DF-IV Unit Root Tests Using Stationary Instrument Variables

Kyung So Im[a], Junsoo Lee[b,*], Vladimir Arčabić[c], Mansik Hur[d]

[a] *Retired*
8104 Narrow Branch Ct.
Gainesville, VA 20155, USA
[b] *Department of Economics, Finance & Legal Studies*
University of Alabama
263 Alston Hall, Box 870224, Tuscaloosa, AL 35487, USA
[c] *Faculty of Economics and Business, University of Zagreb*
Trg J.F.Kennedyja 6, 10000 Zagreb, Croatia
[d] *Department of Management and Marketing*
California State University, Bakersfield, Bakersfield, CA 93311, USA

Abstract

We propose new unit root tests using stationary instrumental variables in the framework of the Dickey-Fuller (DF) regression. The most noteworthy feature of the suggested tests is that they are free of nuisance parameters. Under the null hypothesis, the proposed test statistic converges to the standard normal distribution regardless of various types of linear deterministic trends or structural breaks in the time series.

Keywords: Unit Root Tests, Dickey Fuller, Instrumental Variables, Standard Normal Distributions, t-test
JEL Classification: C12, C15, C22

1. Introduction

The limiting distributions of the pioneering Dickey-Fuller [1] (DF) unit root tests are non-standard. They are expressed as functionals of Brownian motions. The standard model specifications involve a constant and/or a trend function in the testing regression. However, we often need to allow for various types of deterministic terms. One popular example includes the models with break dummy variables. It is well known that the DF test statistics will depend on the location of breaks in the mean-shift or trend-shift terms. The location poses a nuisance parameter, and things will become complicated in the presence of multiple

*Corresponding author
Email addresses: kyungsoim@gmail.com (Kyung So Im), jlee@cba.ua.edu (Junsoo Lee), varcabic@efzg.hr (Vladimir Arčabić), mansikhur@gmail.com (Mansik Hur)

breaks; as the number of breaks increases, it might become infeasible to obtain the relevant critical values for all possible combinations of multiple breaks. That is, each of different cases changes the DF distribution and requires simulating new critical values. The same situation occurs when stationary covariates are added to the testing regression. The DF tests depend on the nuisance parameter reflecting the contribution of stationary covariates, and new critical values should be tabulated. Things become more complicated when we have mixed cases. For example, consider the case where breaks occur in the data and stationary covariates are added. In such case, obtaining the proper critical values would be difficult, if not impossible.

Although it has been common to have non-standard distributions for unit root tests (also cointegration tests) when we deal with non-stationary data and test the long-run parameters, the source of the non-standard distributions is not necessarily non-stationarity of the data. If different estimation strategies were employed, it would be possible to obtain the usual results where the standard normal distribution or chi-square distribution are utilized in testing on the long-run parameter when dealing with potentially non-stationary variables. For example, So and Shin [2] suggest using an instrumental variable (IV) method using a (nonlinear) sign function as an instrumental variable. Their IV statistics follow the standard normal distribution. The suggested idea is intuitive and appealing. Phillips *et al.* [3] extend the idea further and consider another types of nonlinear IV tests. These works are enlightening and they demonstrate that it is possible to perform valid statistical inference based on the standard distribution theory when testing for a unit root, if we use nonlinear instruments.[1]

In this paper, we suggest new unit root tests that modify the conventional DF tests. We employ *stationary instrumental variables* that can be defined from the Dickey-Fuller type regression. We call these as DF-IV tests since we utilize IV estimation instead of OLS estimation. We adopt the moment conditions that are naturally embodied in the DF testing

[1]Harris *et al.* [4] examine the sample autocovariance function $E(y_t y_{t-k})$ and its long-run variance estimate. It is encouraging to find that the standardized statistic using these estimates has a standard normal distribution. Their method relies on a nonparametric approach.

regression. As we demonstrate, the asymptotic distributions of the new tests are *standard normal*, regardless of different deterministic terms in the underlying model. Therefore, it is possible to use a standard t-test in testing stationarity. What is even more appealing, our tests have a standard normal distribution regardless of different deterministic terms (constant, trend, or dummy variables). It is especially convenient when testing for a structural breaks, because non-standard test statistics depend on a location of the break. Furthermore, since the instruments are naturally given from the regressors in a linear model framework rather than relying on nonlinear functions, they can be easily extended to more general models, which might include the models with stationary covariates and even cointegration models with breaks. In addition, our suggested IV tests do not require the additional recursive procedure to achieve normality. Indeed, the underlying scheme to achieve normality is different from other nonlinear IV tests.

It will be helpful to explain intuitively why the standard normal result will follow in our IV tests. The moment conditions based on IV estimation are different from those based on the Ordinary Least Squares (OLS) estimation. Suppose that we have the data $\{y_t : t = 1, 2, ..., T\}$ which follows a pure stochastic AR(1) process. The conventional DF unit root tests are essentially viewed as using the moment conditions $E(y_{t-1}\Delta y_t) = 0$, $t = 1, ..., T$, under the null hypothesis, against $E(y_{t-1}\Delta y_t) < 0$ under the alternative hypothesis. Various extensions of the existing unit root tests adopt similar moment conditions when y_{t-1} is replaced with $\widetilde{y}_{t-1}$, where $\widetilde{y}_{t-1}$ is the residual from the regression of y_{t-1} on certain deterministic terms. Traditional DF tests adopt the OLS method to utilize these moment conditions. Although the moment $E(y_{t-1}\Delta y_t)$ of the OLS based tests are natural to implement, they result in non-standard distributions because moment conditions depend on the non-stationary term y_{t-1} under the null hypothesis. Thus, the dependency of the moment conditions on y_{t-1} are the source of the non-standard distributions in the usual unit root tests. In contrast, the IV tests that we adopt in this paper are based on moment condition $E[(y_{t-1} - y_{t-1-m})\,\Delta y_t] = 0$. We may wish to utilize the term $y_{t-1} - y_{t-1-m}$ which is a stationary process even when y_{t-1} is non-stationary. It is easy to see that the sample moment $\sqrt{T}\sum_{t=1}^{T}(y_{t-1} - y_{t-1-m})\Delta y_t$ converges to the normal distribution. Moreover, this outcome continues to hold when different deterministic terms are included in the model. Testing can be undertaken with the usual t-tests from the IV estimation of the DF type regression using $y_{t-1} - y_{t-1-m}$ as an instrument. Clearly, a standard normal distribution is free of any nuisance parameters. The normality result also holds in various extended models with breaks or stationary covariates which can lead to the dependency on the nuisance parameters otherwise. This is an important feature of the IV tests. Thus, we do not need to simulate new critical values for the extended models. This feature can offer a great flexibility in the applied research.

The remainder of the paper is organized as follows. In Section 2, we provide test statistics and their relevant asymptotic results in three basic models and the model with structural breaks. In Section 3, we examine the small sample performance of the tests via simulations. Section 4 provides concluding remarks.

2. Stationary IV Test

Suppose we have data y_t, for $t = 1, 2, ..., T$, generated as

$$y_t = d_t + x_t, \tag{1}$$

where d_t is a deterministic component, and x_t is a stochastic component of a time series. Let x_t component follow an autoregressive process

$$x_t = \phi x_{t-1} + \varepsilon_t, \tag{2}$$

where ε_t is the innovation term which is assumed to have zero mean and satisfy the following assumption.

Assumption 1 $\{\varepsilon_t\}$ is a martingale difference process satisfying

$$E(\varepsilon_t|\varepsilon_{t-1,}\varepsilon_{t-2,}...) = 0, \text{ and } E(\varepsilon_t^2|\varepsilon_{t-1,}\varepsilon_{t-2,}...) = \sigma^2, \text{ for } t = 1, 2, ..., . \text{ with } 0 < \sigma^2 < \infty.$$

Here, we can consider various models when the deterministic component d_t is properly defined with different types of deterministic functions. We assume that the initial value y_0 is finite such that $y_0 = O_p(1)$. However, we do not assume that $y_0 = 0$, and it can take any large value. In general, the initial value is not a nuisance parameter under the null, but the power of the corresponding tests can depend on this value. Combining (1) and (2), we have

$$(1-\phi L)y_t = (1-\phi L)d_t + \varepsilon_t, \tag{3}$$

and the testing regression model is

$$\Delta y_t = \beta y_{t-1} + (1-\phi)d_t + \phi\Delta d_t + \varepsilon_t, \tag{4}$$

where $\beta = \phi - 1$. Interest focus is on testing the null hypothesis $\beta = 0$ against the alternative hypothesis $\beta < 0$. We let $z_t = (d_t,\ \Delta d_t)'$. Note that the term Δd_t drops out from the regression when d_t is a function of a polynomial of t, but it remains in z_t when d_t contains dummy variables to capture structural breaks. In this paper, we examine four different models. In the simplest case, we consider a model with zero mean and no trend, where $d_t = 0$. Then we examine models with a non-zero mean and/or a linear trend. We also consider models with structural breaks.

$$\text{Model 1} \quad : \quad \Delta y_t = \beta y_{t-1} + \varepsilon_t \tag{5}$$

$$\text{Model 2} \quad : \quad \Delta y_t = \beta y_{t-1} + \gamma_0 + \varepsilon_t \tag{6}$$

$$\text{Model 3} \quad : \quad \Delta y_t = \beta y_{t-1} + \gamma_0 + \gamma_1 t + \varepsilon_t \tag{7}$$

$$\text{Model 4} \quad : \quad \Delta y_t = \beta y_{t-1} + \gamma_0 + \gamma_1 t + \gamma_2(t \times D_t) + \gamma_3\Delta D_t + \varepsilon_t. \tag{8}$$

For Model 4, we assume that a break occurs between $t = T_B$ and $T_B + 1$, and introduce a dummy variable as

$$D_t = \begin{cases} 0,\ t < T_B \\ 1,\ t \geq T_B + 1 \end{cases} \tag{9}$$

Model 4 is a general model which allows for shifts in both level and trend. When $\gamma_2 = 0$, we have a crash model as a special case, which captures a shift in the mean. Here, $\Delta D_t = 1$ for $t = T_B + 1$, and zero otherwise. Note that ΔD_t should not be omitted in the above regression. It is a one-point dummy variable whose effect can vanish asymptotically. However, if ΔD_t is excluded in the above regression, the resulting test becomes invalid. Lee and Strazicich

[5] and [6] note that some popular unit root tests can suffer from spurious rejections when this term is excluded from the testing regression. One motivation to examine Model 4 is to show the invariance result, implying that the distribution of the IV unit root test does not depend on the nuisance parameter indicating the location of a break, $\lambda = T_B\ /T$. This result can be easily generalized to models with multiple structural breaks. This is a very appealing feature of this test, since applied researchers do not need to simulate new critical values for different types of structural breaks that can occur at different locations.

In this section, we consider the DF testing procedure using IV estimation. We express the DF type regression model as

$$\Delta y_t = \beta y_{t-1} + z_t'\gamma + \varepsilon_t,\ \ t = 1, 2, ..., T, \tag{10}$$

where we let $z_t = 0$, $[1]$, $[1, t]'$, and $[1, t, D_t, tD_t, \Delta D_t]'$ for Models 1, 2, 3 and 4, respectively. A standard estimation procedure is to adopt the least squares method. It is well known that the OLS estimator of β and the corresponding t-statistic will converge to a functional of Wiener processes under the null hypothesis. Moreover, the asymptotic non-standard distribution depends on the type of deterministic variables that are included in the regression. To utilize instrumental variable estimation in regression (10), we define an instrumental variable for the regressor y_{t-1} as

$$w_t = y_{t-1} - y_{t-1-m}, \tag{11}$$

where m is a finite positive integer. Note that the instrumental variable w_t is stationary. Straightforward algebra shows that

$$\hat{\beta}_{DF-IV} = \frac{\sum_{t=1}^{T} w_t \Delta y_t - \sum_{t=1}^{T} w_t z_t' \left(\sum_{t=1}^{T} z_t z_t' \right)^{-1} \sum_{t=1}^{T} z_t \Delta y_t}{\sum_{t=1}^{T} w_t y_{t-1} - \sum_{t=1}^{T} w_t z_t' \left(\sum_{t=1}^{T} z_t z_t' \right)^{-1} \sum_{t=1}^{T} z_t y_{t-1}} \tag{12}$$

and the corresponding t-statistic is given by

$$t_{DF-IV} = \frac{\sum_{t=1}^{T} w_t \Delta y_t - \sum_{t=1}^{T} w_t z_t' \left(\sum_{t=1}^{T} z_t z_t' \right)^{-1} \sum_{t=1}^{T} z_t \Delta y_t}{\hat{\sigma} \sqrt{\sum_{t=1}^{T} w_t^2 - \sum_{t=1}^{T} w_t z_t' \left(\sum_{t=1}^{T} z_t z_t' \right)^{-1} \sum_{t=1}^{T} z_t w_t}} \tag{13}$$

where $\hat{\sigma}^2$ is a consistent estimator of the variance of the error terms in the model, σ^2. It is obtained by

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \left(\Delta y_t - \hat{\beta} y_{t-1} - z_t \hat{\gamma} \right)^2.$$

The distribution of the IV statistic is described as follows.

Theorem 1. *Under Assumption 1 and the null hypothesis, the limiting distribution of the DF type IV statistic converges in distribution to*

$$\sqrt{T} \hat{\beta}_{DF-IV} \xrightarrow{d} \frac{W(1)}{\sigma \left(\frac{1}{2} \left\{ W(1)^2 + 1 \right\} \right)} \tag{14}$$

for Model 1. For Model 2, the limiting distribution is given as

$$\sqrt{T} \hat{\beta}_{DF-IV} \xrightarrow{d} \frac{W(1)}{\sigma \left(\frac{1}{2} \left\{ W(1)^2 + 1 \right\} - W(1) \int_0^1 W(r) dr \right)}. \tag{15}$$

The limiting distribution for Model 3 is

$$\sqrt{T} \hat{\beta}_{DF-IV} \xrightarrow{d} \tag{16}$$
$$\frac{W(1)}{\sigma \left(\begin{array}{c} [W(1)^2 + 1] - W(1) \int_0^\infty W(r) dr + 3[\int_0^1 W(r) dr]^2 + \\ 3W(1) \int_0^1 r W(r) dr - 6 \int_0^1 W(r) dr \int_0^1 r W(r) dr \end{array} \right)}.$$

For Model 4, it follows that

$$\sqrt{T}\hat{\beta}_{DF-IV} \overset{d}{\to} \tag{17}$$
$$\frac{W(1)/\sigma}{\left(\begin{array}{c} [W(1)^2+1]- \\ \left\{3W(\lambda)/\lambda^2+6\left(\lambda W(1)-\lambda\int_0^1 W(r)dr\right)/\lambda^3\right\}\int_0^\lambda rW(r)dr \\ +\left\{2W(\lambda)/\lambda-3\left(\lambda W(1)-\lambda\int_0^1 W(r)dr\right)/\lambda^2\right\}\int_0^\lambda W(r)dr \\ +\left\{\begin{array}{c} 6\left((1-\lambda)W(1)-(1-\lambda)\int_0^1 W(r)dr\right)/ \\ (1-\lambda)^3-3W(1-\lambda)/(1-\lambda)^2 \end{array}\right\}\int_\lambda^1 rW(r)dr \\ +\left\{\begin{array}{c} -\left(3(1-\lambda)W(1)-(1-\lambda)\int_0^1 W(r)dr\right)/ \\ (1-\lambda)^2+2W(1-\lambda)/(1-\lambda) \end{array}\right\}\int_\lambda^1 W(r)dr \end{array}\right)}$$

where $W(\cdot)$ denotes a standard Wiener process; and where the notation $\overset{d}{\to}$ denotes convergence in distribution, $\lambda = \lim_{T\to\infty} T_B/T$, and $W_1(\cdot)$ and $W_2(\cdot)$ are independent Wiener processes. Also, for all four models, the distribution of the corresponding t-statistic follows

$$t_{DF-IV} \overset{d}{\to} W(1). \tag{18}$$

Proof. See the Appendix ■

Therefore, it is clear that the limiting distribution of the *t-statistic* does not depend on the parameters in the deterministic terms. Thus, the distribution is free of any nuisance parameters. When structural breaks in the level or trend are allowed in the testing regression, the distribution is still standard normal and free of the nuisance parameter λ which indicates the location of the break. The intuitive reason for normality is clear from the expressions for the t-statistics in (13). The first term in the numerator of (13) is expressed as the stationary moment condition, $\sqrt{T}\sum_{t=1}^{T}(y_{t-1}-y_{t-1-m})\Delta y_t$, which follows a normal distribution. Note that the first term in the denominator in (13) is stationary. Moreover, the second terms in the numerator and denominator in (13) are asymptotically degenerate. The deterministic components in the models are relegated to these asymptotically negligible terms. As a consequence, the normality of the t-test is applied directly to all four models and other extended models. However, we note that the distribution of the estimated coefficient, $\hat{\beta}_{DF-IV}$, does not converge to a normal distribution. They can be referred to as the coefficient tests. It is shown that $\sqrt{T}\hat{\beta}_{DF-IV}$ converges to functionals of Wiener processes. This result is not surprising. Intuitively speaking, the source for the non-standard distribution for $\hat{\beta}_{DF-IV}$ lies in the first term $\sum_{t=1}^{T} w_t y_{t-1}$ in the denominator of (12), which is given as a functional of a non-stationary term y_{t-1}.

We note that the instrument w_t is asymptotically uncorrelated with y_{t-1} under the null hypothesis.[2] On the other hand, under the alternative hypothesis the correlation coefficient between w_t and y_{t-1} is $1 - \phi^m$. This result essentially shows the consistency of the test. Also, we note that if stationary covariates are used as regressors, their effects will be captured in the second terms of the numerator and the denominator in (13). The resulting t-statistic will have a standard normal distribution and is free of nuisance parameters. This property significantly differs from that of OLS based unit root tests using covariates, whose distribution is a convex combination of the non-standard Dickey-Fuller distribution and the standard normal distribution. Using stationary IV tests will achieve power without incurring this nuisance parameter. Hansen [7] initially suggests that the usual unit root tests can become more powerful if the information in related time series is utilized. He suggests adding stationary variables, if available, to the unit root testing regression. The required condition is that these added covariates are stationary and highly correlated with y_{t-1} but not correlated with Δy_t. The question is whether such variables are readily available in practical applications. However, if they are available we can also improve the power of our IV tests.

In a more general case where the error terms are serially correlated, we can simply adopt an augmented version of the Dickey-Fuller test to control for serial correlations. If we allow p lagged terms of y_{t-1} in the regression, the instrumental variable needs to be adjusted to $w_t = y_{t-1} - y_{t-p-m}$. The asymptotic normality of test statistics in this case is rather obvious.

[2] In finite sample, w_t is correlated with y_{t-1}. The covariance between w_t and y_{t-1} under the null is $m\sigma^2$, and the corresponding correlation coefficient converges to 0 as t increases.

Although the asymptotic result is rather straightforward for any finite m, the size property of the test in finite samples depends on the selected value of m. At the same time, a moderately bigger value of m is necessary for obtaining desirable power properties of the tests. While in theory any value of m can be used to achieve the asymptotic normality result, it is desirable to choose the value of m properly to minimize finite sample biases. In this paper, as a practical guide, we suggest using the value that minimizes the sum of squared residuals.

Owing to the standard normal result, it is obvious that the asymptotic distribution is free of any nuisance parameters in (5) – (8). In this paper, we illustrate this point by examining the model with structural change, but our tests can be easily extended to other settings including seasonality, cointegration and nonlinear models. Adding more regressors in the testing regression may affect the size and power properties in finite samples, but the normality result will still hold.

3. Simulations

In this section, we investigate the small sample properties of the DF-IV unit root tests through Monte Carlo simulations. The data generating process (DGP) implies (1) and (2). We note that all tests are invariant to the parameters γ in the corresponding DGP for which we utilize $d_t = z_t'\gamma$. Thus, using $\gamma = 0$ or any values of γ in the DGP will not change the simulation results when the corresponding expression of z_t is used for each of the models. We use pseudo-*iid* N(0,1) random numbers from the Gauss 15.0.10 RNDNS procedure. We consider two different values for the error variance of the initial value with $\sigma^2 = 1$ and $\sigma^2 = 5$, to see their effects on the power of the tests. All the simulation results are based on 10,000 replications. We use the asymptotic one-sided critical values of the standard normal distribution at the 5% significance level; it is -1.645 in each case.

The instrumental variable w_t is given in (11). We examine the effect of choosing the value of m that minimizes the sum of squared residuals. We also use five different fixed values of m = 1, 2, 3, 4, and 5. We do not examine Model 1, but consider three models, Model 2, 3 and 4, as specified in (5) - (8). We refer to these models as *drift*, *trend*, and *trend-shift*. We use $\phi = 1$ to calculate the empirical size, and $\phi = 0.9$ to compute the power of the tests. We experiment with cases of $T = 50, 100, 300$ and 1000. In the models with level or trend shifts, we assume that the break occurs in the middle of the time period, such that $\lambda = 0.5$.

In Table 1, we report the simulation results for the DF-IV tests. These results can provide a practical guideline to select a value of m that gives the best possible correct size for each of the tests. We notice that when $T = 50$, using $m = 2$ or 3 gives rejection rates that are closer to the nominal sizes than using higher values of m for the model with drift. For the models with trend, level-shift or trend-shift, $m = 1$ gives the lowest size distortions. The tests tend to exhibit size distortions when a higher value of m is used. When $T = 100$, using $m = 3$ gives more or less correct sizes for the model with drift but $m = 1$ gives a better size of the test for the model with trend, level-shift, or trend-shift. When $T = 1000$, choosing $m = 5$ gives more or less correct sizes for the model with drift, but a lower value of m is needed to warrant good size under the null. It is clear that the values of m that lead to correct size increase as the sample size increases.

We now look at the power of the test. The results in Table 2 show that the power increases as m increases. This occurs in all cases. Also, we note that the power increases significantly when the error variance, σ^2 increases, while the size of the tests is not affected by σ^2. This is a good feature of the DF type tests; on the contrary, the LM or DF-GLS type tests have the shortcoming that they tend to lose power under the same situation.

4. Summary and Concluding Remarks

In this paper, we have developed new unit root tests using stationary instrumental variables. Our new unit root tests are based on the moments $E[(y_{t-1} - y_{t-1-m})\,\Delta y_t]$. In contrast, most existing unit root tests are essentially based on the moment conditions $E(y_{t-1}\Delta y_t) = 0$. Since these tests can be undertaken with the usual t-tests in the unit root regression using $y_{t-1} - y_{t-1-m_T}$ as an instrument, we call these tests as 'stationary IV tests'. The asymptotic distribution of our stationary IV statistics is standard normal under the null hypothesis regardless of differing deterministic terms or detrending methods in the underlying model.

As such, with our new testing procedures it is possible to perform valid statistical inference based on the standard distribution theory when testing for a unit root. Also, we demonstrated that the coefficient tests do not have the normal distribution as they depend on y_{t-1} under the null hypothesis.

This paper utilizes the DF type regression. However, it is possible to consider other testing strategies using the Lagrange Multiplier (LM, Schmidt and Phillips [8]) or the DF-GLS detrending methods (Elliott *et al.* [9]). Although the tests adopting such methods may look more powerful than the DF version tests, the power comparison depends on whether the initial value of a time series is large or not. When the initial value is large, these other unit root tests lose power. Then, the DF version of the test is preferred in this situation. Still, such tests can be desired in other situations, and they remain as future research topics.

References

[1] Dickey, D.A. and W.A. Fuller. "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, 74.366a (1979); 427-431.

[2] So, B.S. and D.W. Shin. "Cauchy estimators for autoregressive processes with applications to unit root tests and confidence intervals," *Econometric Theory*, 15.2 (1999); 165-176.

[3] Phillips, P.C.B., J.Y. Park, and Y. Chang. "Nonlinear instrumental variable estimation of an autoregression," *Journal of Econometrics*, 118.1 (2004); 219-246.

[4] Harris, D., B. McCabe, and S. Leybourne. "Some limit theory for autocovariances whose order depends on sample size, " *Econometric Theory*, 19.5 (2003); 829-864.

[5] Lee, J. and M. Strazicich. "Break point estimation and spurious rejections with endogenous unit root tests," *Oxford Bulletin of Economics and Statistics,* 63.5 (2001); 535-558.

[6] Lee, J. and M. Strazicich. "Minimum LM unit root tests with two structural breaks," *Review of Economics and Statistics*," 85.4 (2003); 1082-1089.

[7] Hansen, B. "Rethinking the univariate approach to unit root tests: How to use covariates to increase power," *Econometric Theory*, 11.5 (1995); 1148-1171.

[8] Schmidt, P. and P.C.B. Phillips. "LM tests for a unit root in the presence of deterministic trends," *Oxford Bulletin of Economics and Statistics*, 54.3 (1992); 257-287.

[9] Elliott, G., T.J. Rothenberg, and J. Stock. "Efficient tests for an autoregressive unit root," *Econometrica*, 64.4 (1996); 813-836.

[10] Hamilton, J. "*Time series analysis*." Princeton: Princeton University Press, 1994.

[11] White, H. "*Asymptotic theory for econometricians*," Academic Press, 2001.

Tables

Table 1: Size of the test

	$T = 50$		$T = 100$		$T = 300$		$T = 1000$	
	$\sigma^2 = 1$	$\sigma^2 = 5$	$\sigma^2 = 1$	$\sigma^2 = 5$	$\sigma^2 = 1$	$\sigma^2 = 5$	$\sigma^2 = 1$	$\sigma^2 = 5$
Model 2: With Drift								
$m = 1$	0.023	0.022	0.013	0.014	0.009	0.010	0.007	0.005
$m = 2$	0.044	0.044	0.032	0.030	0.023	0.025	0.018	0.015
$m = 3$	0.060	0.058	0.046	0.043	0.033	0.032	0.025	0.022
$m = 4$	0.071	0.072	0.054	0.051	0.038	0.038	0.031	0.029
$m = 5$	0.082	0.082	0.060	0.060	0.043	0.046	0.035	0.034
Model 3: With Trend								
$m = 1$	0.061	0.060	0.045	0.041	0.026	0.026	0.017	0.018
$m = 2$	0.106	0.104	0.076	0.076	0.047	0.048	0.033	0.035
$m = 3$	0.139	0.135	0.097	0.092	0.059	0.061	0.042	0.047
$m = 4$	0.162	0.162	0.118	0.113	0.071	0.072	0.051	0.053
$m = 5$	0.194	0.187	0.133	0.131	0.080	0.079	0.057	0.054
Model 4: With Trend Shift								
$m = 1$	0.110	0.113	0.066	0.063	0.035	0.033	0.025	0.021
$m = 2$	0.212	0.208	0.112	0.113	0.060	0.054	0.040	0.043
$m = 3$	0.296	0.290	0.150	0.151	0.077	0.072	0.056	0.055
$m = 4$	0.376	0.372	0.189	0.190	0.091	0.090	0.058	0.059
$m = 5$	0.447	0.449	0.225	0.229	0.107	0.107	0.065	0.068

Table 2: Power of the test

	$T = 50$		$T = 100$		$T = 300$		$T = 1000$	
	$\sigma^2 = 1$	$\sigma^2 = 5$	$\sigma^2 = 1$	$\sigma^2 = 5$	$\sigma^2 = 1$	$\sigma^2 = 5$	$\sigma^2 = 1$	$\sigma^2 = 5$
Model 2: With Drift								
$m = 1$	0.074	0.650	0.105	0.632	0.209	0.558	0.491	0.500
$m = 2$	0.119	0.663	0.165	0.655	0.327	0.594	0.733	0.592
$m = 3$	0.149	0.660	0.214	0.657	0.436	0.604	0.859	0.645
$m = 4$	0.181	0.651	0.246	0.648	0.514	0.609	0.930	0.675
$m = 5$	0.201	0.637	0.276	0.638	0.584	0.611	0.964	0.693
Model 3: With Trend								
$m = 1$	0.101	0.384	0.119	0.517	0.224	0.518	0.489	0.485
$m = 2$	0.160	0.405	0.188	0.544	0.344	0.552	0.730	0.577
$m = 3$	0.197	0.396	0.245	0.548	0.439	0.569	0.857	0.628
$m = 4$	0.235	0.382	0.281	0.539	0.513	0.575	0.928	0.663
$m = 5$	0.273	0.359	0.312	0.521	0.588	0.574	0.962	0.680
Model 4: With Trend Shift								
$m = 1$	0.157	0.308	0.142	0.456	0.215	0.501	0.482	0.468
$m = 2$	0.262	0.335	0.220	0.489	0.332	0.537	0.716	0.568
$m = 3$	0.353	0.342	0.287	0.490	0.428	0.555	0.852	0.619
$m = 4$	0.443	0.341	0.345	0.485	0.504	0.559	0.925	0.655
$m = 5$	0.517	0.343	0.390	0.478	0.575	0.558	0.961	0.673

Appendix A.

Lemma 1: Define a partial sum process $S_t = \sum_{j=1}^{t} \varepsilon_j$, and $\xi_t = \varepsilon_{t-1} + ... + \varepsilon_{t-m}$, where m is a finite positive integer. Under Assumption 1,

$$T^{-1} \sum_{t=1}^{T} S_{t-1}\varepsilon_t \xrightarrow{d} \frac{1}{2}\sigma^2 \left[W(1)^2 - 1\right] \tag{A1}$$

$$T^{-3/2} \sum_{t=1}^{T} S_{t-1} \xrightarrow{d} \sigma \int_0^1 W(r)dr \tag{A2}$$

$$T^{-5/2} \sum_{t=1}^{T} tS_{t-1} \xrightarrow{d} \sigma \int_0^1 rW(r)dr \tag{A3}$$

$$T^{-1} \sum_{t=1}^{T} \xi_t S_{t-1} \xrightarrow{d} \frac{1}{2}m\sigma^2 \left[W(1)^2 + 1\right] \tag{A4}$$

$$T^{-1/2} \sum_{t=1}^{T} \xi_t \xrightarrow{d} m\sigma W(1) \tag{A5}$$

$$T^{-3/2} \sum_{t=1}^{T} \xi_t t \xrightarrow{d} m\sigma \left[W(1) - \int_0^1 W(r)dr\right] \tag{A6}$$

$$T^{-1/2} \sum_{t=1}^{T} \xi_t\varepsilon_t \xrightarrow{d} \sqrt{m}\sigma^2 W(1). \tag{A7}$$

$$T^{-1} \sum_{t=1}^{T} \xi_t^2 \xrightarrow{p} m\sigma^2 \tag{A8}$$

Proof. See Hamilton [10] (Proposition 17.1, p.506) for (A1) - (A3). For (A4), note that $\xi_t S_{t-1} = \xi_t(S_{t-1} - S_{t-1-m} + S_{t-1-m}) = \xi_t(\xi_t + S_{t-1-m})$. $T^{-1} \sum_{t=1}^{T} \xi_t^2 \xrightarrow{p} m\sigma^2$ by the strong law of large numbers and $T^{-1} \sum_{t=1}^{T} \xi_t S_{t-1-m} \xrightarrow{d} \frac{1}{2}m\sigma^2 \left[W(1)^2 - 1\right]$ from (A1). (A7) follows since $\{\xi_t\varepsilon_t\}_1^\infty$ is a martingale difference series with variance $m\sigma^4$; see, for example, White [11] (p.133). (A8) follows from the strong law of large numbers. ∎

Lemma 2: Let

$$A_T = \sum_{t=1}^{T} w_t y_{t-1} - \sum_{t=1}^{T} w_t z_t' \left(\sum_{t=1}^{T} z_t z_t' \right)^{-1} \sum_{t=1}^{T} z_t y_{t-1} \tag{A9}$$

$$B_T = \sum_{t=1}^{T} w_t \Delta y_t - \sum_{t=1}^{T} w_t z_t' \left(\sum_{t=1}^{T} z_t z_t' \right)^{-1} \sum_{t=1}^{T} z_t \Delta y_t \tag{A10}$$

$$C_T = \sum_{t=1}^{T} w_t^2 - \sum_{t=1}^{T} w_t z_t' \left(\sum_{t=1}^{T} z_t z_t' \right)^{-1} \sum_{t=1}^{T} z_t w_t \tag{A11}$$

Under the null hypothesis and Assumption 1, $T^{-1}A_T$ converges in distribution to

$$2m\sigma^2 \left([W(1)^2 + 1] \right), \tag{A12}$$

for Model 1, and

$$2m\sigma^2 \left([W(1)^2 + 1] - \frac{1}{2} W(1) \int_0^1 W(r)dr \right), \tag{A13}$$

for Model 2, respectively. $T^{-1}A_T$ converges to

$$2m\sigma^2 \left(\begin{array}{c} [W(1)^2 + 1] - W(1) \int_0^\infty W(r)dr + 3[\int_0^1 W(r)dr]^2 + \\ 3W(1) \int_0^1 rW(r)dr - 6 \int_0^1 W(r)dr \int_0^1 rW(r)dr \end{array} \right), \tag{A14}$$

for Model 3. For Model 4, it converges to

$$2m\sigma^2 \left(\begin{array}{c} [W(1)^2 + 1] - \left\{ 3W(\lambda)/\lambda^2 + 6 \left(\begin{array}{c} \lambda W(1) - \\ \lambda \int_0^1 W(r)dr \end{array} \right) /\lambda^3 \right\} \int_0^\lambda rW(r)dr \\ + \left\{ 2W(\lambda)/\lambda - 3 \left(\lambda W(1) - \lambda \int_0^1 W(r)dr \right) /\lambda^2 \right\} \int_0^\lambda W(r)dr \\ + \left\{ \begin{array}{c} 6 \left((1-\lambda)W(1) - (1-\lambda) \int_0^1 W(r)dr \right) / \\ (1-\lambda)^3 - 3W(1-\lambda)/(1-\lambda)^2 \end{array} \right\} \int_\lambda^1 rW(r)dr \\ + \left\{ \begin{array}{c} -\left(3(1-\lambda)W(1) - (1-\lambda) \int_0^1 W(r)dr \right) / \\ (1-\lambda)^2 + 2W(1-\lambda)/(1-\lambda) \end{array} \right\} \int_\lambda^1 W(r)dr \end{array} \right), \tag{A15}$$

Also, for all four models,

$$T^{-1/2} B_T \xrightarrow{d} \sqrt{m}\sigma^2 W(1), \tag{A16}$$

and

$$T^{-1}C_T \xrightarrow{p} m\sigma^2. \tag{A17}$$

Proof. **<u>Distribution of $T^{-1}A_T$</u>**

We consider projecting the matrix z_t on w_t in order to remove the effects of the deterministic terms. We note that in Models 1, 2 and 3 for which no structural break is allowed for, such projections will completely purge the deterministic components of w_t. In Model 4, the projection does not remove the deterministic component completely. The problem remains for the $m-1$ periods for $t = T_B+1, ..., T_B+m-1$. But, the remaining term appears only for a finite time period, and the effect is ignorable asymptotically. We, therefore, have under the null hypothesis

$$T^{-1}A_T = T^{-1}\sum_{t=1}^{T}\xi_t S_{t-1} - T^{-1}\sum_{t=1}^{T}\xi_t z_t' \left(\sum_{t=1}^{T} z_t z_t'\right)^{-1} \sum_{t=1}^{T} z_t S_{t-1} + o_p(1) \tag{A18}$$

where S_{t-1} and ξ_t are defined in Lemma 1. For the first term, we have the result in (A4). This is a case for Model 1. For the asymptotic distribution of Models 2 - 4, we note that the distribution of the second term depends on z_t. We need to examine each case separately. We re-write the second term in (A18) as

$$-\sum_{t=1}^{T} D_T \xi_t z_t' \left(\sum_{t=1}^{T} D_T z_t z_t' D_T\right)^{-1} T^{-1}\sum_{t=1}^{T} D_T z_t S_{t-1} + o_p(1) \tag{A19}$$

where D_T is defined as below.

Model 2: when $z_t = 1$, we let $D_T = [T^{-1/2}]$. Then, it is easy to see from (A2) and (A5) that

$$T^{-1}\sum_{t=1}^{T}\xi_t \left(\frac{1}{T}\right)\sum_{t=1}^{T} S_{t-1} \xrightarrow{d} m\sigma^2 W(1)\int_0^{\infty} W(r)dr$$

Applying the above result and (A4) to (A18), we can obtain the desired result of (A13).

Model 3: when $z_t = [1, t]$, we let $D_T = diag[T^{-1/2}, T^{-1/3}]$. It follows from (A2), (A3), (A5), and (A6) that

$$\begin{bmatrix} T^{-1/2} \\ T^{-3/2} \end{bmatrix} \sum_{t=1}^{T} \xi_t z_t' \xrightarrow{d} \begin{bmatrix} m\sigma W(1) \\ m\sigma \left[W(1) - \int_0^1 W(r)dr \right] \end{bmatrix} \tag{A20}$$

$$\left(\sum_{t=1}^{T} D_T z_t z_t' D_T \right)^{-1} \xrightarrow{d} \begin{pmatrix} 4 & -6 \\ -6 & 12 \end{pmatrix}$$

$$\begin{bmatrix} T^{-3/2} \\ T^{-5/2} \end{bmatrix} \left(\sum_{t=1}^{T} z_t S_{t-1} \right)' \xrightarrow{d} \begin{bmatrix} \sigma \int_0^1 W(r)dr \\ \sigma \int_0^1 rW(r)dr \end{bmatrix}$$

Then, we obtain the result in (A14) by evaluating the matrix multiplication of the above terms in (A20).

Model 4: when $z_t = [1, t, D_t, tD_t, \Delta D_t]$, we drop ΔD_t for simplicity in deriving the asymptotic distribution. Including ΔD_t in the regression has the effect of removing one observation from the series, hence it should have no effects on the asymptotic distribution. Therefore, we have $z_t = [1, t, D_t, tD_t]$. We let $\ddot{z}_t = [D_t, D_t t, D_{2t}, D_{2t} t]$, where $D_{2t} = \begin{cases} 1, \ t \leq T_B \\ 0, \ t \geq T_B + 1 \end{cases}$. Note that the projection of z_t has the same effect as the projection of $\ddot{z}_t$. Then, we consider $D_T = diag[T^{-1/2}, T^{-1/3}, T^{-1/2}, T^{-1/3}]$ to have $\sum_{t=1}^{T} D_T \xi_t z_t' \left(\sum_{t=1}^{T} D_T z_t z_t' D_T \right)^{-1}$ $\sum_{t=1}^{T} D_T z_t S_{t-1} D_T = \sum_{t=1}^{T} \xi_t \ddot{z}_t' \left(\sum_{t=1}^{T} \ddot{z}_t \ddot{z}_t' \right)^{-1} \sum_{t=1}^{T} \ddot{z}_t S_{t-1}$. But, $\sum_{t=1}^{T} \ddot{z}_t \ddot{z}_t'$ is block diagonal. It is straightforward to see that

$$\left(\sum_{t=1}^{T} D_T \ddot{z}_t \ddot{z}_t' D_T \right)^{-1} = \begin{bmatrix} 4\lambda^{-1} & -6\lambda^{-2} & 0 & 0 \\ -6\lambda^{-2} & 126\lambda^{-3} & 0 & 0 \\ 0 & 0 & 4(1-\lambda)^{-1} & -6(1-\lambda)^{-2} \\ 0 & 0 & -6(1-\lambda)^{-2} & 12(1-\lambda)^{-3} \end{bmatrix}. \tag{A21}$$

and

$$
D_T \sum_{t=1}^{T} \xi_t \ddot{z}_t' \xrightarrow{d} \tag{A22}
$$

$$
\begin{bmatrix} m\sigma W(\lambda) \\ m\sigma \left[\lambda W(1) - \lambda \int_0^1 W(r)dr\right] \\ m\sigma W(1-\lambda) \\ m\sigma \left[(1-\lambda)W(1-\lambda) - (1-\lambda)\int_0^1 W(r)dr\right] \end{bmatrix}
$$

$$
T^{-1} D_T \left(\sum_{t=1}^{T} \ddot{z}_t S_{t-1}\right)' \xrightarrow{d} \begin{bmatrix} \sigma \int_0^\lambda W(r)dr \\ \sigma \int_0^\lambda rW(r)dr \\ \sigma \int_\lambda^1 W(r)dr \\ \sigma \int_\lambda^1 rW(r)dr \end{bmatrix}
$$

Applying (A4) and the above results to (A18), we obtain the desired result of (A15).

Distribution of $T^{-1/2}B_T$

We have under the null hypothesis

$$
T^{-1/2} B_T = T^{-1/2} \sum_{t=1}^{T} \xi_t \varepsilon_t - T^{-1/2} \sum_{t=1}^{T} \xi_t z_t' \left(\sum_{t=1}^{T} z_t z_t'\right)^{-1} \sum_{t=1}^{T} z_t \varepsilon_t + o_p(1) \tag{A23}
$$

The result follows for the first term using (A7). The second term is ignorable asymptotically since $T^{-1/2} \sum_{t=1}^{T} \xi_t z_t' \left(\sum_{t=1}^{T} z_t z_t'\right)^{-1} \sum_{t=1}^{T} z_t \varepsilon_t = O_p(T^{-1/2})$.

Distribution of $T^{-1}C_T$

We have under the null hypothesis

$$
T^{-1} C_T = T^{-1} \sum_{t=1}^{T} \xi_t^2 - T^{-1} \sum_{t=1}^{T} \xi_t z_t' \left(\sum_{t=1}^{T} z_t z_t'\right)^{-1} \sum_{t=1}^{T} z_t \xi_t + o_p(1). \tag{A24}
$$

The result follows from (A8), noticing that the second term on the right hand side is negligible since $T^{-1} \sum_{t=1}^{T} \xi_t z_t' \left(\sum_{t=1}^{T} z_t z_t'\right)^{-1} \sum_{t=1}^{T} z_t \xi_t = O_p(T^{-1})$.

■

Proof of Theorem 1: Note

$$\sqrt{T}\hat{\beta}_{DF-IV} = \frac{T^{-1/2}B_T}{T^{-1}A_T}, \tag{A25}$$

where A_T and B_T are defined in (A9) and (A10) of Lemma 2. Apply the results for $T^{-1}A_T$ in (A12), (A13) and (A14) for Models 1,2 and 3, respectively, and the result in (A23) for $T^{-1/2}B_T$. Proof is complete from continuous mapping theorem.

We have

$$t_{DF-IV} = \frac{B_T}{\hat{\sigma}\sqrt{C_T}} \xrightarrow{d} \frac{\sqrt{m}\sigma^2 W(1)}{\sigma\sqrt{m\sigma^2}} = W(1) \tag{A26}$$

where C_T is defined in (A11). Apply the results for $T^{-1}A_T$ in (A12), (A13) and (A14) for Models 1,2 and 3, respectively, and the result in (A24) for $T^{-1/2}C_T$. Proof is complete from continuous mapping theorem.