMMA 3.3: *Let* $\tilde{S} = \sum_{h=-m}^{m} k(h, m)\tilde{R}(h)$, *where*

$$\tilde{R}(h) = \sum_{t=h+1}^{T} (Z_{1t}'\hat{e}_t, y_t\hat{e}_t)'(Z_{1t-h}'\hat{e}_{t-h}, y_{t-h}\hat{e}_{t-h}) \quad (h \geqslant 0), \tag{3.3}$$

$$\tilde{R}(h) = \tilde{R}(-h)' \quad (h < 0),$$

and $\hat{e}_t$ *is a residual calculated using a* $\hat{\beta}$ *for which* $C_T(\hat{\beta} - \beta)$ *is bounded in probability. Let* $\hat{S} = D_T^{-1}\tilde{S}D_T^{-1}$. *Then under Assumptions* 1, 3, 4, *and* 6, *if* $m \to \infty$ *as* $T \to \infty$ *in such a way that* m *is* $o(T^{1/2})$, plim $\hat{S} = S$.

In Lemma 3.3, the dependence of $\tilde{S}$ on m is suppressed for notational convenience. The residual $\hat{e}_t$ used in computing $\tilde{S}$ may of course be calculated from the Theorem 3.1 estimator, using any instrument matrix that satisfies the assumptions of the theorem.

It is not difficult to show that if the number of nonzero autocorrelations of $(Z_{1t}'e_t e_t)'$ is known a priori to be, say, m^*, then, under Assumptions 1, 3, 4, and 6, plim $D_T^{-1}\tilde{S}D_T^{-1} = S$, where

$$\tilde{S} = \sum_{h=-m^*}^{m^*} \tilde{R}(h) = Z'\hat{\Omega}Z, \tag{3.4}$$

and $\hat{\Omega}$ is a band diagonal matrix with $\hat{e}_t\hat{e}_s$ in its (t, s) position for $0 \leqslant |t-s| \leqslant m^*$, zero elsewhere. Note that since $\tilde{S}$ as defined in equation (3.4) may not be positive definite in any given finite sample, it may be desirable to use the Lemma 3.3 estimator even when the number of nonzero autocorrelations is known a priori.

If $EZ_{1t}e_t(Z_{1t-h}e_{t-h})' = EZ_{1t}Z_{1t-h}'Ee_te_{t-h}$ and $EZ_{1t}e_te_{t-h} = EZ_{1t}Ee_te_{t-h}$ for all h, the formulas in Lemma 3.3 and equation (3.4) may be changed so that normalized cross products of instruments enter separately from sample moments of residuals. Since this will be convenient in the important special case when $m^* = 0$, it seems worth noting this as a separate proposition for that case.

LEMMA 3.4: *Suppose that* $S = \sigma(0)B$, B *defined in Lemma* 3.2. *Let* $\hat{\sigma}(0) = T^{-1}\sum\hat{e}_t^2$, $\hat{e}_t$ *the residual calculated using a* $\hat{\beta}$ *for which* $C_T(\hat{\beta} - \beta)$ *is bounded in probability. Let* $\tilde{S} = \hat{\sigma}(0)Z'Z$, $\hat{S} = D_T^{-1}\tilde{S}D_T^{-1}$. *Then under Assumptions* 1, 3, *and* 4, plim $\hat{S} = S$.

In the remainder of this section it will be useful to maintain the following assumption:

ASSUMPTION 7: $\tilde{S}$ *is an* $(r+1) \times (r+1)$ *matrix such that* plim $D_T^{-1}\tilde{S}D_T^{-1} = S$.

Candidates for $\tilde{S}$ include those defined in Lemmas 3.3 and 3.4, and equation (3.4).

LEMMA 3.5: *Let*

$$\tilde{V} = \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1} X'Z(Z'Z)^{-1}\tilde{S}(Z'Z)^{-1}Z'X \times \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1}.$$

Then under Assumptions 1, 3, 4, *and* 7, $C_T\tilde{V}C_T$ *is a consistent estimate of the covariance matrix of the Theorem* 3.1 *estimator*, $\operatorname{plim}[C_T\tilde{V}C_T] = (AB^{-1}A')^{-1}AB^{-1}SB^{-1}A'(AB^{-1}A')^{-1}$.

It is well known that in a stationary environment, the two stage least squares estimator is not optimal in general (Hansen (1982)); if the equation is overidentified $(r > k)$ and the assumptions of Lemma 3.4 do not hold, a two step estimator is preferable. This holds true here as well.

THEOREM 3.2: *Let* $\hat{\beta}$ *be calculated by two-step, two-stage least squares*, $\hat{\beta} = [X'Z\tilde{S}^{-1}Z'X]^{-1}X'Z\tilde{S}^{-1}Z'W$. *Then under Assumptions* 1, 3, 4, 5, *and* 7, $C_T(\hat{\beta} - \beta)$ *converges in distribution to a* $N\{0,(AS^{-1}A')^{-1}\}$ *random variable. The asymptotic variance covariance matrix is consistently estimated by* $C_T\tilde{V}C_T$, *where*

$$\tilde{V} = [X'Z\tilde{S}^{-1}Z'X]^{-1}.$$

THEOREM 3.3: *Let* $\hat{\beta}$ *be estimated as in Theorem* 3.2. *Let* $\hat{e} = W - X\hat{\beta}$ *be the corresponding vector of residuals. Then under Assumptions* 1, 3, 4, 5, *and* 7, $\hat{e}'Z\tilde{S}^{-1}Z'\hat{e}$ *converges in distribution to a* $\chi^2(r-k)$ *random variable.*

Theorem 3.3 establishes that Hansen's (1982) test of instrument residual orthogonality can be used as usual.

The following theorem establishes that the usual Wald test of linear restrictions yields legitimate test statistics.

THEOREM 3.4: *Consider testing* H_0: $G\beta - g_0 = 0$, *where* G *is a* $q \times (k+1)$ *matrix of known constants of rank* q *and* g_0 *is* $(q \times 1)$ *vector of known constants. Let* $\hat{\beta}$ *be estimated under the assumptions of Theorem* 3.1 *or* 3.2, *and let* $\tilde{V}$ *be such that* $C_T\tilde{V}C_T$ *is a consistent estimate of the asymptotic variance-covariance matrix of* $C_T(\hat{\beta} - \beta)$. *Then if the null hypothesis* $G\beta - g_0 = 0$ *is true,*

$$\left(G\hat{\beta} - g_0\right)'[G\tilde{V}G']^{-1}\left(G\hat{\beta} - g_0\right)$$

converges in distribution to a $\chi^2(q)$ *random variable.*

It was noted in the introduction that analysis of equation (1.1) makes it straightforward to analyze an equation that has multiple nonstationary regressors with a common unit root. This follows formally from Corollary 3.1.

COROLLARY 3.1: *Let P be $(k+1)\times(k+1)$, Q be $(r+1)\times(r+1)$, with P and Q nonsingular matrices of constants. Let $Z^* = ZQ^{-1}$, with* tth *row $Z_t^{*\prime}$, $X^* = XP^{-1}$, $\beta^* = P\beta$, so that the regression model can be rewritten $W = X^*\beta^* + e$. For $\hat{\beta}^*$ defined below, let $\hat{e}^* = W - X^*\hat{\beta}^*$. Let $\tilde{S}^* = \Sigma_{h=-m}^{m} k(h,m)\tilde{R}^*(h)$, $\tilde{R}^*(h) = \Sigma_{t=h+1}^{T}(Z_t^*\hat{e}_t^*)(Z_t^*\hat{e}_t^*)'$.*

(a) *Let*

$$\hat{\beta}^* = \left[X^{*\prime}Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}X^*\right]^{-1}X^{*\prime}Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}W,$$

$$\tilde{V}^* = \left[X^{*\prime}Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}X^*\right]^{-1}X^{*\prime}Z^*(Z^{*\prime}Z^*)^{-1}\tilde{S}^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}X^* \times\left[X^{*\prime}Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}X^*\right]^{-1}.$$

Then under Assumptions 1, 3, 4, 5, *and* 6, *if $m\to\infty$ as $T\to\infty$ such that m is $o(T^{1/2})$, $T^{1/2}(\beta^* - \beta^*)$ converges in distribution to a normally distributed random variable with covariance matrix* plim $T\tilde{V}^*$.

(b) *Let*

$$\hat{\beta}^* = [X^{*\prime}Z^*\tilde{S}^{-1}Z^{*\prime}X^*]^{-1}X^{*\prime}Z^*\tilde{S}^{-1}Z^{*\prime}W,$$

$$\tilde{V}^* = [X^{*\prime}Z^*\tilde{S}^{-1}Z^{*\prime}X^*]^{-1}.$$

Then under Assumptions 1, 3, 4, 5, *and* 6, *if $m\to\infty$ as $T\to\infty$ such that m is $o(T^{1/2})$, $T^{1/2}(\hat{\beta}^* - \beta^*)$ converges in distribution to a normally distributed random variable with covariance matrix* plim $T\tilde{V}^*$.

(c) *Let $\hat{\beta}^*$ be calculated as in (b). Then under Assumptions* 1, 3, 4, 5, *and* 6, *$\hat{e}^{*\prime}Z^{*\prime}\tilde{S}^{*-1}Z^{*\prime}\hat{e}^*$ converges in distribution to a $\chi^2(r-k)$ random variable.*

(d) *Let $\hat{\beta}^*$ and $\tilde{V}^*$ be calculated either as in (a) or (b). Let G be $q\times(k+1)$ matrix of known constants of rank q, g_0 a $q\times 1$ vector of known constants. Then if H_0: $G\beta^* - g_0 = 0$ is true, under Assumption* 1, 3, 4, 5, *and* 6, *$(G\hat{\beta}^* - g_0)'(G\tilde{V}^*G')^{-1}(G\hat{\beta}^* - g_0)$ converges in distribution to a $\chi^2(q)$ random variable.*

Suppose we are estimating $W = X^*\beta^* + e$, with an instrument matrix Z^*. Corollary 3.1 implies that if there are nonsingular P and Q that transform X^* and Z^* such that $X\equiv X^*P$ and $Z\equiv Z^*Q$ satisfy the relevant assumptions, least squares and linear instrumental variables estimation and inference on β^* can proceed as usual. While not stated explicitly in Corollary 3.1, the usual formulas for estimating $\tilde{V}^*$ apply when e_t is iid [e.g., if $X^* = Z^*$, $\tilde{V}^* = \hat{\sigma}(0)^*(X^{*\prime}X^*)^{-1}$, $\hat{\sigma}(0)^* = T^{-1}\Sigma e_t^{*2}$], or if the number of nonzero autocorrelations of $(Z_{1t}e_t', e_t)'$ is known a priori to be $m^* < \infty$ [$\tilde{S}^* = \Sigma_{h=-m^*}^{m^*}\tilde{R}^*(h)$].

It is not relevant that P and Q are not uniquely defined, and may depend on unknown parameters. All that matters is that it is known a priori that such P and Q exist. There may be multiple nonstationary instruments in Z^*. If so, the fact that $Z\equiv Z^*Q$ has only one such instrument means that Corollary 3.1 is applicable only if there are stationary linear combinations of the nonstationary instruments. That is, the nonstationary instruments must be cointegrated. Similarly, nonstationary regressors in X^* (if there is more than one) must be cointegrated. This is the case with a wide class of structural models; see equations (1.2) and

(1.3). If there is more than one nonstationary regressor, each regression coefficient has a nondegenerate limiting distribution when normalized by $T^{1/2}$.[4] Of course, a certain linear combination of regression coefficients has a nondegenerate limiting distribution when normalized by $(\Sigma t^2)^{1/2}$ (specifically, $(0\,0 \ldots 0\,1)P^{-1}(\beta^* - \beta^*) \equiv (\hat{\gamma} - \gamma)$). Note that if there is more than one nonstationary instrument then, in general, all such instruments must be uncorrelated with e_t for there to exist Q such that the resulting Z_{1t} is uncorrelated as well.

In connection with this last remark, a simple extension of these results is worth noting. Asymptotically equivalent estimators of $\hat{\beta}$, and consistent estimators of its covariance matrix, are obtained when the instrument vector is $[Z_{1t}', t]'$ or $[Z_{1t}', y_{t-j}]'$ instead of $[Z_{1t}', y_t]'$. Asymptotic equivalence of $[Z_{1t}', t]'$ is useful to note because this instrument vector may have better small sample properties; see Section 4. Asymptotic equivalence of $[Z_{1t}', y_{t-j}]'$ is useful to note in application of Corollary 3.1 to models with multiple nonstationary instruments. Suppose, for example, that $w_t = \beta_1^* + \beta_2^* y_{t-1} + \beta_3^* y_t + e_t$, and e_t is correlated with y_t but uncorrelated with y_{t-j} for $j > 0$. Let X^* be the $T \times 3$ matrix whose tth row is $[1, y_{t-1}, y_t]$). Corollary 3.1 will not apply directly for analysis of ordinary least squares estimates of this system: no matter what Q matrix one uses to transform X^*, $EZ_{1t}e_t \neq 0$ since $Ee_t y_t \neq 0$. Suppose instead that one estimates using $T \times 3$ instrument matrix Z^* whose tth row is, say $[1, y_{t-1}, y_{t-2}]$. Then $EZ_{1t}e_t = 0$ if Q transforms Z^* so that Z has, say, tth row $[Z_{1t}', y_{t-2}]$, with $Z_{1t}' = [1, y_{t-1} - y_{t-2}]$. Since using an instrument vector $[1, y_{t-1} - y_{t-2}, y_{t-2}]'$ is asymptotically equivalent to using $[1, y_{t-1} - y_{t-2}, y_t]'$, Corollary 3.1 may be applied, and one may conduct inference as usual on

$$\hat{\beta}^* = \left[X^{*\prime}Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}X^*\right]^{-1}X^{*\prime}Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}W.$$

4. MONTE CARLO EVIDENCE

This section describes the results of a small Monte Carlo experiment. The aim is to see whether the asymptotic approximation to the finite sample distribution can be useful for hypothesis testing and inference, and not to determine with a high degree of precision the small sample distribution of the estimators. It therefore was sufficient to use only a small number of parameter values, chosen for their empirical plausibility.

The model used in the experiment states that expected returns on a stock are constant:

$$w_t = bE_t(w_{t+1} + y_{t+1}) = (1+r)^{-1}E_t(w_{t+1} + y_{t+1}), \tag{4.1}$$

[4] Let β_i^* be the ith element of β^*. It follows from Corollary 3.1 that $T^{1/2}(\beta_i^* - \beta_i^*)$ has a limiting normal distribution (possibly degenerate, with zero variance), $i = 1, \ldots, k+1$. Since the $(k+1) \times (k+1)$ matrix plim $T\tilde{V}^*$ is positive semidefinite of rank k (see the proofs of Corollary 3.1 and Theorem 3.4), either k or $k+1$ of the regression coefficients have nondegenerate distributions (i.e., at most one of the diagonal elements of plim $T\tilde{V}^*$ is zero). It is straightforward to show that there are k such distributions if there is exactly one nonstationary regressor, $k+1$ if there is more than one such regressor. Incidentally, this $T^{1/2}$ convergence can also occur in models with multiple nonstationary regressors with zero drift. See Theorem 8.5.1 in Fuller (1976).

where w_t is the real stock price, y_t is the real dividend, b is the constant disc‹ rate, $0<b<1$, r the constant expected return, $r>0$, E_t the mathema expectation of the market. Equation (4.1) may be written as

$$(4.2)\qquad \begin{aligned} & w_t = r^{-1}y_t + r^{-1}E_t(\Delta w_{t+1} + \Delta y_{t+1}) \Rightarrow \\ & w_t = \alpha + \gamma y_t + e_t, \\ & \alpha = r^{-1}E(\Delta w_{t+1} + \Delta y_{t+1}), \quad \gamma = r^{-1}, \\ & e_t = r^{-1}[E_t(\Delta w_{t+1} + \Delta y_{t+1}) - E(\Delta w_{t+1} + \Delta y_{t+1})]. \end{aligned}$$

γ (and α) can be estimated from (4.2), if y_t satisfies the Section 2 assumpti This is done in Campbell and Shiller (1986).

Consider instead rewriting (4.1) as

$$(4.3)\qquad \begin{aligned} & w_t = b(w_{t+1} + y_{t+1}) + u_{t+1}, \\ & u_{t+1} = -bv_{t+1}, \\ & v_{t+1} = (w_{t+1} + y_{t+1}) - E_t(w_{t+1} + y_{t+1}) \\ & \qquad = (\Delta w_{t+1} + \Delta y_{t+1}) - E_t(\Delta w_{t+1} + \Delta y_{t+1}). \end{aligned}$$

The parameter b can be estimated from (4.3), if $w_{t+1} + y_{t+1}$ satisfies the Secti‹ assumptions for y_t. This is done in West (1987, 1988).

In the experiment, the (w_t, y_t) process was generated by

$$(4.4a)\qquad y_t = \mu + y_{t-1} + q_{t-1} + \eta_{1t},$$

$$(4.4b)\qquad q_t = \phi q_{t-1} + \eta_{2t},$$

$$(4.4c)\qquad \begin{aligned} & w_t = \alpha + \gamma y_t + e_t, \\ & \alpha = \mu b/(1-b)^2, \quad \gamma = b/(1-b), \\ & e_t = kq_t, \quad k = b/[(1-b)(1-b\phi)], \\ & E\eta_{1t}\eta_{2s} = 0 \quad \text{for all } t, s, \quad \eta_{1t} \sim N(0, \sigma^2); \quad \eta_{2t} \sim N(0, \sigma^2). \end{aligned}$$

The values of α, γ, and k written out above are such that $w_t = \Sigma_1^\infty b^j E_t y$ where $E_t y_{t+j} = E[y_{t+j}|1, y_{t-1}, w_{t-1}, \ldots, y_0, w_0]$. This simple parameterization chosen because it implies that $\Delta y_t \sim$ ARMA(1,1), $e_t \sim$ AR(1), both of which app to be reasonable characterizations of Shiller's (1981) long-term annual data the Standard and Poor's index.[5]

The parameters b, μ, ϕ, and σ were set to round numbers that were roug consistent with these data. The parameter b was set to .95, μ to .04. Two va of ϕ were tried, $\phi = .1$ and $\phi = .7$. The former is roughly consistent with sample autocorrelations of Δy_t, the latter with those of $w_t - \gamma y_t$ $(= w_t - 19$ Two values of σ were tried, $\sigma = .1$ and $\sigma = .2$. The latter is roughly consist

[5] This is not to say that 4.4(a) to 4.4(c) adequately describe these data: the empirical estimat‹ 4.4(a) are not consistent with those of 4.4(c) if, indeed, the model is the overly simple one presum‹ the experiment. See West (1987, 1988).

with the sample variance of Δy_t (using $\text{var}[\Delta y_t] = \sigma^2/[1-\phi^2] + \sigma^2$); the former was tried for comparison.

In sum, then, four sets of parameter values were tried: $\phi = .1$ and $\phi = .7$, $\sigma = .1$ and $\sigma = .2$, with $b = .95$ and $\mu = .04$ throughout. For each of the four sets of parameter values, the following was done 1000 times. Two vectors of 100 independent $N(0, \sigma^2)$ variables were drawn. These were the values of η_{1t} and η_{2t}. y_t, q_t, and w_t were generated by (4.4a), (4.4b), and (4.4c), with $y_0 = q_0 = 0$.[6] Equations (4.2) and (4.3) were estimated by OLS and IV. This produced four sets of estimates for each random draw and each set of parameter values. A constant and trend were the instruments used in IV estimation of (4.2), a trend the sole instrument used for (4.3).

For the two estimates of γ from (4.2), calculation was done of $\hat{\gamma} - 19$ divided by the estimated standard error of $\hat{\gamma}$. (19 is the true value of γ, since $b = .95$.) The standard error was computed as described in Section 2, using 9 lags of the autocovariances of $\hat{e}_t$ ($m = 9$, m defined as in Lemma 2.3; 9 is approximately $T^{1/2} = 10$). For the two estimates of b from (4.3), calculation was done of $\hat{b} - .95$ divided by the estimated standard error of $\hat{b}$. The usual standard error was used here. All four t statistics of course are asymptotically distributed as standard normal random variables.

Table IA has information on the empirical distributions of the t statistics for OLS estimation of (4.2), Table IB for IV estimation of (4.2), Table IC for OLS estimation of (4.3), Table ID for IV estimation of (4.3). To illustrate how to read the tables, consider the entry in the first line of Table IA. The "Percentile" columns say that of the 1000 sets of w_t and y_t generated with $\phi = .1$ and $\sigma = .1$, OLS estimates of (4.2) yielded a t statistic less than -2.81 in 25 samples, less than -2.50 in 50 samples, ..., less than 1.92 in 975 samples. The "Size of a nominal five percent test" column says that when $\hat{\gamma} - 19$ was divided by its estimated standard error, the resulting statistic was greater than 1.65 in about 3 percent of the samples, less than -1.65 in about 15 percent of the samples, and greater than 1.96 in absolute value in about 13 percent of the samples. The comparable values for a standard normal variable, which are reproduced for convenience at the foot of the table, are -1.96, $-1.65, \ldots, 1.96$ for "Percentile," .05, .05, .05 for "Size."

A comparison of the entries in the tables to those for a standard normal variable suggests that t statistics from OLS estimates of (4.3) (Table IC) are sharply biased downwards, those from OLS estimates of (4.2) (Table IA) and IV estimates of (4.3) (Table ID) are slightly biased downwards, those from IV estimation of (4.2) have no noticeable bias. Actual sizes of tests, then, appear unlikely to equal nominal sizes. The divergence between actual and nominal size is probably larger for one tailed than for two tailed tests. The downward bias of

[6] Strictly speaking, to be consistent with Assumption 1.c, q_0 should have been drawn from its steady state $N[0, \sigma^2/(1-\phi^2)]$ distribution. Stock's (1987) results for iid innovations can, however, be combined with the argument of Lemma 2.1 to show that the OLS and IV estimates generated with $q_0 = 0$ are still asymptotically normal.

TABLE I

EMPIRICAL DISTRIBUTION OF t STATISTICS

Table IA

OLS, Equation (4.2)

		Percentile					Size of nominal 5% test		
ϕ	σ	2.5	5	50	95	97.5	$\gamma < 19$	$\gamma > 19$	$\gamma = 19$
.1	.1	−2.81	−2.50	−.50	1.41	1.92	.03	.15	.13
.1	.2	−3.13	−2.65	−.81	1.08	1.33	.01	.21	.15
.7	.1	−2.31	−1.89	−.31	1.56	1.96	.04	.08	.07
.7	.2	−2.29	−1.90	−.35	1.37	1.85	.03	.08	.06

Table IB

IV, Equation (4.2)

		Percentile					Size of nominal 5% test		
ϕ	σ	2.5	5	50	95	97.5	$\gamma < 19$	$\gamma > 19$	$\gamma = 19$
.1	.1	−2.23	−2.00	−.01	1.87	2.26	.07	.07	.10
.1	.2	−1.95	−1.72	−.00	1.72	1.99	.06	.06	.05
.7	.1	−2.22	−1.90	.00	1.78	2.17	.07	.08	.08
.7	.2	−2.24	−1.86	.04	1.83	2.19	.06	.07	.08

Table IC

OLS, Equation (4.3)

		Percentile					Size of nominal 5% test		
ϕ	σ	2.5	5	50	95	97.5	$b < .95$	$b > .95$	$b = .95$
.1	.1	−2.18	−1.86	−.31	1.20	1.46	.01	.07	.05
.1	.2	−2.62	−2.39	−.75	.38	.49	.00	.19	.12
.7	.1	−2.39	−2.15	−.72	.37	.49	.00	.15	.09
.7	.2	−2.74	−2.53	−1.20	−.15	−.03	.00	.27	.16

Table ID

IV, Equation (4.3)

		Percentile					Size of nominal 5% test		
ϕ	σ	2.5	5	50	95	97.5	$b < .95$	$b > .95$	$b = .95$
.1	.1	−2.03	−1.65	−.02	1.64	1.90	.05	.05	.05
.1	.2	−2.19	−1.91	−.14	1.40	1.66	.03	.07	.06
.7	.1	−2.09	−1.83	−.21	1.23	1.48	.01	.07	.04
.7	.2	−2.40	−2.06	−.40	1.09	1.33	.01	.10	.07

Note: Distributions are from 1000 replications with a sample size of 100, with $b = .95$, $\mu = .04$. Figures for a standard normal random variable:

Percentile 2.5	5	50	Size of nominal 5% test 95	97.5	< > =
−1.96	−1.65	0	1.65	1.96	.05 .05 .05

course means that OLS estimates of γ (from (4.2)) and IV and OLS estimates of b (from (4.3)) tend to be low. This is especially true of OLS estimates of b (Table IC).[7]

[7] Incidentally, while the Table I t statistics for OLS were in general less satisfactory than those for IV, the OLS point estimates were generally more satisfactory (i.e., closer to the true parameter value). So if one's only aim is to obtain a point estimate of b or γ to use in subsequent regressions, the OLS estimator appears to be preferable.

While the asymptotic normal approximation is far from perfect, Table I suggests that inference using the approximation is unlikely to be terribly misleading. Indeed, the approximation seems more accurate here than in some stationary environments (Evans and Savin (1984)). The adequacy of the approximation seems especially reassuring since the data in the experiment were fairly noisy. The population ratio of the mean to the standard deviation of Δy_t, for example, ranged from about .1 to about .3; an arbitrarily small value of this ratio of course results in a small sample distribution of $\hat{\gamma}$ arbitrarily similar to the nonnormal distribution that results when the ratio is exactly zero. In Monte Carlo simulations using less noisy data (e.g., quarterly aggregate U.S. consumption data, for which this ratio is about .6), the asymptotic normal approximation appears to be even more accurate (Stock and West (1988)).

CONCLUSION

The basic result of this paper is that under fairly general circumstances ordinary least squares and linear instrumental variables estimators of a model that can be transformed to have a single nonstationary right hand side variable are asymptotically normal. This applies even if the disturbance is conditionally heteroskedastic and serially correlated, with the heteroskedasticity and serial correlation of unknown form. The usual formulas for calculating the variance covariance matrix are correct. So hypothesis tests and confidence intervals are easily constructed, and test statistics yield asymptotically correct inferences whether or not the regressor is nonstationary. This applies not only to tests about the coefficient on the nonstationary variable, but to hypotheses about the model as a whole. In estimation and test of such a model, there is no need to take special steps to handle the assumed nonstationarity. See West (1987, 1988) for illustrations of these points.

Useful extensions of the present paper's environment include allowing for systems with multiple equations, multiple unit roots, and with nonstationary variables whose unconditional mean is zero. OLS estimators of such systems are studied in Phillips and Park (1986) and Sims, Stock, and Watson (1986). See Phillips and Park (1986) for systems in which none of the regressors are cointegrated, Sims, Stock, and Watson (1986) for systems with disturbances that are conditionally homoskedastic and serially uncorrelated.

Woodrow Wilson School, Princeton University, Princeton, NJ 08544, U.S.A.

Manuscript received May, 1986; final revision received September, 1987.

APPENDIX

The Appendix of the working paper version of this paper has more detailed versions of the proofs. The following notation is used below: $\Delta Y_t = \Delta y_t - \mu$ is the indeterministic portion of the Δy_t process; $Y_t = y_t - t\mu = \Sigma_{s=1}^{t} \Delta Y_s$ is the indeterministic portion of the y_t process; E, var, cov are unconditional expectations, variances, and covariances; "$d_T \overset{A}{\sim} N(0, G)$" means "the random variable D_T converges in distribution to a $N(0, G)$ random variable;" unless otherwise noted, all summations run over t from 1 to T; all limits and probability limits are taken as the sample size $T \to \infty$. Whenever it is necessary to show that a random variable, say, d_T, converges in probability to a constant, say, d, this is done by showing that $\lim Ed_T = d$, $\lim \text{var}(d_T) = 0$.

The proofs to Lemmas A.1 to A.3 may be found in the working paper version of this paper.

LEMMA A.1: *Let z_t be a zero mean, fourth order stationary random variable such that the $(\Delta Y_t, z_t)$ process has absolutely summable second central moments and fourth cumulants. Then* $\lim E(T^{-1-\delta}\Sigma Y_t z_t) = \lim \text{var}(T^{-1-\delta}\Sigma Y_t z_t) = 0$ *for any* $\delta > 0$.

LEMMA A.2: *Let ΔY_t have absolutely summable second central moments and fourth cumulants. Then* $\text{plim}\, T^{-2-\delta}\Sigma Y_t = 0$ *for any* $\delta > 0$.

LEMMA A.3: *Let z_t be stationary and ergodic, with absolutely summable second central moments, and let the Δy_t process have absolutely summable second central moments and fourth cumulants. Then* $\text{plim}(T\Sigma t^2)^{-1/2}\Sigma y_t z_t = (\sqrt{3}/2)\mu E z_t$.

LEMMA 2.1: Special case of Lemma 3.1.

LEMMA 2.2: What must be shown is that (a) $\text{plim}\Sigma y_t/[T\Sigma t^2]^{1/2} = \mu(\sqrt{3}/2)$, and (b) $\text{plim}\Sigma y_t/\Sigma t^2 = \mu^2$. Point (a) follows from Lemma A.3, with z_t equal to the constant term. Point (b): We have

$$y_t^2 = (\mu t + Y_t)^2 = \mu^2 t^2 + 2\mu t Y_t + Y_t^2 \Rightarrow \sum y_t^2 = \mu^2 \sum t^2 + 2\mu \sum t Y_t + \sum Y_t^2.$$

The limit of $(\Sigma t^2)^{-1}$ times the first term is μ^2. The expectation of the second term is zero; it is straightforward to show that since $\Sigma_{j=0}^{\infty}|E\Delta Y_t \Delta Y_{t-j}| < \infty$, $\lim \text{var}[(\Sigma t^2)^{-1}\Sigma t Y_t] = 0 \Rightarrow [(\Sigma t^2)^{-1} 2\mu \Sigma t Y_t] \xrightarrow{p} 0$. By Lemma A.2, $[(\Sigma t^2)^{-1}\Sigma Y_t^2] \xrightarrow{p} 0$.

THEOREM 2.1: Special case of Theorem 3.1.

LEMMA 2.3: Special case of Lemma 3.3.

LEMMA 3.1: $D_T^{-1}Z'e = [T^{-1/2}\Sigma Z_{1t}'e_t,\ (\Sigma t^2)^{-1/2}\Sigma y_t e_t]' = [T^{-1/2}\Sigma Z_{1t}'e_t, (\Sigma t^2)^{-1/2}\Sigma \mu t e_t + (\Sigma t^2)^{-1/2}\Sigma Y_t e_t]'$. But

$$\text{plim}\left(\sum t^2\right)^{-1/2}\sum Y_t e_t = 0$$

by Lemma A.1. So

$$\text{plim}\ D_T^{-1}Z'e = \text{plim}\left[T^{-1/2}\sum Z_{1t}'e_t, \left(\sum t^2\right)^{-1/2}\sum \mu t e_t\right]'.$$

Asymptotic normality of this last random vector follows directly from the proof of Theorem 2(a) in Hannan (1973).

To show that S is positive definite, I need to show $c'Sc > 0$ for any nonzero $(r+1)\times 1$ vector c. Let $c = [c_1' c_2]'$ where c_2 is a scalar. Since S_{11} is positive definite by Assumption 5.b, $c'Sc > 0$ if $c_2 = 0$. So assume $c_2 \neq 0$. Let K be the $(r+1)\times(r+1)$ spectral density matrix of $\{Z_{1t}'e_t, e_t\}'$ evaluated at frequency zero, which is positive semidefinite by Assumption 5.b. Then $c'Sc = [c_1'(\sqrt{3}/2)\mu c_2]K[c_1'(\sqrt{3}/2)\mu c_2]' + s(1/4)\mu^2 c_2^2 > 0$ since the first factor is nonnegative and the second is strictly positive.

LEMMA 3.2: By definition

$$C_T^{-1}X'ZD_T^{-1} = \begin{vmatrix} T^{-1}\sum X_{1t}Z_{1t}' & \left(T\sum t^2\right)^{-1/2}\sum X_{1t}y_t \\ \left(T\sum t^2\right)^{-1/2}\sum y_t Z_{1t}' & \left(\sum t^2\right)^{-1}\sum y_t^2 \end{vmatrix}.$$

Assumption 4.a implies that $\text{plim}\ T^{-1}\Sigma X_{1t}Z_{1t}' = EX_{1t}Z_{1t}'$. Lemma A.3 implies that $\text{plim}(T\Sigma t^2)^{-1/2}\Sigma y_t Z_{1t}' = (\sqrt{3}/2)\mu EZ_{1t}'$, $\text{plim}(T\Sigma t^2)^{-1/2}\Sigma X_{1t}y_t = (\sqrt{3}/2)\mu EX_{1t}$. Lemma 2.2 shows that $\text{plim}(\Sigma t^2)^{-1}\Sigma y_t^2 = \mu^2$. This establishes that $\text{plim}\, C_T^{-1}X'ZD_T^{-1} = A$. A similar argument establishes that $\text{plim}\ D_T^{-1}Z'ZD_T^{-1} = B$.

B may be shown to be positive definite by an argument like that used in the proof of Lemma 3.1 to show that S is positive definite.

THEOREM 3.1: Let $A(T) = C_T^{-1}X'ZD_T^{-1}$, $B(T) = D_T^{-1}Z'ZD_T^{-1}$. Then $C_T(\hat{\beta} - \beta) = [A(T)B(T)^{-1}A(T)']^{-1}A(T)B(T)^{-1}D_T^{-1}Z'e$. The result now follows from Lemmas 3.1 and 3.2.

LEMMA 3.3: It is necessary to show that $\hat{S}_{ij}$, the (i, j) element of $\hat{S}$, converges in probability to S_{ij}, the (i, j) element of S for $1 \leqslant i, j \leqslant r+1$. Let z_{it} be the ith element of the instrument vector, with $z_{r+1,t} = y_t$. Then $\hat{S}_{ij}$ is

$$f_T^{-1}\left\{\sum_{t=1}^{T} z_{it}z_{jt}\hat{e}_t^2 + \sum_{h=1}^{m} k(h,m)\left[\sum_{t=h+1}^{T}(z_{it}z_{jt-h} + z_{it-h}z_{jt})\hat{e}_t\hat{e}_{t-h}\right]\right\}$$

$$= f_T^{-1}\left\{\sum_{t=1}^{T} z_{it}z_{jt}(\hat{e}_t^2 - e_t^2)\right.$$

$$\left. + \sum_{h=1}^{m} k(h,m)\left[\sum_{t=h+1}^{T}(z_{it}z_{jt-h} + z_{it-h}z_{jt})(\hat{e}_t\hat{e}_{t-h} - e_te_{t-h})\right]\right\}$$

$$+ f_T^{-1}\left\{\sum_{t=1}^{T} z_{it}z_{jt}e_t^2 + \sum_{h=1}^{m} k(h,m)\left[\sum_{t=h+1}^{T}(z_{it}z_{jt-h} + z_{it-h}z_{jt})e_te_{t-h}\right]\right\}$$

where (a) $f_T = T$ if $1 \leqslant i,\ j \leqslant r$, (b) $f_T = \Sigma t^2$ if $i = r+1 = j$, (c) $f_T = (T\Sigma t^2)^{1/2}$ if $i = r+1 \neq j$ or $i \neq r+1 = j$. By considering cases (a), (b), and (c), separately, it is straightforward though extremely tedious to show that the first term in braces after the equality converges in probability to zero, the second to S_{ij}. The details are in the Appendix to the working paper version of this paper.

LEMMA 3.4: By Lemma 3.2, $D_T^{-1}Z'ZD_T^{-1} \xrightarrow{p} B \Rightarrow$ it suffices to show $\hat{\sigma}(0) \xrightarrow{p} \sigma(0)$. Let $\hat{e} = W - X\hat{\beta}$. We have

$$\sigma(0) = T^{-1}\hat{e}'\hat{e} = T^{-1}(W - X\hat{\beta})'(W - X\hat{\beta}) = T^{-1}[e - X(\hat{\beta} - \beta)]'[e - X(\hat{\beta} - \beta)]$$

$$= T^{-1}e'e - 2[T^{-1}X'e]'(\hat{\beta} - \beta) + (\hat{\beta} - \beta)'(T^{-1}X'X)(\hat{\beta} - \beta)$$

$$= T^{-1}e'e - 2[T^{-1}C_T^{-1}X'e]'C_T(\hat{\beta} - \beta)$$

$$+ C_T(\hat{\beta} - \beta)'[T^{-1}C_T^{-1}X'XC_T^{-1}]C_T(\hat{\beta} - \beta).$$

The first term converges in probability to $\sigma(0)$, by ergodicity. We have $T^{-1}C_T^{-1}X'e \xrightarrow{p} 0$ since $T^{-1/2}[T^{-1}\Sigma X_{1t}e_t] \xrightarrow{p} 0$ by ergodicity and $T^{-1}[(\Sigma t^2)^{-1/2}\Sigma y_te_t] \xrightarrow{p} 0$ by Hannan (1973) and Lemma A.2. Arguments similar to those in Lemma 3.2 can be used to establish that $C_T^{-1}X'XC_T^{-1}$ converges in probability to a matrix of constants $\Rightarrow T^{-1}[C_T^{-1}X'XC_T^{-1}] \xrightarrow{p} 0$. Since $C_T(\hat{\beta} - \beta)$ is bounded in probability, the second and third terms in the last line of the equation for $\hat{\sigma}(0)$ converge in probability to zero.

LEMMA 3.5: Follows immediately from Lemma 3.2 and Theorem 3.1.

THEOREM 3.2: Let $A(T) = C_T^{-1}X'ZD_T^{-1}$, $B(T) = D_T^{-1}Z'ZD_T^{-1}$. Then $C_T(\hat{\beta} - \beta) = [A(T)D_T\tilde{S}^{-1}D_TA(T)']^{-1}A(T)D_T\tilde{S}^{-1}D_T(D_T^{-1}Z'e)$. Since $D_T^{-1}\tilde{S}D_T^{-1}$ converges in probability to a nonsingular matrix by Assumption 7, and D_T is nonsingular, $D_T\tilde{S}^{-1}D_T$ converges in probability to $\operatorname{plim}(D_T^{-1}\tilde{S}D_T^{-1})^{-1} = S^{-1}$. The result now follows from Lemmas 3.1 and 3.2.

THEOREM 3.3: Let $A(T) = C_T^{-1}X'ZD_T^{-1}$, $\tilde{V} = [X'Z\tilde{S}^{-1}Z'X]^{-1}$. We have

$$\hat{e} = e - X(\hat{\beta} - \beta) = e - X\tilde{V}X'Z\tilde{S}^{-1}Z'e$$

$$\Rightarrow Z'\hat{e} = [I - Z'X\tilde{V}X'Z\tilde{S}^{-1}]Z'e$$

$$\Rightarrow D_T^{-1}Z'\hat{e} = [I - A(T)'C_T\tilde{V}C_TA(T)D_T\tilde{S}^{-1}D_T]D_T^{-1}Z'e.$$

So $D_T^{-1}Z'\hat{e}$ converges in distribution to a $N(0, Q)$ random variable,

$$Q = [I - A'VAS^{-1}]S[I - A'VAS^{-1}]' = S - A'VA,$$

$$V = \text{plim}\, C_T\tilde{V}C_T = \left(AS^{-1}A'\right)^{-1}.$$

Let $\tilde{S}^{-1/2}$ be such that $\text{plim}\, \tilde{S}^{-1/2}D_T = S^{-1/2}$, where $(S^{-1/2})'S^{-1/2} = S^{-1}$. Then $(\tilde{S}^{-1/2}D_T)D_T^{-1}$. is asymptotically normal with variance-covariance matrix $I - S^{-1/2}A'VA(S^{-1/2})'$. Since this cov ance matrix is idempotent and of rank $(r-k)$ (see Hansen (1982)), the result follows.

THEOREM 3.4: This will first be shown for two cases, which are not necessarily mutually exclusi It will then be shown that the assumptions of the Theorem imply that all tests fall in at least one the two cases.

Case (a): The rank of the first k columns of G is q. Partition $\tilde{V}$ so that $C_T\tilde{V}C_T$ can be writ

$$C_T\tilde{V}C_T = \begin{vmatrix} T\tilde{V}_{11} & \left(T\sum t^2\right)^{1/2}\tilde{V}_{12} \\ \left(T\sum t^2\right)^{1/2}\tilde{V}_{12}' & \left(\sum t^2\right)\tilde{v}_{22} \end{vmatrix}.$$

Note that because $C_T\tilde{V}C_T$ converges in probability to a matrix of constants, $\text{plim}(T\tilde{V}_{12})$ $\text{plim}(T\tilde{v}_{22}) = 0$. Define the $(k+1)\times(k+1)$ matrix Q as

$$Q = \begin{vmatrix} \text{plim}\left(T\tilde{V}_{11}\right) & 0 \\ 0 & 0 \end{vmatrix}.$$

So $\sqrt{T}(\hat{\beta} - \beta)$ is asymptotically normal with variance covariance matrix Q (Rao (1973, p. 122, $x($

Under the null hypothesis, $\sqrt{T}(G\hat{\beta} - g_0) = G\sqrt{T}(\hat{\beta} - \beta) \Rightarrow \text{plim}[\sqrt{T}(G\hat{\beta} - g_0)] = G\,\text{plim}[\sqrt{T}(\hat{\beta} -$ $\overset{A}{\sim} N(0, GQG')$. GQG' is nonsingular since $\text{plim}(T\tilde{V}_{11})$ and the first q columns of G each have rank So $[\sqrt{T}(G\hat{\beta} - g_0)]'[GQG']^{-1}[\sqrt{T}(G\hat{\beta} - g_0)]$ *is asymptotically* $\chi^2(q)$, as is $[\sqrt{T}(G\hat{\beta} - g_0)]'[TP]^{-1}[\sqrt{T}($ $- g_0)]$ for any matrix P such that $\text{plim}[TP - GQG'] = 0$ (White (1984, Corollary 4.24 and Theor 4.30)). It is easy to verify that this is satisfied in particular for $P = G\tilde{V}G'$, since $\text{plim}(T\tilde{V}_{12})$ $\text{plim}(T\tilde{v}_{22}) = 0$.

Case (b): There is a nonsingular $(q\times q)$ matrix R such that RG has zeros in (i) row q, column to k, and (ii) column $k+1$, rows 1 to $q-1$.

Let $F_T = \text{diag}[T^{1/2},\ldots,T^{1/2},(\sum t^2)^{1/2}]$ be $q\times q$ (so, e.g., $q = k+1 \rightarrow F_T = C_T$). Then $F_T(RG)$ $(RG)C_T$ by the assumed diagonal structure of RG. So

$$\begin{aligned} \text{plim}\left[F_TR\left(G\hat{\beta} - g_0\right)\right] &= \text{plim}\, F_TRG\left[\left(\hat{\beta} - \beta\right)\right] \\ &= RG\left[\text{plim}\, C_T\left(\hat{\beta} - \beta\right)\right] \overset{A}{\sim} N\left[0, RG\left(\text{plim}\, C_T\tilde{V}C_T\right)G'R'\right] \\ &\Rightarrow \left[F_TR\left(G\hat{\beta} - g_0\right)\right]'\left[RGC_T\tilde{V}C_TG'R'\right]^{-1}\left[F_TR\left(G\hat{\beta} - g_0\right)\right] \end{aligned}$$

is asymptotically $\chi^2(q)$. But

$$\begin{aligned} &\left[F_TR\left(G\hat{\beta} - g_0\right)\right]'\left[RGC_T\tilde{V}C_TG'R'\right]^{-1}\left[F_TR\left(G\hat{\beta} - g_0\right)\right] \\ &\quad = \left(G\hat{\beta} - g_0\right)'\left[G\tilde{V}G'\right]^{-1}\left(G\hat{\beta} - g_0\right). \end{aligned}$$

To show that all tests fall in at least one of the two cases: Suppose that the rank of $G(\alpha)$, t $(q\times k)$ matrix consisting of the first k columns of G, is less than q. There is then one row in $G($ say, the last row, that is a linear combination of the other $(q-1)$ rows. Then there exists nonsingular $(q\times q)$ matrix, say, H, such that HG has zeros in row q, columns 1 to k, and, since G of rank q, a nonzero number in row q, column $k+1$. Given this, there is a nonsingular $(q\times$ matrix, say, J, such that the last row and column of JHG have zeros everywhere except for row column $k+1$. $R = JH$ is then a matrix that satisfies case (b).

COROLLARY 3.1: (a) As in Theorem 3.1, let $\hat{\beta} = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'W$. Sin $\hat{\beta}^* = P\hat{\beta}$, $\hat{e}_t^* = \hat{e}_t \Rightarrow \tilde{V}^* = P\tilde{V}P'$. Since $T^{1/2}(\hat{\beta} - \beta) \overset{A}{\sim} N(0, \text{plim}\, T\tilde{V})$ (see the proof of Theorem 3. $T^{1/2}(\hat{\beta}^* - \beta^*) \overset{A}{\sim} N(0, \text{plim}\, T\tilde{V}^*)$.

(b) Similar to the proof of (a).

(c) This statistic is algebraically identical to the Theorem 3.3 statistic.

(d) This statistic is algebraically identical to the Theorem 3.4 statistic for H_0: $GP\beta - g_0 = 0$.

REFERENCES

ALTONJI, J., AND O. ASHENFELTER (1980): "Wage Movements and the Labour Market Equilibrium Hypothesis," *Economica*, 47, 217–245.

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

CAMPBELL, J. Y. (1987): "Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis," *Econometrica*, 55, 1249–1273.

CAMPBELL, J. Y., AND R. J. SHILLER (1986): "Cointegration and Tests of Present Value Models," manuscript, Princeton University.

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 Assocation*, 74, 427–431.

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

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflations," *Econometrica*, 50, 987–1008.

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

EVANS, G. B. A., AND N. E. SAVIN (1981): "Testing for Unit Roots: I," *Econometria*, 49, 753–779.

——— (1984): "Testing for Unit Roots: II," *Econometrica*, 52, 1241–1260.

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

——— (1985): "Nonstationary Autoregressive Time Series," in *Handbook of Statistics, Vol.* 5, eds. E. J. Hannan, P. R. Krishnaiah, M. M. Rao. Amsterdam: Elsevier Science Publishers.

FULLER, W. A., D. P. HASZA, AND J. GOEBEL (1981): "Estimation of the Parameters of Stochastic Difference Equations," *Annals of Statistics*, 9, 531–543.

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

——— (1973): "Central Limit Theorems for Time Series Regression," *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 26, 157–170.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

KLEIDON, A. W. (1987): "Variance Bounds Tests and Stock Price Valuation Models," *Journal of Political Economy*, 94, 953–1001.

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

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

NEWEY, W. K., AND K. D. WEST (1987): "A Simple Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

PHILLIPS, P. C. B., AND J. Y. PARK (1986): "Statistical Inference in Regressions with Integrated Processes: Part 1," Cowles Foundation Paper No. 811.

RAO, C. R. (1973): *Linear Statistical Inference and its Applications*. New York: John Wiley and Sons.

ROWLEY, J. R., AND P. K. TREVEDI (1975): *Econometrics of Investment*. New York: John Wiley and Sons.

SAID, S. E., AND D. E. DICKEY (1984): "Testing for Unit Roots in ARMA Models of Unknown Order," *Biometrika*, 71, 599–607.

SARGENT, T. J. (1978): "Estimation of Dynamic Labor Demand Schedules under Rational Expectations," in *Rational Expectations and Econometric Practice*, ed., by Robert E. Lucas, Jr. and Thomas J. Sargent. Minneapolis: The University of Minnesota Press, 1981.

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

SIMS, C., J. STOCK, AND M. WATSON (1986): "Inference in Linear Time Series Models with Some Unit Roots," manuscript.

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

STOCK, J., AND K. D. WEST (1988): "Integrated Regressors and Tests of the Permanent Income Hypothesis," forthcoming, *Journal of Monetary Economics*, 21, 17–33.

WEST, K. D. (1987): "A Specification Test for Speculative Bubbles," *Quarterly Journal of Economics*, 102, 553–580.

——— (1988): "Dividend Innovations and Stock Price Volatility," *Econometrica*, 56, 37–61.

WHITE, H. (1984): *Asymptotic Theory for Econometricians*. New York: Academic Press.