

A SIMPLE APPROACH TO THE STATISTICAL INFERENCE IN LINEAR TIME SERIES MODELS WHICH MAY HAVE SOME UNIT ROOTS*

TAKU YAMAMOTO

Abstract

This paper proposes a simple method to circumvent the difficulties encountered in the statistical inference for linear time series models which may have some unit roots. Specifically, it proposes adding certain artificial regressors in the regression model in addition to the original ones. The method can be applied without a priori knowledge of whether the true process is stationary, integrated, or cointegrated. It is shown that the linear (or non-linear) restrictions of coefficient estimates have an asymptotically chi-square distribution. Thus, by using this method, the testing problem can be handled as a standard asymptotic theory. The power of the test of the proposed method is rather low in the small sample, but the simplicity of the method may be proved to be useful in situations where the conventional approach is difficult to apply. Thus, it may be useful for analysis in financial market where a large body of data is available.

1. *Introduction*

Linear time series models with some unit roots have been a focus of econometrics last ten years. Most notably, various tests of unit root and tests of cointegrations have been proposed. Among them are Fuller (1976), Dickey and Fuller (1979), Phillips (1987), Engle and Granger (1987), and Johansen (1991). Significant general results have been derived on the statistical inference for linear time series models by Phillips and Durlauf (1988), Park and Phillips (1989), and Sims, Stock and Watson (1990, hereafter SSW). Recently, extensive results have been given by Toda and Phillips (1993) on the Granger causality in vector autoregressive (VAR) models. However, as is well known, there are several difficulties in the statistical inference for these models. Firstly, the asymptotic distribution of the ordinary least squares (OLS) estimator may be non-normal. Secondly, the pretests such as tests for unit roots and/or cointegrations are necessary for determining a proper representation of the model.

There have been a few attempts to obtain a standard asymptotic theory for time series models with unit roots. Choi (1993) has derived asymptotic normality of the OLS estimator in univariate autoregressive models. In particular, he has shown that the usual t test statistic for

* The earlier version of the paper was presented at Econometrics Conference at Hitotsubashi University and the Annual Meeting of the Japan Association of Economics and Econometrics. I am grateful to Peter Phillips, Mitsuhiro Odaki, Katsuto Tanaka, Hiro Toda for valuable comments on earlier drafts. The research reported here was financially supported by Nihon Keizai Kenkyu Shorei Zaidan.

unit root is asymptotically normal, if an extra lagged variable is added in the regression model. Along the same line, Toda and Yamamoto (1995) have shown that the usual Wald test statistic of general non-linear restrictions for a VAR model is asymptotically chi-square when extra lagged variables are added in the model. Lütkepohl and Reimers (1992) has shown, based upon the result in Toda and Phillips (1993), that the Wald statistic for testing the Granger causality in a cointegrated bivariate VAR model is always asymptotically chi-square. Phillips (1995) has recently proposed a powerful method which can induce the Wald statistic whose limit distribution is a linear combination of independent chi-square variates, even when the restriction involves coefficients of both stationary and non-stationary variables.

However, when a restriction involves the coefficients of a constant term, the above approaches are not valid, since the limiting distribution of the estimated coefficients of a constant term is non-normal, even if extra lagged variables are added. In such a case, the usual Wald test statistic is no longer asymptotically chi-square, even if extra lagged variables are added. For example, we cannot make the F-type tests for unit roots in Hasza and Fuller (1979) or Dickey and Fuller (1981) to be asymptotically chi-square, by adding extra lagged variables. Further, as is well known, the convergence speed of the estimated coefficients of a constant term is generally slower than those of lagged endogenous variables. It means the second difficulty. As Theorem 2 of SSW shows, if a single linear or non-linear restriction involves estimated coefficients that exhibit different rates of convergence, then the estimated coefficients with the slowest rate of convergence will dominate the test statistic. A Typical example of such restriction occurs in the test of the present value model, such as Ito (1986) and Campbell and Shiller (1987). They escaped the difficulty by assuming stationarity or a certain type of cointegration for the process. The different rates of convergence of estimated coefficients also bothers, for example, when constructing the asymptotic confidence interval of prediction, or when conducting the usual F type test of structural change.

In this paper, we propose a simple method to circumvent the above difficulties. We propose to add a slightly disturbed explanatory variable for each original explanatory variable including a constant term. Thus, we shall have a regression model that has twice as many explanatory variables as the original one. It can be applied to an integrated, a cointegrated or a stationary model without any a priori knowledge. The usual Wald statistic is shown to valid in the sense it has a chi-square distribution asymptotically. Thus, by using the method, we can deal with the testing problem or the construction of prediction interval as if the process were purely stationary.

The obvious merit of the method proposed here is its simplicity. However, as shown in section 3, its size of the test is slightly distorted and its power of the test is rather low in the small sample. Thus, the method is proposed *not* as an alternative to conventional tests for unit roots or cointegrations whose critical values have already been tabulated, *but* as tests of more complicated restrictions that involve coefficients of a constant term and whose critical values have not been tabulated.

In section 2, we propose a simple transformation of the model by which the usual Wald test statistic converges to a chi-square asymptotically. In section 3, we present the results of some experiments that exhibit the small sample properties of the method. The concluding comments are given in section 4.

2. Statistical Inference in a Multivariate Model

We shall consider the Wald test of linear restrictions in a multivariate time series model. Let n variate process $\{X_t = [x_{it}]\}$ be generated by

$$X_t = \delta_0 + \delta_1 t + \omega_t \quad (t = 0, 1, 2, \dots), \tag{2.1}$$

where $\delta_0 = [\delta_{i0}]$ and $\delta_1 = [\delta_{i1}]$ are coefficient vectors, and $\{\omega_t = [\omega_{it}]\}$ is a p -th order autoregressive process

$$\omega_t = \sum_{k=1}^p A_k \omega_{t-k} + \eta_t, \tag{2.2}$$

where $A_k = [a_{ijk}]$ ($k = 1, 2, \dots, p$) are coefficient matrices, and $\eta = [\eta_{it}]$ ($t = 1, 2, \dots, T$) are independently identically distributed with mean zero and covariance matrix $\Sigma = [\sigma_{ij}]$ such that $E|\eta_{it}|^{2+\delta} < \infty$ for some $\delta > 0$. We assume that p is known a priori and $A(z) = 0$ lie either on or outside a unit circle where $A(z) = I_n - \sum_{k=1}^p A_k z^k$. That is, each variable of $\{\omega_t\}$ may be stationary or non-stationary with a unit root. There may be some cointegrations among variables. Substituting $\omega_t = X_t - \delta_0 - \delta_1 t$ into (2.2), we get

$$X_t = \sum_{k=1}^p A_k X_{t-k} + \gamma_0 + \gamma_1 t + \eta_t, \tag{2.3}$$

where $\gamma_0 = [\gamma_{i0}] = A(1)\delta_0 - A'(1)\delta_1$ and $\gamma_1 = [\gamma_{i1}] = A(1)\delta_1$ are coefficient vectors.

In what follows, it is convenient to discuss using the following streamlined notation. We express the model (2.3) as

$$X_t = A^{+'} Y_{t-1}^+ + \eta_t, \tag{2.4}$$

where $Y_{t-1}^+ = [X'_{t-1}, X'_{t-2}, \dots, X'_{t-p}, 1, t]'$, and $A^{+'} = [A_1, A_2, \dots, A_p, \gamma_0, \gamma_1] = [\beta_i^{+'}]$ is the $n \times (np + 2)$ coefficient matrix, and $\beta_i^{+'}$ is the coefficient vector of the i -th equation, i.e., the i -th row of $A^{+'}$. We can rewrite (2.4) in an alternative formula as

$$X = Y_{-1}^+ A^+ + \eta, \tag{2.5}$$

where $X = [X_{p+1}, X_{p+2}, \dots, X_T]'$, $Y_{-1}^+ = [Y_p^+, Y_{p+1}^+, \dots, Y_{T-1}^+]'$, and $\eta = [\eta_{p+1}, \eta_{p+2}, \dots, \eta_T]'$. In a single equation formula, we get

$$s = [I_n \otimes Y_{-1}^+] \beta^+ + u, \tag{2.6}$$

where, $s = \text{Vec}(X)$, $\beta^+ = \text{Vec}(A^+)$, $u = \text{Vec}(\eta)$, \otimes is the kronecker product, and $\text{Vec}(\cdot)$ is the column stacking operator.

Now consider the test of a general linear hypothesis given as follows:

$$H_0: R^+ \beta^+ = r, \text{ vs.} \tag{2.7}$$

$$H_1: R^+ \beta^+ \neq r,$$

where R^+ is the $m \times 2n(np + 2)$ restriction matrix with $\text{rank}(R^+) = m$, and r is the $m \times 1$ vector.

In the traditional approach, first, the rank of cointegration space is evaluated, and then statistical inference is made upon the appropriate model, such as the error-correction model, as in Johansen (1991). Other approaches, such as Phillips (1991), are also available. In what follows, we propose an alternative approach that is based upon the following augmented model:

$$X_t = \sum_{k=1}^p A_k K_{t-k} + \gamma_0 + \gamma_1 t + \sum_{k=1}^p A_k^* X_{t-k}^* + \gamma_0^* 1^* + \gamma_1^* t^* + \eta_t, \tag{2.8}$$

where $X_{t-k}^* = X_{t-k} + T^{-\lambda} \varepsilon_{kt}$ ($k = 1, 2, \dots, p$), $1^* = 1 + T^{-\lambda} \varepsilon_{0t}$ and $t^* = t + T^{-\lambda} \varepsilon_{p+1t}$ are added regressors, ε_{0t} and ε_{p+1t} are scalar random variables and $\varepsilon_{kt} = [\varepsilon_{kjt}]$ ($k = 1, 2, \dots, p$) are $n \times 1$ vectors of random variables which are independently identically distributed with mean zero and covariance matrix $\Phi^k = \text{diag} \{ \Phi_{ii}^k \}$ ($k = 0, 1, 2, \dots, p + 1$) such that $|\varepsilon_{0t}|^{2+\delta} < \infty$, $|\varepsilon_{p+1t}|^{2+\delta} < \infty$ and $|\varepsilon_{kjt}|^{2+\delta} < \infty$ for some $\delta > 0$, and γ_0^* , γ_1^* , and A_k^* ($k = 1, 2, \dots, p$) are zero. We assume that $0 < \lambda < 1/2$. The added regressors X_{t-k}^* ($k = 1, 2, \dots, p$), 1^* , and t^* are artificial ones which are constructed by adding random variates to the original regressors. In practice, we may obtain these random variates by drawing computer generated pseudo normal numbers.

The augmented model (2.8) is compactly expressed as

$$X_t = A' Y_{t-1} + \eta_t, \tag{2.9}$$

where $Y_{t-1} = [Y_{t-1}^+, Y_{t-1}^{*'}]'$, $Y_{t-1}^+ = Y_{t-1} + T^{-\lambda} \varepsilon_t$, where $\varepsilon_t = [\varepsilon'_{1t}, \varepsilon'_{2t}, \dots, \varepsilon'_{pt}, \varepsilon_{0t}, \varepsilon_{p+1t}]'$ is $(np + 2) \times 1$ vector of random variables which are independently identically distributed as $N(0, V_{++})$ and $V_{++} = \text{diag} \{ \Phi^k \}$, $A' = [A^{+'}, A^{*'}]$, and $A^{*'} = [A_1^*, A_2^*, \dots, A_p^*, \gamma_0^*, \gamma_1^*] = [\beta_i^{*'}]$ where $\beta_i^{*'}$ is the i -th row of $A^{*'}$.

We can rewrite (2.9) in an alternative formula as

$$X = Y_{-1} A + \eta \tag{2.10}$$

where $Y_{-1} = [Y_p, Y_{p+1}, \dots, Y_{T-1}]'$. In a single equation formula, we get

$$s = [I_n \otimes Y_{-1}] \beta + u, \tag{2.11}$$

where $\beta = \text{Vec}(A)$. We have $\beta^+ = [I_n \otimes (I_{np+2}, 0)] \beta$. The OLS estimator $\hat{\beta}$ of β is given by

$$\hat{\beta} = [I_n \otimes (Y_{-1}' Y_{-1})^{-1} Y_{-1}'] s \tag{2.12}$$

The general linear hypothesis in (2.7) is written in terms of the coefficient vector β of the augmented model as follows:

$$H_0: R\beta = r, \text{ vs.} \tag{2.13}$$

$$H_1: R\beta \neq r.$$

where $R = R^+ [I_n \otimes (I_{np+2}, 0)]$ is the $m \times 2n(np+2)$ restriction matrix which has zeros for the coefficients of the added regressors. Obviously, we have $R\beta = R^+ \beta^+$.

Consider the following transformation

$$Z_t = DY_t, \tag{2.14}$$

where

$$D = \begin{bmatrix} I_{np+2} & -I_{np+2} \\ D^* & 0 \end{bmatrix},$$

and D^* is an appropriate rotation matrix that separates different stochastic order components. See, for example, (2.7) of SSW. Then, we have

$$Z_t = \begin{bmatrix} Z_t^+ \\ Z_t^* \end{bmatrix} = \begin{bmatrix} -T^{-\lambda} \varepsilon_t \\ D^* Y_t^+ \end{bmatrix}.$$

Correspondingly, the model (2.9) is represented as

$$X_t = \Delta' Z_{t-1} + \eta_t, \tag{2.16}$$

where $\Delta' = A'D^{-1} = [\delta'_i]$, $\delta'_i = [\delta_i^{+'}, \delta_i^{*'}]$ is a coefficient vector of the i -th equation of the above model, and δ_i^+ and δ_i^* are both $(np+2) \times 1$ vectors. In a single equation formula, the above model is written, instead of (2.11), as

$$s = [I_n \otimes Z_{-1}] \delta + u, \tag{2.17}$$

where $Z_{-1} = [Z_p \ Z_{p+1}, \dots, Z_{T-1}]'$, $[I_n \otimes Z_{-1}] = [I_n \otimes Y_{-1}][I_n \otimes D'] = [I_n \otimes Y_{-1}D']$, $\delta = \text{Vec}(\Delta) = [I_n \otimes D'^{-1}] \beta$. For the limit of the moment matrix, we have

$$\gamma^{-1} Z'_{-1} Z_{-1} \gamma^{-1} \Rightarrow V = \begin{bmatrix} V_{++} & 0 \\ 0 & V_{**} \end{bmatrix}, \tag{2.18}$$

where the symbol \Rightarrow indicates convergence in distribution, V_{++} is defined in (2.9) and V_{**} is the $O_p(1)$ square matrix of dimension $np+2$, and

$$\gamma = \begin{bmatrix} T^{1/2-\lambda} I_{np+2} & 0 \\ 0 & \gamma_* \end{bmatrix},$$

and γ_* is an appropriate scaling matrix with $\gamma_*^{-1} = O_p(T^{-1/2})$. See, for example, the scaling matrix γ_T in p.121 of SSW.

Let us define

$$\begin{aligned}\delta^+ &= [I_n \otimes (I_{np+2}, 0)] \delta = [\delta_1^{+'}, \delta_2^{+'}, \dots, \delta_p^{+'}]', \text{ and} \\ \delta^* &= [I_n \otimes (0, I_{np+2})] \delta = [\delta_1^{*'}, \delta_2^{*'}, \dots, \delta_p^{*'}]'. \end{aligned} \quad (2.19)$$

Let $\hat{\delta}$ be the OLS estimator of δ . Since V in (2.18) is block-diagonal, it is easily seen that

$$T^{1/2-\lambda} (\hat{\delta}^+ - \delta^+) \Rightarrow N(0, \Sigma \otimes V_{++}^{-1}). \quad (2.20)$$

Since $\beta^+ = [I_n \otimes (I_{np+2}, 0)] \beta$ and $\beta = [I_n \otimes D'] \delta$, we have

$$\begin{aligned}\beta^+ &= [I_n \otimes (I_{np+2}, 0)] [I_n \otimes D'] \delta = [I_n \otimes (I_{np+2}, D^{*'})] \delta \\ &= \delta^+ + (I_n \otimes D^{*'}) \delta^* \end{aligned} \quad (2.21)$$

Since, as was shown by SSW, $\hat{\delta}^* - \delta^* = O_p(T^{-1})$, we have

$$\begin{aligned}T^{1/2-\lambda} (\hat{\beta}^+ - \beta^+) &= T^{1/2-\lambda} [(\hat{\delta}^+ - \delta^+) + (I_n \otimes D^{*'}) (\hat{\delta}^* - \delta^*)] \\ &= T^{1/2-\lambda} (\hat{\delta}^+ - \delta^+) + o_p(1) \\ &\Rightarrow N(0, \Sigma \otimes V_{++}^{-1}). \end{aligned} \quad (2.22)$$

Now consider the usual Wald test statistic for the hypotheses in (2.13):

$$W = (R\hat{\beta} - r)' [R \{ \hat{\Sigma} \otimes (Y'_{-1} Y_{-1})^{-1} \} R']^{-1} (R\hat{\beta} - r), \quad (2.23)$$

where $\hat{\Sigma} = \sum_{t=p+1}^T (X_t - \hat{A}' Y_{t-1})(X_t - \hat{A}' Y_{t-1})' / \{T - p - 2(np + 2)\}$. Since $R\beta = R^+ \beta^+$, we have from (2.22)

$$\begin{aligned}T^{1/2-\lambda} R (\hat{\beta} - \beta) &= T^{1/2-\lambda} R^+ (\hat{\beta}^+ - \beta^+) \\ &\Rightarrow N(0, R^+ (\Sigma \otimes V_{++}^{-1}) R^+). \end{aligned} \quad (2.24)$$

On the other hand, for the middle bracket of the right hand side in (2.23), we have, by $Z_t = DY_t$,

$$\begin{aligned}T^{-(1-2\lambda)} [R \{ \hat{\Sigma} \otimes D' (\sum_{t=p+1}^T D Y_{t-1} Y'_{t-1} D')^{-1} D \} R']^{-1} \\ = [T^{1-2\lambda} R \{ \hat{\Sigma} \otimes D' \gamma^{-1} \gamma (\sum_{t=p+1}^T Z_{t-1} Z'_{t-1})^{-1} \gamma \gamma^{-1} D \} R']^{-1} \\ = [T^{1-2\lambda} R \{ \hat{\Sigma} \otimes D' \gamma^{-1} (\gamma^{-1} \sum_{t=p+1}^T Z_{t-1} Z'_{t-1} \gamma^{-1})^{-1} \gamma^{-1} D \} R']^{-1}. \end{aligned} \quad (2.25)$$

Noting (2.18) and $R = R^+ [I_n \otimes (I_{np+2}, 0)]$, we have

$$\begin{aligned}[T^{1-2\lambda} R \{ \hat{\Sigma} \otimes D' \gamma^{-1} (\gamma^{-1} \sum_{t=p+1}^T Z_{t-1} Z'_{t-1} \gamma^{-1})^{-1} \gamma^{-1} D \} R']^{-1} \\ = [R^+ \{ \Sigma \otimes T^{1-2\lambda} (I_{np+2}, 0) D' \gamma^{-1} (V^{-1} + o_p(1)) \gamma^{-1} D (I_{np+2}, 0)' \} R^+]^{-1} \\ = [R^+ \{ \Sigma \otimes (I_{np+2}, T^{1/2-\lambda} D^{*'} \gamma_*^{-1}) (V^{-1} + o_p(1)) \\ \quad \times (I_{np+2}, T^{1/2-\lambda} D^{*'} \gamma_*^{-1})' \} R^+]^{-1} \\ = [R^+ \{ \Sigma \otimes (V_{++}^{-1} + T^{1-2\lambda} (D^{*'} \gamma_*^{-1} V_{**}^{-1} \gamma_*^{-1} D^*) + o_p(1)) \} R^+]^{-1} \\ \rightarrow [R^+ \{ \Sigma \otimes V_{++}^{-1} \} R^+]^{-1}, \end{aligned} \quad (2.26)$$

where the symbol \rightarrow indicates convergence in probability, and the last relation comes from $\gamma_*^{-1} = O_p(T^{-1/2})$. Then, from (2.24) and (2.26), we have that W converges to a chi-square variable with m degrees of freedom:

$$W \Rightarrow \chi_m^2. \tag{2.27}$$

Example: We shall illustrate the merit of the proposed method in the following example. Consider a scalar process $\{x_t\}$ given by

$$x_t = \gamma_0 + \alpha_1 x_{t-1} + \gamma_1 t + \eta_t \quad (t = 0, 1, 2, \dots), \tag{2.28}$$

which is a scalar version of (2.3) with $p = 1$. In this case, we have $\gamma_0 = (1 - \alpha_1) \delta_0 + \alpha_1 \delta_1$ and $\gamma_1 = (1 - \alpha_1) \delta_1$. When $|\alpha_1| < 1$, $\{x_t\}$ is a stationary process around a linear trend $\delta_0 + \delta_1 t$. When $\alpha_1 = 1$, the model is reduced to $x_t = \delta_1 + \alpha_1 x_{t-1} + \eta_t$, that is, $\{x_t\}$ is a random walk with drift δ_1 . Assume that we know a priori that $\delta_0 = 0$ for simplicity, and suppose that we are interested in the slope of a linear trend. In particular, suppose that we are interested in testing for $\delta_1 = d$, i.e., the slope of a linear trend being d ($d \neq 0$). It should be noted that here we are *not* interested in whether the process is difference stationary or trend stationary, i.e. $\alpha_1 = 1$ or not. When $\delta_0 = 0$, we have that $\gamma_0 = \delta_1 \alpha_1$ and $\gamma_0 + \gamma_1 = \alpha_1 \delta_1 + (1 - \alpha_1) \delta_1 = \delta_1$. Thus, the null and the alternative hypotheses are formally given as

$$\begin{aligned} H_0: & \gamma_0 - d\alpha_1 = 0 \text{ and } \gamma_0 + \gamma_1 = d \\ H_1: & \gamma_0 - d\alpha_1 \neq 0 \text{ and/or } \gamma_0 + \gamma_1 \neq d \end{aligned} \tag{2.29}$$

The Wald statistic proposed in the paper can be derived from the OLS estimates of the following augmented model:

$$x_t = \gamma_0 + \alpha_1 x_{t-1} + \gamma_1 t + \gamma_0^* 1^* + \alpha_1^* x_{t-1}^* + \gamma_1^* t^* + \eta_t. \tag{2.30}$$

In the conventional approach, we first test for a unit root in (2.28). When an appropriate unit root test is rejected, the Wald test for (2.29) is constructed based upon the OLS estimates of (2.28). It has the standard distribution, i.e. a chi-square distribution. When the unit root test is accepted, we test for

$$\begin{aligned} H_0: & \gamma_0 = d \\ H_1: & \gamma_0 \neq d \end{aligned} \tag{2.31}$$

based upon the OLS estimates of (2.28). In this case, $T^{1/2}(\hat{\gamma}_0 - \gamma_0)$ has the non-standard distribution. Alternatively, we may test for the hypotheses (2.31) based upon OLS estimates of the following regression model:

$$x_t = \gamma_0 + \alpha_1 x_{t-1} + \eta_t \quad (t = 0, 1, 2, \dots).$$

In this case, $T^{1/2}(\hat{\gamma}_0 - \gamma_0)$ has the standard distribution, i.e., a normal distribution. See, for example, West (1988). The obvious drawback of the conventional approach is that it requires

a pretest for a unit root, whereas the proposed method can be applied without paying any attention whether $\alpha_1 = 1$ or not, and thus is free of error in the pretest.

Several remarks are in order:

Remark 1: We may remark on the range of λ . If $\lambda \geq 1/2$, we have $\sum_{i=2}^T (T^{-\lambda} \varepsilon_{it})^2 = T^{-2\lambda} \sum_{i=2}^T \varepsilon_{it}^2 = O_p(1)$ ($i = 0, 1, \dots, P, P+1$). Then, the central limit theory in (2.20) does not work. On the other hand, if $\lambda \leq 0$, (2.22) and (2.26) do not hold. Namely, terms involving $(\delta^* - \delta^*)$ and V_{**}^{-1} in (2.22) and (2.26) are not smaller order of convergence.

Remark 2: If a restriction involves only coefficients of lagged variables X_{t-k} ($k = 1, 2, \dots, p$), we just need to add augmented variables to lagged variables as a special case of (2.8) as follows:

$$X_t = \sum_{k=1}^p A_k X_{t-k} + \gamma_0 + \gamma_1 t + \sum_{k=1}^p A_k^* (X_{t-k} + T^{-\lambda} \varepsilon_{kt}) + \eta_t \quad (2.32)$$

However, for the present case, the following regression model has been recently proposed by Toda and Yamamoto (1995), which includes only an extra lagged variable vector to (2.3):

$$X_t = \sum_{k=1}^p A_k X_{t-k} + A_{p+1} X_{t-p-1} + \gamma_0 + \gamma_1 t + \eta_t. \quad (2.33)$$

In the paper, it was shown that the conventional Wald statistic is asymptotically chi-square. This approach is better than one proposed in this paper, since it saves a number of parameters. Further, the convergence speed of this approach is $O_p(T^{-1/2})$, while that of the present paper is $O_p(T^{-(1/2-\lambda)})$. It means that it has higher efficiency in a large sample.

If a restriction involves coefficients of lagged variables and a trend variable but not a constant term, we may also estimate the following regression model as an extension of the above:

$$X_t = \sum_{k=1}^p A_k X_{t-k} + A_{p+1} X_{t-p-1} + \gamma_0 + \gamma_1 t + \gamma_1^* (t + \varepsilon_t) + \eta_t. \quad (2.34)$$

where ε_t is independently identically distributed as similarly defined in (2.8) and is independent of η_t . An important point here is that, if the restriction does not involve γ_0 , we can easily obtain the estimates of other parameters whose convergence speed is $O_p(T^{-1/2})$.

Remark 3: The essential feature of the proposed is to create estimates which have the same and slow convergence speed. As is clear from the above remark, use of $T^{-\lambda} \varepsilon_t$ in augmented regressors are necessary, when a restriction involves a coefficient of a constant term. This is because its OLS estimate is non-normal and its convergence speed is $O_p(T^{-1/2})$, when there is a unit root in the process. In order to get normality, we need estimates which have a slower convergence speed than $O_p(T^{-1/2})$. It implies that the method proposed in this paper is less efficient than the conventional approach, and consequently possesses lower power of the test. Thus, merit of the present method is its simplicity and is free from possible errors in pretests at sacrifice of its efficiency.

Remark 4: The method proposed in this paper can be easily generalized to higher order unit roots and/or higher trend models. It can be also applied to models with exogenous variables. In such cases, it is generally hard to apply the conventional inference strategy, since there can be too many possible parameter values, and it is difficult to obtain necessary critical values by simulations.

Remark 5: Adding one augmented variable to each existing regressor may easily be a cause of a severe multicollinearity. That may occur in a small sample, but is unlikely in a moderate sample, since an added disturbance vector ε_i has diagonal covariance matrix V_{++} .

Remark 6: The diagonal covariance matrix of ε_i also implies diagonality of the covariance matrix of the OLS estimates $\hat{\beta}_i^+$ of the i -th equation. Thus, inference can be further simplified, if we fix variance of a random number to a specific value, say, $\phi_{ii}^k = c^2$ for all k and i . In this case, we have $V_{++} = c^2 I_{np+2}$. The statistical inference is quite simplified, since the asymptotic covariance matrix of $\hat{\beta}^+$ in (2.22) is simplified to

$$\Sigma \otimes V_{++}^{-1} = c^{-2} \Sigma \otimes I_{np+2}. \tag{2.35}$$

Remark 7: Some may be bothered by using the computer generated artificial variates. However, the essential properties needed for the added ε_{it} ($i = 0, 1, \dots, p, p + 1$) are that they are stationary with mean zero and uncorrelated with η_t . Thus, for example, for practical applications, we may obtain ε_{it} as $\varepsilon_{it} = \Delta^2 X_{t-i-1}$ ($i = 0, 1, \dots, p, p + 1$). In such a case, V_{++} will not be diagonal. But it will not pose any problem.

3. Some Experiments

In this section, we show some experimental results which exhibit finite sample properties of the proposed Wald statistic (2.23) in a single equation model. Here, we are interested in its speed of convergence to a chi-square distribution and in its empirical size and power. All the experiments were executed using GAUSS version 3.0.

3.1 Unit Root Test

Design of the Experiment: While the proposed method is *not* suitable for testing a unit root, it is examined for the sake of comparison of the empirical size and power of the test. We generate data from the following simplest random walk model:

$$x_t = x_{t-1} + \eta_t, \tag{3.1}$$

where η_t is independently identically distributed as $N(0, 1)$. The regression model we consider for testing for a unit root is given by

$$x_t = \gamma_0 + \alpha_1 x_{t-1} + \gamma_1 t + \eta_t. \tag{3.2}$$

The true value of $\beta^{+'} = [\gamma_0, \alpha_1, \gamma_1]$ is $[0, 1, 0]$. The null and the alternative hypotheses for the test are formally given as

$$\begin{aligned} H_0: & \beta^{+'} = [0, 1, 0], \text{ and} \\ H_1: & \beta^{+'} = [0, \alpha_1, 0] \quad (0 < \alpha_1 < 1). \end{aligned} \tag{3.3}$$

The augmented model is given by

$$x_t = \gamma_0 + \alpha_1 x_{t-1} + \gamma_1 t + \gamma_0^* (1 + T^{-\lambda} \varepsilon_{1t}) + \alpha_1^* (x_{t-1} + T^{-\lambda} \varepsilon_{2t}) \tag{3.4}$$

$$+ \gamma_1^* (t + T^{-\lambda} \varepsilon_{3t}) + \eta_t.$$

The true value of $\beta' = [\beta^{+'}, \beta^{*'}] = [\gamma_0, \alpha_1, \gamma_1, \gamma_0^*, \alpha_1^*, \gamma_1^*]$ is $[0, 1, 0, 0, 0, 0]$, and $\varepsilon_t = [\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t}]'$ is independent of η_t , and is independently identically distributed as $N(0, I_3)$. Sample sizes are $T = 50, 250,$ and 1000 . The number of replication is 10,000 for each sample size T . The shorter samples are first parts of the longest one in each replication. For the sake of comparison, we experiment with three values of λ , that is, $\lambda = 0.1, 0.25,$ and 0.4 .

Convergence to a Chi-square Distribution: Figure 1 to Figure 3 show cumulative distributions of the Wald statistic W in (2.23) for $\lambda = 0.1, 0.25,$ and 0.4 , respectively. The solid line in each figure is the theoretical cumulative distribution of a chi-square variable with 3 degrees of freedom. Comparing these figures it is seen that convergence to the chi-square distribution is

FIGURE 1. EMPIRICAL DISTRIBUTIONS: CHI-SQUARE STATISTICS WITH 3dfs, $\lambda = 0.1$

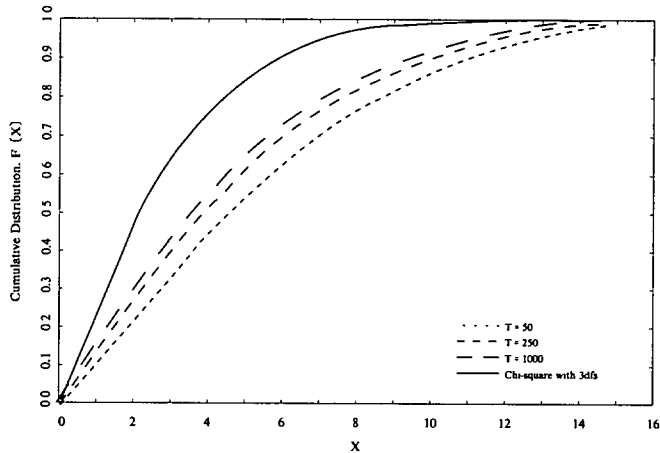


FIGURE 2. EMPIRICAL DISTRIBUTIONS: CHI-SQUARE STATISTICS WITH 3dfs, $\lambda = 0.25$

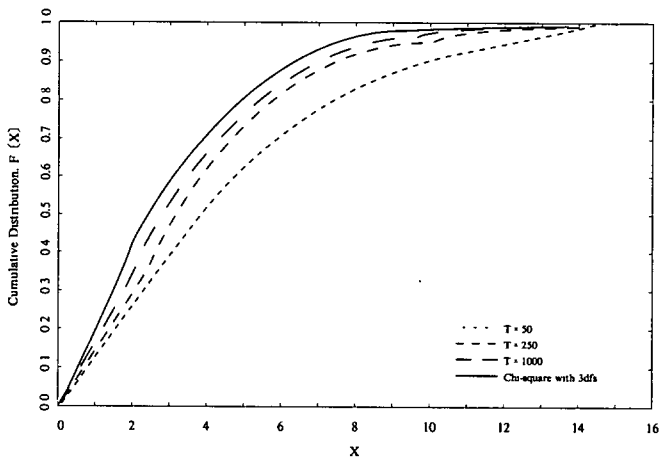
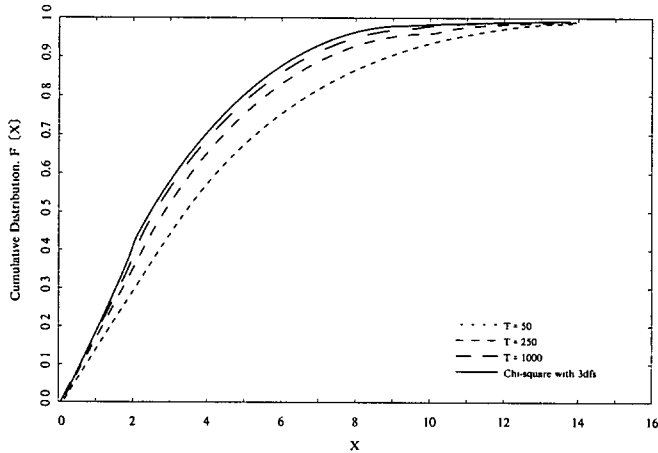


FIGURE 3. EMPIRICAL DISTRIBUTIONS: CHI-SQUARE STATISTICS WITH 3dfs, $\lambda = 0.4$



faster when λ is larger. When λ is small, i.e., $\lambda = 0.1$, there remains a wide discrepancy even $T = 1000$. It may be a result of the fact that the second term in (2.22) disappears faster for larger λ .

Empirical Size and Power of a Unit Root Test: Table 1 reports empirical size and power of a unit root test for (3.3). DF in the table indicates the corresponding Dickey-Fuller test statistic Φ_2 in Dickey and Fuller (1981). The column of $\alpha_1 = 1.00$ shows the empirical size of the test. The empirical size of the proposed method is generally larger than the corresponding nominal

TABLE 1. EMPIRICAL SIZE AND POWER OF CHI-SQUARE TEST FOR A UNIT ROOT WITH 5% SIGNIFICANCE LEVEL

Method \ α_1	0.8	0.9	0.95	1.00
T=50				
DF	0.09	0.04	0.03	0.06
$\lambda = 0.1$	0.19	0.17	0.18	0.22
0.25	0.12	0.12	0.13	0.15
0.4	0.08	0.09	0.09	0.11
T=250				
DF	0.99	0.43	0.09	0.05
$\lambda = 0.1$	0.27	0.14	0.12	0.19
0.25	0.10	0.08	0.08	0.12
0.4	0.06	0.06	0.06	0.07
T=1000				
DF	1.00	1.00	1.00	0.05
$\lambda = 0.1$	0.73	0.25	0.11	0.16
0.25	0.14	0.08	0.06	0.08
0.4	0.07	0.06	0.05	0.05

Notes: 1. DF indicates the Dickey-Fuller Φ_2 test in Dickey and Fuller (1981).
 2. The number of replications is 10000.

size. This size distortion is greater when λ is smaller, as expected from the figures discussed above. The size distortion for $\lambda = 0.1$ is large, even when sample size is relatively large, i.e., $T = 1000$, while it disappears for $\lambda = 0.4$.

The other columns in Table 1 shows the empirical power of the proposed test for various values of α_1 , i.e., $\alpha_1 = 0.8, 0.9$, and 0.95 . We find that power of the proposed test is quite low in comparison with the Dickey-Fuller test, when sample size is 250 or larger. From these observations, the proposed test is not recommended as an alternative to conventional unit root tests such as various Dickey-Fuller tests. We note that power is particularly low large λ , while it is relatively high for small λ . Thus, there appears to be a trade-off between size-distortion and low power of test in terms of value of λ . Thus, in experiments of the next sub-section, we will adopt $\lambda = 0.25$ as a compromise.

3.2 . Test of Linear Restrictions of Coefficient Parameters with Different Speed of Convergence

Design of Experiment: We now consider a test of linear restrictions of coefficients of a model which may or may not have a unit root. Actually, the hypothesis (2.29) for the model (2.28) discussed in the example in the previous section is examined.

We set $d = 1$ in (2.29). Several values of α_1 and δ_1 , taht is, $\alpha_1 = 0.5, 0.9, 1.0$, and $\delta_1 = 0.0, 0.5, 1.0, 1.5, 2.0$ are selected for the experiment. The number of replication is 3000 for each configuration of α_1 and δ_1 .

Empirical Size and Power of the Test: Table 2 reports the empirical size and power of the test described above. The columns of $\delta_1 = 1.0$ show the empirical size for various values of α_1 . When α_1 is equal or close to unity and the sample size is small, the empirical sizes are generally larger than the corresponding nominal sizes. When α_1 is small, i.e., $\alpha_1 = 0.5$, the empirical sizes

TABLE 2. TEST OF LINEAR RESTRICTIONS

Model: $x_t = \gamma_0 + \alpha_1 x_{t-1} + \gamma_1 t + \eta_t$ where $\gamma_0 = (1 - \alpha_1)\delta_0 + \alpha_1\delta_1$ and $\gamma_1 = (1 - \alpha_1)\delta_1$.									
$H_0: \gamma_0 - \alpha_1 = 0$ and $\gamma_0 + \gamma_1 = 1$ (i.e., $\delta_1 = 1$),									
$H_1: \gamma_0 - \alpha_1 \neq 0$ and/or $\gamma_0 + \gamma_1 \neq 1$ (i.e., $\delta_1 \neq 1$).									
α_1	0.5			0.9			1.0		
δ_1	1.0	1.5	2.0	1.0	1.5	2.0	1.0	1.5	2.0
Significance level 10%									
T = 25	.12	.20	.58	.16	.24	.44	.22	.29	.46
250	.11	.32	.76	.11	.35	.80	.17	.39	.76
1000	.09	.52	.97	.09	.58	.99	.13	.59	.97
Significance level 5%									
T = 25	.07	.12	.32	.10	.16	.34	.14	.21	.37
250	.05	.21	.65	.06	.24	.71	.11	.29	.69
1000	.04	.39	.95	.04	.45	.97	.06	.49	.96
Significance level 1%									
T = 25	.02	.05	.15	.03	.06	.18	.05	.09	.22
250	.01	.08	.43	.01	.10	.49	.04	.15	.51
1000	.01	.18	.84	.01	.24	.91	.02	.28	.89

Note: The number of replication is 3,000.

are relatively close to the corresponding nominal size. The other columns show empirical power of the test. It was found that the empirical power is symmetric around $\delta_1 = 1.0$, that is, results of $\delta_1 = 0.0$ and 0.5 are similar to those of 2.0 and 1.5 respectively, those of $\delta_1 = 0.0$ and 0.5 are omitted from the table. The power of the test is slightly higher for a larger α_1 . Thus, there exists a slight trade-off between size-distortion and power of the test in terms of value of α_1 . As in the previous experiment, the empirical power is generally low. However, when $\delta_1 = 2$ and the sample size is large, i.e. $T = 1000$, the test shows reasonable power.

4. Concluding Comments

In this paper, we proposed a simple method to circumvent the difficulties in the statistical inference for linear time series models which may have some unit roots. Specifically, we proposed to add a slightly disturbed explanatory variable for each original explanatory variable in the regression model. It has been shown that it can be applied to a stationary, integrated, or cointegrated process without any modification. The conventional Wald test statistic has been shown to converge asymptotically to a chi-square variable. Thus, by using the method, we can deal with the testing problem as if the process were purely stationary without worrying about whether the process is stationary or not, or when we are interested in the statistical inference without paying attention on the stationarity of the process.

The experiment has shown that power of the test is very low in comparison with the conventional method in the small sample. It requires a very large sample to be useful. Thus, the proposed method is *not* a reasonable alternative to conventional tests for unit roots whose necessary critical values have been already tabulated. Rather, the proposed method is useful for tests for more complicated restrictions of coefficients, when a large sample is available. For example, it can be useful for evaluation of the confidence interval of prediction or the F type test of structural change in the analysis of financial data.

HITOTSUBASHI UNIVERSITY

REFERENCES

- Campbell, J. Y. & Shiller, R. J. (1987), "Cointegration and Tests of Present Value Models," *Journal of Political Economy* 95, pp. 1062-1088.
- Choi, I. (1993), "Asymptotic Normality of the Least Squares Estimates for Higher Order Autoregressive Integrated Processes with Applications," *Econometric Theory* 9, pp. 263-282.
- Dickey, D. A. & Fuller, W. A. (1979), "Distribution of the Estimators for Autoregressive time Series with a Unit Root," *Journal of the American Statistical Association* 74, pp. 427-431.
- Dickey, D. A. & Fuller, W. A. (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica* 49, pp. 1057-1072.
- Engle, R. F. & Granger, C. W. J. (1987), "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica* 55, pp. 251-277.
- Fuller, W. A. (1976), *Introduction to Statistical Time Series*, New York. Wiley.

- Johansen, S. (1991), "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica* 59, pp. 1151–1580.
- Hasza, D. P. & Fuller, W. A. (1979), "Estimation for Autoregressive Processes with Unit Roots," *Annals of Statistics* 7, pp. 1106–1120.
- Ito, T. (1986), "Use of (Time Domain) Vector Autoregressions to Test Uncovered Interest Rate Parity," *Review of Economics and Statistics* 70, pp. 296–305.
- Lütkepohl, H. & Reimers, H.-E. (1992), "Granger-causality in Cointegrated VAR Processes," *Economic Letters* 40, pp. 263–268.
- Park, J. Y. & Phillips, P. C. B. (1989), "Statistical Inference in Regression with Integrated Processes: Part 2," *Econometric Theory* 5, pp. 95–131.
- Phillips, P. C. B. (1987), "Time Series Regression with a Unit Root," *Econometrica* 55, pp. 277–302.
- Phillips, P. C. B. (1991), "Optimal Inference in Cointegrated Systems," *Econometrica* 59, pp. 283–306.
- Phillips, P. C. B. (1995), "Fully Modified Least Squares and Vector Autoregression," *Econometrica* 63, pp. 1023–1078.
- Phillips, P. C. B. & Durlauf, S. N. (1986), "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies* 53, pp. 473–495.
- Sims, C. A., Stock, J. H. & Watson, M. W. (1990), "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica* 58, pp. 113–144.
- Toda, H. Y. & Phillips, P. C. B. (1993), "Vector Autoregressions and Causality," *Econometrica* 61, pp. 1367–1393.
- Toda, H. Y. & Yamamoto, T. (1995), "Statistical Inference in Vector Autoregressions with Possibly Integrated Processes," *Journal of Econometrics* 66, pp. 225–250.
- West, K. D. (1988), "Asymptotic Normality, When Regressors Have a Unit Root," *Econometrica* 56, pp. 1397–1417.