TESTING FOR UNIT ROOTS USING WEIGHTED MOMENT CONDITIONS

Kyung So Im
University of Central Florida
and
Junsoo Lee*
University of Alabama

October 20, 2006

Abstract

We propose new unit root tests based on instrumental variables estimation utilizing weighted moment conditions. Under the null hypothesis, the asymptotic distribution of the proposed test statistics is standard normal. The normality result holds in more general models using different types of linear deterministic trends and different detrending methods. An important advantage is that the IV unit root tests do not entail nuisance parameters. In addition, the power of the new tests compares quite favorably with existing unit root tests.

JEL Classification: C12, C15, C22
Key Words: Unit Root Tests, Weighted Moments, Instrumental Variables, Standard Normal Distribution

*Corresponding author: Department of Economics, Finance and Legal Studies, Box 870224, University of Alabama, Tuscaloosa, AL 35487, e-mail: jlee@cba.ua.edu, phone: 205-348-8978, fax: 205-348-0590.

1 Introduction

The limiting distributions of the usual unit root tests are typically nonstandard and expressed as functionals of Brownian motions. Since the pioneering work by Dickey and Fuller (1979, DF henceforth), many authors have derived and examined various nonstandard distributions whenever new model specifications are considered. Since the specific expressions of the nonstandard asymptotic distributions vary over different models, different sets of critical values are provided for each type of model using different deterministic components and detrending methods. A fundamental issue is that the limiting distributions often depend on various nuisance parameters and the dependency on nuisance parameters can pose a problem in some extended models. For example, consider the models with multiple breaks, stationary covariates, endogenous breaks, or panel models with breaks. In these models, it becomes cumbersome or infeasible to obtain the relevant asymptotic critical values. Clearly, there are many cases where it is desirable or necessary to utilize the tests that do not depend on nuisance parameters.

In this paper, we propose new unit root tests whose asymptotic distributions are standard normal. Obviously, a standard normal distribution is free of any nuisance parameters. Our goal is to develop convenient tests that remain valid in more general models without sacrificing the power significantly, but maintaining desirable properties. To achieve this outcome, we adopt new procedures based on instrumental variables (IV) estimation utilizing weighted moment conditions. Our departure differs from the traditional testing scheme. Suppose for simplicity that we have data $\{y_t\colon\ t = 0, 1, .., T\}$, which follows a pure AR(1) process with no deterministic components. It can be shown that the usual unit root tests are essentially based on the moment conditions $E(y_{t-1}\Delta y_t) = 0$, $t = 1, .., T$, under the null hypothesis, against $E(y_{t-1}\Delta y_t) < 0$ which holds under the alternative hypothesis. The popular tests advanced by Dickey and Fuller (1979) can be seen as utilizing these moment conditions. Various extensions of the DF tests using the least squares method can be viewed as adopting similar moment conditions. Although the moments $E(y_{t-1}\Delta y_t)$ of such tests are very natural to implement, they result in non-standard distributions. This is so because the moment conditions depend on the non-stationary term y_{t-1} under the null hypothesis.

In contrast, we suggest to use the moments $E[(y_{t-1} - y_{t-1-m})\,\Delta y_t]$. As we shall show in more detail in the next section, using these different moments leads to the standard normal asymptotic result. The underlying intuition is clear. For a moment, suppose that m is a fixed finite number. Then, $y_{t-1} - y_{t-1-m}$ is a stationary process and we can see that the sample moment $\sqrt{T}\sum_{t=1}^{T}(y_{t-1} - y_{t-1-m})\Delta y_t$ converges to a normal distribution. Therefore, it can be expected that the asymptotic distributions of the corresponding unit root t-statistics will be standard normal. The point of this treatment can be seen more intuitively when our moments are compared to those in the usual unit root tests. Since $E(y_{t-1}\Delta y_t) = E\left[(y_{t-1} - y_{t-1-m})\Delta y_t\right] + E(y_{t-1-m}\Delta y_t)$, using the moments $E\left[\left(y_{t-1} - y_{t-1-m}\right)\Delta y_t\right]$ amounts to truncating the second term which does not contribute to the ability to reject the null when the alternative is true. It is easy to see that the correlation between Δy_t and y_{t-1-m} dies out exponentially as m gets larger under the alternative hypothesis. Thus, it is sensible to construct new tests that utilize the moments $E\left[(y_{t-1} - y_{t-1-m})\Delta y_t\right]$ after truncating $E(y_{t-1-m}\Delta y_t)$. 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 instrumental variable, we call these as "IV tests".[1] [2]

[1] Hall (1989) provided an IV unit root test in the presence of moving average errors. He suggested using y_{t-m}, $m > q$, as an instrument for y_{t-1} when there exists an $MA(q)$ errors in the DF type regression. Lee and Schmidt (1994) adopted a similar procedure. The motivation of these IV tests is to correct for moving average errors and their distribution is nonstandard.

[2] We wish to point out that there are other ways to obtain the normality result of unit root tests. So and Shin (1999) suggest using a sign function as an instrumental variable in the Dickey-Fuller type testing regression. Phillips, Park and Chang (2004) have shown that an integrable function of the lagged dependent variable can be used as an instrument. The resulting t-statistic for a unit root converges to a standard normal distribution. These approaches are novel and have helped to motivate the present paper. Our tests differ from these other tests in some important respects. First, it can be viewed that these other approaches utilize $E\left[f\left(y_{t-1}\right)\Delta y_t\right]$, which is a nonlinear function of an integrated process (y_{t-1}) under the null. Our moment conditions are different. Second, in other IV type tests, a recursive demeaning procedure is often necessary to achieve the normality result. The procedure works well and the standard normality result follows in the basic models. In more general models that include a linear trend, level or trend shifts, nonlinear trends or stochastic regressors, a similar recursive detrending procedure could be possibly considered. However, the recursive detrending procedure can pose a problem in some general models; Sul, Phillips, and Choi (2003) note that a recursive detrending procedure entails a nuisance parameter even for the model with a linear trend. In contrast, no recursive demeaning or detrending procedure is necessary for our IV tests. The approach adopted by Harris, McCabe and Leybourne (2003) also differs from ours. They examine the sample autocovariance function $E(y_t y_{t-k})$ and the standardized statistic using the longrun variance of $E(y_t y_{t-k})$.

The power of the IV tests will increase under the alternative as m increases. In this regard, it is reasonable to make m grow as the sample size increases. Thus, we consider tests where m depends on T such that $m_T = T^\delta$.[3] We use the notation m_T to signify that m is a function of T. In the next section, we will show that the null distribution of tests based on the moments $E\left[(y_{t-1} - y_{t-1-m_T})\Delta y_t\right] = 0$ will converge to a standard normal distribution. In addition, we derive the required condition for this asymptotic result. It turns out that the required condition is shown to be rather mild. As long as m_T is given as $m_T = T^\delta$, with $0 \le \delta < 1$, the standard normal result follows. Therefore, by choosing δ reasonably close to 1, we are able to construct new tests that are more powerful but still follow asymptotically a standard normal distribution.

To improve further the power of our IV unit root tests, we suggest utilizing weighted moments $\sum_{j=1}^{m_T} \psi_j E(\Delta y_{t-j}\Delta y_t)$, where $0 < \psi_j \le 1$ for $j = 1, ..., m_T$. As a matter of fact, when we adopt this strategy, the power of our tests is comparable to the most powerful existing tests in small samples and moderately large sample sizes that we examine. Thus, we focus on these tests that utilize weighted moments. This strategy can be explained intuitively. Since $y_{t-1} - y_{t-1-m} = \sum_{j=1}^{m} \Delta y_{t-j}$, and the correlation between Δy_t and Δy_{t-j} gets smaller for larger j under the alternative hypothesis, it is reasonable to expect that the power of the tests will increase when heavier weights are given to the terms with higher correlations.[4] Using $y_{t-1} - y_{t-1-m_T}$ amounts to setting $\psi_j = 1$ for all j, which is the case of using a uniform window. Our simulation results demonstrate a non-trivial power gain when a Bartlett window is adopted with $\psi_j = 1 - \frac{j-1}{m_T}$. The power gain comes with a small

[3] One limitation of using a finite small value of m is that the order of the test becomes low. For instance, the order of the IV test using $m = 1$ is $\sqrt{T}$; this result can be shown later as a special case. The tests proposed by So and Shin (1999) and Harris *et al.* (2003) have the same limitation since the order of these tests is also $\sqrt{T}$. The lower order can result in loss of power. As such, we adopt a few ways to improve the power of our IV tests and this task is the focus of the present paper. Clearly, the order of the test increases as m increases and it will be shown that the order of the test is $T^{(1+\delta)/2}$ when $m_T = T^\delta$. While the order is still less than T when $\delta < 1$, there is no significant loss of power as we will see in section 3. Instead, our improved IV tests are more powerful than the Dickey-Fuller test whose order is T. This outcome is possible when we utilize weighted moments for our IV tests.

[4] We note that the nonlinear IV test suggested by Phillips, Park and Chang (2004) implicitly imposes different weights on highly correlated terms. This treatment leads to increased power. This insight helped us develop our new tests.

price of having to deal with some small sample bias. While our results regarding the normal distributions hold asymptotically, the lower order terms in the asymptotic distributions are not negligible in small samples when m_T depends on T. This can lead to size distortions with a large value of m_T. Since we use the asymptotic critical value of the standard normal distribution, say -1.645 at the 5% significance level in all cases, it is important to have the correct size under the null. As such, we suggest modifying the test statistics to correct for small sample bias. We provide details of how to correct for the small sample bias in Section 2. With this correction, and when the Bartlett window is used, the size of our IV tests is fairly good and the power is closely comparable to the most powerful existing unit root tests. This is an encouraging property of our IV tests, since the desirable normality result can be obtained without inducing any significant loss of power.

The most important advantage of using our proposed IV tests is that they do not entail nuisance parameters. This benefit of our IV tests is not readily available with existing tests using the least squares method. One popular example of having nuisance parameters in unit root tests is when structural changes are allowed. When structural breaks are considered, the distributions of unit root test statistics typically depend on the number and location of the breaks. Then, one needs to simulate critical values for different break locations and different types of models. Finding the asymptotic distributions for these models can be done, but the dependence on nuisance parameters can pose a problem when testing is extended to more general models. For example, consider panel models with breaks. When each cross-section unit can have different numbers and different break points, it will be extremely difficult to obtain proper critical values for panel version tests. Such panel version tests are infeasible unless underlying univariate tests are invariant to nuisance parameters.[5] We demonstrate the invariance result of our IV unit root tests with the structural change models with level and trend shifts.

[5] The dependency on nuisance parameters induces a significant problem in endogenous unit root tests that employ estimated break points. Since the null distributions of the so-called minimum or maximum tests depend on the break location and magnitudes, it is often *assumed* that there are no breaks under the null; otherwise the tests diverge. Then, these endogenous tests exhibit serious size distortions when there are breaks under the null. The source of size distortions of these tests is their dependency on the nuisance parameters of the corresponding exogenous version tests; see Nunes, Newbold and Kuan (1997) and Lee and Strazicich (2001) for more details. In section 4, we discuss more examples and possible extensions of our IV tests where nuisance parameters pose a problem.

Note that no previous tests with trend-shifts are invariant to the nuisance parameter. The distributions of our IV tests are invariant to the number of breaks and break locations. This outcome makes a contrast with the property of the Perron (1989) type tests whose asymptotic distributions depend on these nuisance parameters. The desirable invariance property of our IV tests appears quite general and is expected to be useful in other models where the dependence on nuisance parameters is a problem. For instance, Saikkonen and Lütkepohl (2002) utilize the invariance property of the LM test which holds only with level-shifts (see Amsler and Lee, 1995) and extend the result to general models with nonlinear deterministic terms.

When a time series contains a non-zero mean, a linear trend or other deterministic terms, we need to control for these effects. This is the detrending method. There are a few different detrending methods available in the literature. The DF type detrending method has been popular but the LM detrending method advocated by Schmidt and Phillips (1992) yields more powerful tests, as noted by Vougas (2003). Actually, it turns out that the IV tests based on the LM detrending method are more powerful than the IV tests using the DF type detrending method.[6] In addition, although it is feasible to consider the IV tests using the GLS detrending method, the IV tests based on the LM and GLS detrending methods produce very similar properties in terms of size and power. Thus, we focus on using the IV tests based on the LM detrending method.

The remainder of the paper is organized as follows. In Section 2, we provide the details of test statistics for each model. We also derive conditions for the asymptotic results and provide expressions for the bias term that can exist when m_T is big. In Section 3, we examine the small sample performance of the tests via simulations. Section 4 provides summary and concluding remarks.

2 IV Tests

Suppose that we have data y_t, for $t = 0, 1, 2, ..., T$, which are generated as

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

[6] The power comparison depends on the initial value in the data generating process. We examine this issue in section 3.

Here, x_t is the stochastic component of the series following an autoregressive process

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

where ε_t is the innovation term and is assumed to satisfy the following condition.

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

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

Here, d_t is the deterministic component of y_t in which we can consider general models with various types of deterministic functions. We assume that the initial value y_0 is finite such that $y_0 = O_p(1)$. 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

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

where $\beta = \phi - 1$. Interest centers on testing the null hypothesis $\beta = 0$ against the alternative hypothesis $\beta < 0$. Note that the term Δd_t drops out from the regression when d_t is a function of a polynomial of t, but it does not disappear when d_t contains dummy variables to capture structural breaks. In this case, omitting Δd_t can drive the resulting test statistic to diverge. We examine four different models.

2.1 Results in the Basic Model

In the simplest case, we consider an AR(1) model with zero mean and no trend such that $d_t = 0$

$$\Delta y_t = \beta y_{t-1} + \varepsilon_t, \; t = 1, 2, ..., T, \tag{5}$$

where $\{\varepsilon_t : t = 1, 2, ..., T\}$ is a differenced martingale sequence. Then, one may construct an instrumental variable for y_{t-1} as $w_{t-1} = y_{t-1} - y_{t-1-m}$, where m is a finite number. By applying IV estimation, we can obtain a t-statistic for β. It is a valid IV statistic that follows a standard normal distribution.[7] We can show that the t-statistic utilizes the moment condition $E[(y_{t-1} - y_{t-1-m})\,\Delta y_t]$. As noted in the previous section, we can improve the power of the IV tests by increasing m so that we can utilize all possible useful information. To do so, we utilize the weighted average of stationary moments in such a way to put more weight on the terms that are highly correlated with Δy_t, and less or no weight on the terms that are less important. Accordingly, letting $m_T = T^{\delta}$, we define the instrumental variable for y_{t-1} in (5) as

$$w_{t-1} = \begin{cases} \sum_{j=1}^{t-1} \psi_{t-j} \Delta y_j, & \text{for } t = 2, ..., m_T, \\ \sum_{j=1}^{m_T} \psi_{m_T+1-j} \Delta y_{t-(m_T+1)+j}, & \text{for } t = m_T+1, m_T+2, ..., T, \end{cases} \tag{6}$$

with $0 < \psi_j \leq 1$, for $j = 1, 2, ..., m_T$. Here, ψ_j is used as a weight for each term. From 2SLS using w_{t-1} as an instrument for y_{t-1}, we obtain the t-statistic

$$t_{IV} = \frac{\sum_{t=1}^{T} w_{t-1} \Delta y_t}{\hat{\sigma} \sqrt{\sum_{t=1}^{T} w_{t-1}^2}}, \tag{7}$$

where $\hat{\sigma}$ is a consistent estimator of σ.[8] Then, we have the following result.

Theorem 1 *Under Assumption 1 and the null hypothesis of $\beta = 0$ in (5), the t-statistic in (7) follows*

$$t_{IV} \xrightarrow{d} N(0, 1) \tag{8}$$

if $m_T = T^{\delta}$, $0 \leq \delta < 1$.

Proof. *See the Appendix.* ■

---

[7] The result continues to hold in the presence of autocorrelations, under which we can construct the IV as $w_{t-1} = y_{t-1} - y_{t-1-p-m}$ where p is an AR term in the augmented regression. The usual arguments and findings regarding the selection of p in the augmented version tests would apply; see Ng and Perron (1995).

[8] One interesting observation emerges since the IV statistic in (7) may look similar to the usual t-statistic on β in the regression, $\Delta y_t = \beta w_{t-1} + e_t$. This regression amounts to using the instrument w_{t-1} as a regressor. For instance, in a special case with $w_{t-1} = \Delta y_{t-1}$, this regression becomes $\Delta y_t = \beta \Delta y_{t-1} + e_t$. However, this regression does not yield any meaningful inference. The difference between the t-statistic from this regression and our IV statistic lies in the estimate of the error variance. The error variance estimate obtained from this invalid regression is not a consistent estimator for σ^2. Standard econometrics wisdom indicates the peril of using an instrument as a regressor and using incorrect error variance for the IV or 2SLS estimator.

Thus, the t-statistic defined in (7) converges in distribution to a standard normal distribution under the very mild condition that $0 \leq \delta < 1$. This is the required condition to obtain the asymptotic standard normality result when we use stationary moments. As such, there are many possible choices for δ. In finite samples, choosing a proper value of δ that ensures an appropriate size and high power is needed. In the next section, we suggest proper values of δ for each of different models[9].

There are many options for ψ_j, $j = 1, 2, ..., m_T$. A uniform window imposes equal weights. To make ψ_j decline in j so that we provide higher weights for more highly correlated pairs under the alternative hypothesis, we consider the Bartlett window. We use these two weighting schemes throughout the paper.

$$\begin{aligned} &\text{Uniform window: } \psi_j = 1, \text{ for } j = 1, 2, ..., m_T \\ &\text{Bartlett window: } \psi_j = 1 - \frac{j-1}{m_T}, \text{ for } j = 1, 2, ..., m_T. \end{aligned}$$

When the uniform window is used, the instrumental variable is simplified as

$$w_{t-1} = \begin{cases} y_{t-1}, \text{ for } t = 1, 2, ..., m_T, \\ y_{t-1} - y_{t-1-m_T}, \text{ for } t = m_T + 1, m_T + 2, ..., T. \end{cases}$$

Although the uniform window is the easiest to implement, the resulting tests are typically less powerful than using the Bartlett window. Through extensive simulations, we have examined several other possible windows, such as the Pazen, Quadratic, and other exponential type windows. It turns out that the results are not too sensitive to the choice of these windows, but we found that the Bartlett window works slightly better.

[9]The order of our IV test is $T^{(1+\delta)/2}$ with $0 \leq \delta < 1$. This result can be easily obtained from (i) and (ii) of Lemma 1 in the Appendix. In the extreme case when $\delta = 0$ (with $m = 1$), the order of the test is $\sqrt{T}$. However, we recommend using $\delta = 0.7$ to 0.9, depending on specific models. Then, the order of the test will be $T^{0.85} \sim T^{0.95}$. In these cases, the improvement in power is significant. The question is how the power of the IV tests is compared with the power of the OLS based tests whose order is T. As we will see in the next section, the power of the IV tests is comparable to the most powerful tests (say, GLS tests) in small samples as well as moderately large sample sizes, say $T = 1,000$, which should be reasonably large to see some influence of the large sample property of the tests. Note that our IV tests will have power against the local alternative of $\phi = 1 - c/T^{(1+\delta)/2}$. We will examine the power issue in more details in section 3.

2.2 Detrending and Bias Correction

In practice, a time series typically contains some deterministic components. In this paper, we consider the following three popular cases, but the IV tests can be utilized in other general models.

$$\begin{aligned} \text{Drift:} \quad & d_t = \gamma_0, \\ \text{Linear trend:} \quad & d_t = \gamma_0 + \gamma_1 t, \\ \text{Structural break:} \quad & d_t = \gamma_0 + \gamma_1 t + \gamma_2 D_t + \gamma_3 (t \times D_t), \end{aligned} \tag{9}$$

where D_t is the dummy variable that captures a break point. When a break occurs between $t = T_B$ and $T_B + 1$, we define

$$D_t = \begin{cases} 0, \ t < T_B + 1, \\ 1, \ t \geq T_B + 1. \end{cases} \tag{10}$$

The interaction term $t \times D_t$ allows for a possible slope change in the linear trend after the break; see Perron (1989). If we omit the interaction term $t \times D_t$ we obtain the model with level shift, which is a special case. Let z_t be the vector of the deterministic terms and γ be the corresponding coefficient vector. Therefore, we let $z_t = [1]$, $[1, t]$, and $[1, t, D_t, \Delta D_t, tD_t]$ for each of three deterministic trends in (9).

There are a few different detrending methods suggested in the literature. One popular method has been to utilize a regression in levels to estimate the coefficients of deterministic terms and use these coefficients to detrend the series. This is the detrending procedure adopted in the DF type tests. Alternatively, one may estimate the coefficients of deterministic terms from the regression with differenced data and use these coefficients to detrend the data. This detrending method was advocated by Bhargava (1986), and Schmidt and Phillips (1992). We call this the LM detrending method since we first impose the null restriction. Hwang and Schmidt (1996) and Elliott, Rothenberg and Stock (1996) adopted a method that is similar to the LM detrending, but allows for a local alternative. Specifically, we first rewrite equation (3) as

$$(1 - \phi L) y_t = (1 - \phi L) z_t \gamma + \varepsilon_t, \tag{11}$$

where $z_t \gamma = d_t$. Then, one can consider the following detrending methods

(i) DF detrending ($\phi = 0$)
(ii) LM detrending ($\phi = 1$)
(iii) GLS detrending ($\phi = a$, close to 1)

(iv) GLS detrending ($\phi = \phi_T = 1 - c/T$).

The difference among these methods lies in how we estimate γ. The DF detrending procedure amounts to detrending the data assuming that $\phi = 0$ in equation (11). On the other hand, the testing procedures for the LM test and the GLS tests are different. For these tests, γ is estimated using the differenced or quasi-differenced data and we detrend the data using the estimate of γ. Then, we run the AR(1) regression of $\Delta\tilde{y}_t$ on $\tilde{y}_{t-1}$, where $\tilde{y}_t$ is the detrended series; $\tilde{y}_t = y_t - z_t\hat{\gamma}$. If we detrend the data using the estimate of γ assuming $\phi = 1$, we have the LM tests developed by Bhargava (1986) and Schmidt and Phillips (1992). That is, γ is estimated from the regression of Δy_t on Δz_t, and we use this estimate of γ to detrend the series; $\tilde{y}_t = y_t - z_t\hat{\gamma}_{LM}$. Note that γ is estimated from using Δz_t, but the detrended series is driven from using z_t. The LM test statistic is obtained from the regression of $\Delta\tilde{y}_t$ on $\tilde{y}_{t-1}$. If we take some constant close to 1, say 0.98, for ϕ, we have the tests proposed by Hwang and Schmidt (1996). Thus, γ is estimated from the regression of $(y_t - \phi y_{t-1})$ on $(z_t - \phi z_{t-1})$, and we use this estimate of γ to detrend the series; $\tilde{y}_t = y_t - z_t\hat{\gamma}_{GLS}$. Very similarly, using $\phi_T = 1 - c/T$ with some constant c, leads to the local alternative GLS test of Elliott, Rothenberg and Stock (1996). In this case, γ is estimated from the regression of $(y_t - \phi_T y_{t-1})$ on $(z_t - \phi_T z_{t-1})$, and the detrended series is accordingly obtained as $\tilde{y}_t = y_t - z_t\hat{\gamma}_{GLS*}$. As in the LM testing procedure, we obtain the GLS test statistics from the regression of $\Delta\tilde{y}_t$ on $\tilde{y}_{t-1}$.

Any of these four detrending methods can be used for our IV tests. The resulting IV tests have the asymptotic standard normal distribution, while the OLS based tests have different nonstandard distributions. We conducted simulations to compare performance of the tests using different detrending methods. We found that the LM and any of the GLS detrending methods (either Hwang and Schmidt or Elliott, Rothenberg and Stock) produce virtually similar results with almost the same size and power properties. But the power is somewhat lower when the DF detrending method is adopted. In the following analysis we will focus on the LM detrending method.

In using the LM detrending method, γ_0 is not identified from the regression of Δy_t on Δz_t. To remove γ_0, we subtract y_0 from y_t, $t = 1, 2, ..., T$. In the conditional maximum likelihood approach, where y_0 is fixed, y_0 is the maximum likelihood estimator of γ_0; see, for example, Pantula, Gonzalez-Farias and Fuller (1994). Define γ^* as γ excluding γ_0, and z_t^* is z_t excluding [1]. In the cases of "linear trend" and "structural break", γ^* is estimated from the regression $\Delta y_t = \Delta z_t^* \gamma^* + error$, $t = 1, ..., T$. We then have the detrended series using the estimated coefficient $\hat{\gamma}_*$ from the regression in difference: $\tilde{y}_t = y_t - z_{*t}\hat{\gamma}_* - y_0$, for $t = 1, ..., T$. The IV unit root test is conducted from the AR(1) regression using the detrended series

$$\Delta\tilde{y}_t = \beta\tilde{y}_{t-1} + error \tag{12}$$

using the instrument

$$\tilde{w}_{t-1} = \begin{cases} \sum_{j=1}^{t-1} \psi_{t-j+1}\Delta\tilde{y}_j, \text{ for } t = 2, ..., m_T, \\ \sum_{j=1}^{m_T} \psi_{m_T+1-j}\Delta\tilde{y}_{t-(m_T+1)+j}, \text{ for } t = m_T + 1, m_T + 2, ..., T. \end{cases} \tag{13}$$

Note that when errors are serially correlated, we add the lagged augmented terms of $\Delta\tilde{y}_t$ in the testing regression, as in the ADF test. When m_T takes a relatively big number and a time trend exists, $E\left(\sum_{t=1}^{T} \tilde{w}_{t-1}\Delta\tilde{y}_t\right)$ is typically $O(m_T)$. Thus, we will need to fix the bias in practice in order to have the correct size. Accounting for the bias correction, we have the test statistic

$$t_{IV}^* = \frac{\sum_{t=2}^{T} \tilde{w}_{t-1}\Delta\tilde{y}_t - bias}{\hat{\sigma}\sqrt{\sum_{t=2}^{T} \tilde{w}_{t-1}^2}}, \tag{14}$$

where $bias$ is the consistent estimate of $E\left(\sum_{t=1}^{T} \tilde{w}_{t-1}\Delta\tilde{y}_t\right)$ under the null hypothesis.

Theorem 2 *Under Assumption 1 and the null hypothesis of $\beta = 0$ in equation (12), the t-statistic defined in (14) converges in distribution to the standard normal distribution*

$$t_{IV*} \xrightarrow{d} N(0, 1) \tag{15}$$

if $m_T = T^\delta$, $0 \leq \delta < 1$.

Proof. *See the Appendix.* ■

Thus, we obtain the same asymptotic normal result when deterministic terms are controlled for. The required condition for δ remains the same as in the case with no deterministic term. The only difference is the presence of the bias term in calculating the t-statistic. The bias correction is not needed in the basic model. We provide details of the bias term for each model in the following. Note that the bias correction is made on the score term, rather than on the coefficient estimators or the entire t-statistic.

2.2.1 Drift: $d_t = \gamma_0$

When a non-zero mean is allowed but no time trend exists, we have $d_t = \gamma_0$. Then, we can use $\tilde{y}_t = y_{t-1} - y_0$, for $t = 1, 2, ..., T$, in (12). In this case, $bias$ = 0. Thus, no bias correction is necessary. Simulation results in the next section suggest that the IV test performs best when the Bartlett window is used with $\delta = 0.9$.

2.2.2 Linear Trend: $d_t = \gamma_0 + \gamma_1 t$

When a linear trend is allowed for in the testing regression, we define $d_t = \gamma_0 + \gamma_1 t$. Then, we have $\hat{\gamma}_1 = \frac{1}{T}\sum_{t=1}^{T} \Delta y_t$, so that $\tilde{y}_t = y_t - t\hat{\gamma}_1 - y_0$, for $t = 1, ..., T$, can be used in (12). In this case, we have

$$bias = -\sigma_1^2 \left[\left(1 - \frac{m_T}{T}\right) \sum_{j=1}^{m_T} \psi_j + \frac{1}{T} \sum_{t=2}^{m_T} \sum_{j=2}^{t} \psi_{j-1} \right]. \tag{16}$$

In the Appendix, we show how the above result is derived. Note that $\sigma_1^2 = \sigma^2$ when errors are serially uncorrelated, and that in the presence of serially correlated errors, σ_1^2 is estimated as $\hat{\sigma}_1^2 = \hat{\sigma}^2/(1 - b(1))$, where $b(1)$ is the sum of the coefficients of the lagged augmented terms of $\Delta\tilde{y}_t$. When we use the uniform window, where $\psi_j = 1$, for all j, $\sum_{j=1}^{m_T} \psi_j = m_T$, and $\sum_{t=2}^{m_T} \sum_{j=2}^{t} \psi_{j-1} = \frac{1}{2} m_T(m_T - 1)$. Hence, the bias is simplified as

$$bias = -\sigma_1^2 \left[m_T - \frac{1}{2T} m_T(m_T + 1) \right]. \tag{17}$$

When the Bartlett window is used, where $\psi_j = 1 - \frac{j-1}{m_T}$, we can show that $\sum_{j=1}^{m_T} \psi_j = \frac{1}{2}(m_T + 1)$, $\sum_{t=2}^{m_T} \sum_{j=2}^{t} \psi_{j-1} = \frac{1}{3}\left(m^2 - 1\right)$, and

$$bias = -\sigma_1^2 \left[\frac{1}{2}(m_T + 1) - \frac{1}{6T}(m_T + 1)(m_T + 2) \right]. \tag{18}$$

We found from our simulations that the test performs best when the Bartlett window is used with $\delta = 0.7$.

2.2.3 Structural Break: $d_t = \gamma_0 + \gamma_1 t + \gamma_2 D_t + \gamma_3 (t \times D_t)$

When a time series involves a structural break, we have $d_t = \gamma_0 + \gamma_1 t + \gamma_2 D_t + \gamma_3 (t \times D_t)$. It may be informative to show detailed expressions of the detrending procedure. For this model, we have $z_t^* = [t, D_t, tD_t]$ and $\Delta z_t^* = [1, \Delta D_t, \Delta(t \times D_t)]$. Note that $\Delta D_t = 1$ at $t = T_B + 1$, and 0 elsewhere. Then, we obtain numerically identical estimators of γ_1 and γ_3 from the regression Δy_t on $[1, \Delta(t \times D_t)]$ omitting the row at $t = T_B + 1$ in the regression. We obtain $\hat{\gamma}_1 = \overline{\Delta y}_1$, $\hat{\gamma}_3 = \overline{\Delta y}_2 - \overline{\Delta y}_1$, where $\overline{\Delta y}_1 = \frac{1}{T_B} \sum_{t=1}^{T_B} \Delta y_t = (y_{T_B} - y_0)/T_B$ and $\overline{\Delta y}_2 = \frac{1}{T-(T_B+1)} \sum_{t=T_B+1}^{T} \Delta y_t = (y_T - y_{T_B+1})/[T - (T_B + 1)]$. Making use of the fact that the $(T_B + 1)$-th residual is zero when ΔD_t is included in the regression, we obtain $\hat{\gamma}_2 = \Delta y_{T_B+1} - \hat{\gamma}_1 - (T_B + 1)\hat{\gamma}_3$. Using these estimators, we can show that the expression for the detrended series $\tilde{y}_t = y_t - \hat{\gamma}_1 t - \hat{\gamma}_2 D_t - \hat{\gamma}_3 t D_t$ can be alternatively given as

$$\tilde{y}_t = \begin{cases} y_t - t\overline{\Delta y}_1 - y_0, & \text{for } t = 1, 2, ..., T_B \\ (y_t - y_{T_B+1}) - [t - (T_B + 1)]\, \overline{\Delta y}_2, & \text{for } t = T_B + 1, ..., T. \end{cases} \tag{19}$$

Note in particular that $\tilde{y}_{T_B} = \tilde{y}_{T_B+1} = \tilde{y}_T = \tilde{y}_0 = 0$ in the detrended data. The expression for the detrended series in (19) shows that the structural break completely splits the data into two uncorrelated series under the null hypothesis. Therefore, it is more reasonable to choose m_T based on T_B or $T - T_B$ rather than on T. Following our simulation, the size of the test is the most stable when m_T is chosen based on the maximum of T_B and $T - T_B$, namely

$$m_T = [\max(T_B, T - T_B)]^{\delta}. \tag{20}$$

However, note that two split series are correlated under the alternative hypothesis. The bias term is given as

$$bias = -\sigma_1^2 \begin{bmatrix} \left(2 - \frac{m_T}{T_B} - \frac{m_T}{T-T_B-1}\right) \sum_{j=1}^{m_T} \psi_j + \\ \left(\frac{1}{T_B} + \frac{1}{T-T_B-1}\right) \sum_{t=2}^{m_T} \sum_{j=2}^{t} \psi_{j-1} \end{bmatrix}, \tag{21}$$

which is derived in the Appendix. Specifically, when we use the uniform window, it becomes

$$bias = -\sigma_1^2 \left[2m_T - \frac{T-1}{2T_B(T - T_B - 1)} m_T(m_T + 1) \right]. \tag{22}$$

When the Bartlett window is used, the bias is given as

$$bias = -\sigma_1^2 (m_T + 1) \left[1 - \frac{(T-1)(m_T + 2)}{6T_B(T - T_B - 1)} \right]. \tag{23}$$

We note in passing that when there are two structural breaks the expression for the bias will be

$$bias = -\sigma_1^2 \left[\left(3 - \frac{m_T}{T_1} - \frac{m_T}{T_2} - \frac{m_T}{T_3} \right) \sum_{j=1}^{m_T} \psi_j + \left(\frac{1}{T_1} + \frac{1}{T_2} + \frac{1}{T_3} \right) \sum_{t=2}^{m_T} \sum_{j=2}^{t} \psi_{j-1} \right], \tag{24}$$

where T_1, T_2, T_3 denotes the number of data points before the first break, between two breaks and after the second break. A general pattern of the bias will be obvious when there are three breaks or more.[10]

3 Simulations

In this section, we investigate the small sample properties of the IV unit root tests through Monte Carlo simulations. We compare our IV tests with the three widely used tests, DF, LM and GLS tests, for each of three models with different deterministic terms. These tests are based on the OLS estimation. The data generating process implies (1) and (2). In our simulations, d_t is generated by using $\gamma_0 = 10$, $\gamma_1 = 1$, $\gamma_2 = 5$, and $\gamma_3 = 1$ whenever the corresponding deterministic components are included. We note that all tests are invariant to the parameter γ in the corresponding DGP for which we use $d_t = z_t\gamma$. Thus, using $\gamma = 0$ or any other values of γ in the DGP will not change the simulation results when the corresponding expression of z_t is used for each of these models. We use pseudo-*iid* $N(0,1)$ random numbers from the Gauss 6.0.10 RNDNS procedure.[11]

We examine the cases for $\phi = 1$, $0.9, 0.95$ and 0.99 with different sample sizes, $T = 50, 100, 200, 300$, 500, and 1000. Under the null hypothesis with $\phi = 1$, all tests are invariant to the initial value x_0. We can take x_0 simply as a random draw from a standard normal distribution. However, power hinges on the value of x_0. For the results in Tables 1-3, x_0 is given as a random draw from the normal distribution with mean zero and variance $1/\left(1-\phi^2\right)$

[10] We do not pursue the issue of endogenous breaks in this paper. As discussed in the previous section, existing endogenous tests entail certain limitations and difficulties. Given that our IV tests do not depend on the nuisance parameter, it is possible that an extension to endogenous version tests will involve fewer problems and provide better results. However, this issue is usually dealt with separately as the nature of the problems is quite different. Such extensions are best left for future research.

[11] Gauss codes used for our simulations are available from the authors upon request. The calculation of the bias term is rather simple.

to ensure that the variance of x_0 matches with that of other x_t series. In Table 4, we vary the value of x_0 to see the effect of the initial values on the power of the tests.

We report the size, power and the size-adjusted power of the tests at the 5% significance level. We only report the results when m depends on T, but the results using a finite value of m are also available. The power of the tests is computed as the proportion of rejections using the asymptotic critical values. For all IV tests, we use the asymptotic one-sided critical value -1.645 of the standard normal distribution. For other tests, we use the critical values reported in the literature. The size-adjusted power is calculated based on the simulated critical values under the null, which are generated by using 50,000 replications for each case. The size of the tests is calculated from 50,000 replications, and the power and the size-adjusted power are calculated from 10,000 replications. The size-adjusted powers are reported in the parentheses in Tables 1 - 3. In Table 4, all figures are size-adjusted powers.

Table 1 presents the simulation results for the "drift" model ($d_t = \gamma_0$). When there is no time trend, we do not have an LM version test. For the GLS test, we follow Elliott *et al.* (1996) and use $c = 7$, which ensures the 50% asymptotic power against the local alternative $\phi = 1 - c/T$. For the IV tests, we report results from using both the uniform window and the Bartlett window with $m_T = T^{\delta}$ for $\delta = 0.7$, 0.8, and 0.9. These values of δ have been selected from intensive simulations. As seen in Table 1, there is a slight tendency to over-rejections in the IV tests when the sample size is small. The size of the IV tests ranges from 7% to 9% regardless of different types of windows and different values of δ . The size of the tests improves as T gets larger, but the speed of improvement is slow. Although the over-size tendency is a little bit stronger when $\delta = 0.9$, the power is the best when $\delta = 0.9$. As is well known, GLS is shown to be more powerful than DF, especially when $T = 50$ and $T = 100$. But, the comparison is rather mixed when T gets larger, depending on how close $1 - c/T$ is to the true value of ϕ. In contrast, the IV tests based on the Bartlett window with $\delta = 0.9$ are shown to be more powerful in almost all cases. For example, when $T = 100$ and $\phi = 0.9$, the size-adjusted power of the DF, GLS, and IV tests are $0.341, 0.495$ and 0.512, respectively. When $T = 200$ and $\phi = 0.9$, they are $0.877, 0.840$ and 0.936, respectively. This favorable result for the IV tests over the GLS test may be rather surprising since the GLS test is commonly known to be the most powerful test.

Table 2 reports the result for the "linear trend" model ($d_t = \gamma_0 + \gamma_1 t$).

For the GLS test, we use $c = 13.5$ and the small sample critical values as suggested by Elliott *et al.* (1996). Again, the Bartlett window-based IV tests are significantly more powerful than the tests based on the uniform window-based IV tests. The size is the closest to the nominal size when $\delta = 0.7$ in all cases with different T values. The size ranges from 5.3% to 5.7%. Hence, we recommend the Bartlett window in practice with $\delta = 0.7$. Although the power of the DF test remains somewhat lower than the power of the other three tests, the discrepancy is much smaller when compared with the "drift" model. It is interesting to see that the powers of the LM, GLS and IV tests are virtually in a tie with most cases, with a slight edge for GLS. The favorable result in power of the IV tests over the GLS test in the model with drift largely disappears.

In Table 3, we report results for the "trend shift" model ($d_t = \gamma_0 + \gamma_1 t + \gamma_2 D_t + \gamma_3(t \times D_t)$) for two cases with $T_B = 0.5T$ and $T_B = 0.2T$[12]. Since the asymptotic distributions of the DF, LM and GLS tests depend on the location of the breaks, the critical values vary for these tests and no corresponding critical values are readily available. Thus, only the size-adjusted power is reported for these tests. Our main interest is to compare the sensitivity of the IV tests to the DGP with different locations of breaks, and to examine the generic power of the IV tests. For the GLS test, we use $c = 22.5$ which ensures the 50% asymptotic power in the local alternatives of $\phi = 1 - c/T$, as suggested by Perron and Rodriguez (2003). Note that we use $m_T = [\max(T_B, T - T_B)]^\delta$ for the IV tests. In both cases with $T_B = 0.5T$ and $T_B = 0.2T$, the results on the size and power of the IV tests are very similar, implying that the IV tests are free of the nuisance parameter, $\lambda = T_B/T$, indicating the location of the break[13]. The size of the test is the closest to the nominal size when $\delta = 0.7$. Looking at the power, we observe that the Bartlett window provides better power than the uniform window for the IV tests. Comparing the size-adjusted power of the tests, we observe that the GLS test is more powerful than others. The powers of the IV tests and the LM tests are very

[12] We also simulated the cases when $T_B = 0.8T$, but the results are almost identical to those from the case with $T_B = 0.2T$. These results are not reported to conserve space.

[13] This is an encouraging result, since no existing test with trend-shift is invariant to nuisance parameters. The trend-shift model involves both level and trend shifts. As a special case, we also considered the IV test with level shift only. The results are similar to those in Table 2 and are omitted here. The LM tests with level shift already have the invariance property that they are free of the nuisance parameter; see Amsler and Lee (1995).

close, and the DF test shows the least power.

Overall, in most cases, the powers of the IV tests are fairly comparable to the GLS test. This result is encouraging since the IV tests have the additional important feature that their distributions are standard normal and free of nuisance parameters. This desirable property is obtained without a sacrifice in power. Clearly, the IV tests have an operating advantage. It is not necessary for the IV tests to simulate different critical values for various combinations of structural break locations, while such a task is necessary to employ the GLS, LM and DF extension tests. This advantage will become more useful in extended models with multiple breaks and other models.

In Figure 1, we provide the estimated empirical pdfs of various tests under the null hypothesis when $T = 100$. These are based on the kernel estimation using a Gaussian kernel function. Using other kernel functions does not make a big difference. The solid curve depicts the pdf of the standard normal distribution. The curve with short dashes in the upper graph denotes the pdf of the IV statistics of the trend model with $\delta = 0.7$. It is seen that the pdf of the IV statistic is close to that of the standard normal distribution. The pdf of the IV statistic is clearly far off the left-shifted pdfs of the LM (dashed curve) and GLS (dots and dashed) statistics which are based on the OLS estimation. In the lower graph in Figure 1, we provide the pdfs of the IV statistics using different parameter values of $\lambda = 0.5, 0.2$ and 0.8. The pdfs of the IV statistics using $\lambda = 0.2$ and 0.8 overlap exactly and are hardly separable. The pdf of the IV statistic using $\lambda = 0.5$ almost overlaps the pdfs of the IV statistics using $\lambda = 0.2$ and 0.8. These pdfs of the IV statistics are very close regardless of different values of λ, implying the invariance property of the IV tests using different break locations. For the left-shifted pdfs of the LM and the GLS statistics, we use the model with trend shift using the break location parameter $\lambda = 0.5$. The shape and location of these pdfs are different from those in the upper graph of the trend model. This observation is obvious since they are not invariant to λ. On the other hand, for the IV tests, the shape and the location of the pdf (lower graph) of the trend-shift model are similar to those of the pdf (upper graph) of the trend model. In Figure 2, we examine the case with a large sample. We report the estimated pdfs when $T = 500$. The results are similar to those in Figure 1, except that the pdfs of the IV statistics are somewhat closer to the pdf of the standard normal distribution when the sample size is bigger.

We next examine how effectively the augmented version test controls for the effect of autocorrelated errors. We have examined the cases with the

DGP of AR(1) errors and AR(2) errors; $\varepsilon_t = a_1\varepsilon_{t-1} + a_2\varepsilon_{t-2} + u_t$, where u_t is serially uncorrelated with a constant variance. For AR(1) errors, we use the values of $a_1 = 0.8, 0.5, 0.0, -0.5$, and -0.8, and $a_2 = 0$. For the DGP with AR(2) errors, the coefficients are given from $a_1 = c_1 + c_2$, $a_2 = -c_1c_2$,where c_1 and c_2 are the roots of $(\lambda - c_1)(\lambda - c_2) = 0$. We use the values of $(c_1, c_2) = (-0.5, -0.8), (-0.5, 0.8), (0.5, 0.3)$ and $(-0.5, -0.3)$; then, we have $(a_1, a_2) = (-0.3, 0.4), (0.3, 0.4), (0.8, -0.15)$ and $(-0.8, 0.15)$, respectively. For all cases, the IV tests with the Bartlett windows and $\delta = 0.7$ are used. We report the results for the model with a linear trend, but similar results on the effect of serially correlated errors are expected for other models. Table 4 provides the results when $p = 1, 2$, and 3 are used. We are more interested in the cases under the null. Looking at the results with AR(1) errors, we observe that the size properties are fairly reasonable, regardless of different AR(1) coefficients in the DGP. This is so in large samples equal to or greater than 200, although we observe a little size distortion when $T = 100$. In the presence of AR(2) errors, we can see that the IV tests exhibit the same pattern of good size properties when enough augmented lagged terms are used ($p \geq 2$). It is obvious that the tests will show size distortions when $p = 1$ is used. Using more augmentation terms than necessary (that is, using $p \geq 2$ for AR(1) errors, or using $p \geq 3$ for AR(2) errors) can lead to proper sizes. However, we observe a loss of power as compared with the result using the optimal value of p (that is, $p = 1$ for AR(1) errors, and $p = 2$ for AR(2) errors). In general, the order of AR terms is unknown, and the usual data dependent lag selection procedures can be similarly applied to our IV tests; see Ng and Perron (1995) in this regard.

In Table 5, we examine the size-adjusted power of the tests in two models with "drift" and "trend" when the initial values vary. It is known that the power of the GLS tests drops as the initial value of the stochastic process, x_0, gets bigger in absolute terms, while the DF test becomes more powerful; see DeJong *et al.* (1992), Hwang and Schmidt (1996), and Müller and Elliott (2003) for simulation results and detailed discussion. The same result is expected from the LM tests. A similar pattern of power loss is observed for the IV tests as x_0 gets bigger. This is so, since the IV tests are based on the LM detrending method. The IV tests based on the GLS detrending will be subject to the same result. We report results for the IV tests with $\delta = 0.9$ for the "drift" model, and $\delta = 0.7$ for the "trend" model. When $x_0 = 0$, the GLS test is the most powerful in both cases. However, the power of the GLS test drops the most rapidly and more so in the drift model where no time trend is present. When x_0 is non-zero such that $x_0 = 2\sigma/\sqrt{1-\phi^2}$, the GLS test shows the least power. When the initial value problem appears significant, one may alternatively employ the DF type detrending method for the IV tests. A recent paper by Harvey and Leybourne (2005) suggests using a weighted average of the DF and the GLS test. Such a treatment would be possible in the IV framework.

4 Summary and Concluding Remarks

In this paper, we have developed new unit root tests using stationary instrumental variables. Our new unit root tests utilize moment conditions different from those of existing unit root tests. The asymptotic distributions of our IV test statistics are standard normal under the null hypothesis. As such, in light of our new testing procedures it is possible to perform valid statistical inference based on standard distribution theory to test for a unit root. This result appears pretty general in the sense that the result holds in models with differing deterministic terms or using different detrending methods. Also, it is encouraging to see that the normality result is obtained without incurring any notable loss of power. Our IV tests are more powerful than the popular DF tests, and are comparable to the LM tests of Schmidt and Phillips (1992) or the GLS based tests suggested by Elliott, Rothenberg and Stock (1996) and Hwang and Schmidt (1996).

An important feature of our IV unit root tests is that they are free of nuisance parameters. This desirable invariance property will be useful in other extended models, while we have demonstrated this point, as an immediate extension of the IV test, with the models involving structural changes. The IV unit root tests are invariant to nuisance parameters even in the models with trend shift. This case is one important illustration of applying our IV tests to resolve the nuisance parameter dependence problem. The same invariance result holds in models with multiple breaks. There are other important circumstances where the invariance property will be useful. One example includes the model with stationary covariates. As shown by Hansen (1995), the use of stationary covariates can enhance the power of unit root tests. The distribution of the usual tests utilizing stationary covariates becomes dependent on the nuisance parameter. However, the asymptotic distribution of our IV tests using stationary covariates will remain as standard normal. In addition, applying our IV tests to cointegration models appears promising.

The OLS based ECM statistics for cointegration in a single equation model depend on a nuisance parameter; see Kremers, Ericsson and Dolado (1992). The asymptotic distribution of the IV based cointegration ECM test becomes standard normal. This topic is being pursued in a separate work. We expect that other important extensions are possible in seasonality models, dynamic system models, and nonlinear threshold models. Furthermore, when extending our tests to panel unit root tests (with or without breaks), there is an additional advantage. The present paper has already shown that asymptotic normality is obtained when $N = 1$. We leave these topics to future research.

References

[1] Amsler, C. and J. Lee (1995), "An LM Test for a Unit-Root in the Presence of a Structural Change," *Econometric Theory*, 11, 359-368.

[2] Bhargava, A. (1986), "On the Theory of Testing for Unit Roots in Observed Time Series," *Review of Economic Studies*, 52, 369–384.

[3] Dickey, D.A. & W.A. Fuller (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427-431.

[4] DeJong, D., J. Nankervis, N. Savin, and C. Whiteman (1992), "The Power Problems of Unit Root Tests in Time Series with Autoregressive Errors, *Journal of Econometrics*, 53, 323-343.

[5] Elliott, G., T.J. Rothenberg, and J. Stock (1996), "Efficient Tests for an Autoregressive unit Root," *Econometrica*, 64:4, 813-836.

[6] Hall, A. (1989), "Testing for a Unit Root in the Presence of Moving Average Errors," *Biometrika*, 76, 49-56.

[7] Hamilton, J., 1994, Time Series Analysis, Princeton University Press.

[8] Hansen, B. (1995), "Rethinking the univariate approach to unit root tests: How to use covariates to increase power," *Econometric Theory*, 11, 1148-1171.

[9] Harvey, D. and S. Leybourne (2005), "On testing for Unit Roots and the Initial Observation," Working Paper, School of Economics, University of Nottingham.

[10] Harris, D., B. McCabe and S. Leybourne (2003), "Some Limit Theory for Autocovariances Whose Order Depends on Sample Size," Econometric Theory, 19, 829–864.

[11] Hwang, J, and P. Schmidt (1996), "Alternative Methods of Detrending and the Power of Unit Root Tests", *Journal of Econometrics*, 71, 227-248.

[12] Kremers, J.J.M., N.R. Ericsson and J.J. Dolado (1992), "The Power of Cointegration Tests," *Oxford Bulletin of Economics and Statistics*, 54:3, 325-348.

[13] Lee, J. and P. Schmidt (1994), "Unit Root Tests Based on Instrumental Variables Estimation," *International Economic Review*, 35:2, 449-462.

[14] Lee, J. and M.C. Strazicich (2001), "Break Point Estimation and Spurious Rejections with Endogenous Unit Root Tests," *Oxford Bulletin of Economics and Statistics* 63:5, 535-558.

[15] Müller, U., and G. Elliott (2003), "Tests for Unit Roots and the Initial Condition," *Econometrica*, 71, 1269-1286.

[16] Ng, S. and P. Perron (1995), "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association* 90, 269-281.

[17] Nunes, L., P. Newbold, and C.M. Kuan (1997), "Testing for Unit Roots with Breaks: Evidence on the Great Crash and the Unit Root Hypothesis Reconsidered," *Oxford Bulletin of Economics and Statistics* 59:4, 435-448.

[18] Pantula, S. G., G. Gonzalez-Farias and W. Fuller (1994), "A Comparison of Unit Root Test Criteria, *Journal of Business and Economic Statistics*, 12, 449-459.

[19] Perron, Pierre (1989), "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica* 57:6, 1361-1401.

[20] Perron, P. and G. Rodriguez (2003), "GLS Detrending, Efficient Unit Root Tests and Structural Change," *Journal of Econometrics*, 115, 1-27.

[21] Phillips, P.C.B, J. Y. Park and Y. Chang (2004), "Nonlinear Instrumental Variable Estimation of an Autoregression," *Journal of Econometrics*," 118, 219-246.

[22] Saikkonen, P., and H. Lütkepohl (2002), "Testing For a Unit Root In a Time Series With a Level Shift at Unknown Time," *Econometric Theory*, 313-348.

[23] Schmidt, P. and P.C.B. Phillips (1992), "LM Tests for a Unit Root in the Presence of Deterministic Trends," *Oxford Bulletin of Economics and Statistics*, 54, 257-287

[24] So, B.S. and D.W. Shin (1999), "Cauchy Estimators for Autoregressive Processes with Applications to Unit Root Tests and Confidence Intervals," *Econometric Theory*, 15, 165-176.

[25] Sul, D., P.C.B. Phillips and C. Choi (2003), "Prewhitening Bias in HAC Estimation," manuscript.

[26] Vougas, D.V. (2003), "Reconsidering LM Unit Root Testing," *Journal of Applied Statistics*, 30, 727-741.

[27] White, H. (2001), Asymptotic Theory for Econometricians, Academic Press.

5 Appendix

Lemma 1. Let

$$\xi_{t-1} = \begin{cases} \sum_{j=1}^{t-1} \psi_{t-j}\varepsilon_j, \text{ for } t = 2, ..., m_T, \\ \sum_{j=1}^{m_T} \psi_{m_T+1-j}\varepsilon_{t-(m_T+1)+j}, \text{ for } t = m_T+1, m_T+2, ..., T, \end{cases} \tag{25}$$

where $m_T = T^\delta$, for $0 \le \delta < 1$, $0 < \psi_j \le 1$, for $j = 1, 2, ..., m_T$. Define $\bar{\psi}_2 \equiv \frac{1}{m_T}\sum_{j=1}^{m_T}\psi_j^2$ so that $0 < \bar{\psi}_2 \le 1$. Then, under Assumption 1

$$(i) \qquad T^{-(1+\delta)}\sum_{t=2}^{T}\xi_{t-1}\varepsilon_t \xrightarrow{p} 0,$$

$$(ii) \qquad Var\left[T^{-(1+\delta)/2}\sum_{t=2}^{T}\xi_{t-1}\varepsilon_t\right] = \bar{\psi}_2\sigma^4 + O\left(T^{-(1-\delta)}\right),$$

$$(iii) \qquad T^{-(1+\delta)}\sum_{t=2}^{T}\xi_{t-1}^2 \xrightarrow{p} \bar{\psi}_2\sigma^2 + o_p(1).$$

Proof of Lemma 1. Let Γ_T be the $(T-1)\times(T-1)$ lower triangular matrix

$$\Gamma_T = \begin{bmatrix} \psi_1 & & & & & & \\ \psi_2 & \psi_1 & & & & & \\ \vdots & \psi_2 & \ddots & & & & \\ \psi_{m_T} & \vdots & \ddots & \psi_1 & & & \\ 0 & \psi_{m_T} & \cdots & \psi_2 & \psi_1 & & \\ \vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & \psi_{m_T} & \cdots & \psi_2 & \psi_1 \end{bmatrix}. \tag{26}$$

Also, we let $\boldsymbol{\xi}_{-1} = (\xi_1, \xi_2, ..., \xi_{T-1})'$, $\boldsymbol{\varepsilon}_{-1} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_{T-1})'$, and $\boldsymbol{\varepsilon} = (\varepsilon_2, \varepsilon_3, ..., \varepsilon_T)'$. Then, we can show $\boldsymbol{\xi}_{-1} = \Gamma_T\boldsymbol{\varepsilon}_{-1}$ and $\sum_{t=2}^{T}\xi_{t-1}\varepsilon_t = \boldsymbol{\varepsilon}'_{-1}\Gamma'_T\boldsymbol{\varepsilon}$. Since $E\left[\sum_{t=2}^{T}\xi_{t-1}\varepsilon_t\right] = 0$, we have

$$\begin{aligned} Var\left(\sum_{t=2}^{T}\xi_{t-1}\varepsilon_t\right) &= Var\left(\boldsymbol{\varepsilon}'_{-1}\Gamma'_T\boldsymbol{\varepsilon}\right) = \sigma^4 tr\left(\Gamma'_T\Gamma_T\right) \\ &= \sigma^4\sum_{j=1}^{m_T}(T-j)\psi_j^2 = \sigma^4\left(T\sum_{j=1}^{m_T}\psi_j^2 - \sum_{j=1}^{m_T}j\psi_j^2\right). \end{aligned} \tag{27}$$

The second result follows since $T\sum_{j=1}^{m_T}\psi_j^2 = Tm_T\bar{\psi}_2 = O_p\left(T^{1+\delta}\right)$, and $\sum_{j=1}^{m_T} j\psi_j^2 = O_p\left(m_T^2\right) = O_p\left(T^{2\delta}\right)$. The first result follows from the Chebyshev's inequality in view of the second result and $E\left[\sum_{t=2}^{T}\xi_{t-1}\varepsilon_t\right] = 0$. To verify the third result, note that

$$\xi_{t-1}^2 = \begin{cases} \sum_{j=1}^{t-1}\psi_j^2\varepsilon_{t-j}^2 + 2\sum_{j=2}^{t-1}\sum_{i=1}^{j-1}\psi_i\psi_j\varepsilon_{t-i}\varepsilon_{t-j}, \text{ for } t = 2,3,...,m_T, \\ \sum_{j=1}^{m_T}\psi_j^2\varepsilon_{t-j}^2 + 2\sum_{j=2}^{m_T}\sum_{i=1}^{j-1}\psi_i\psi_j\varepsilon_{t-i}\varepsilon_{t-j}, \text{ for } t = m_T+1, m_T+2, ..., T. \end{cases}$$

Hence,

$$T^{-(1+\delta)}\sum_{t=2}^{T}\xi_{t-1}^2 = T^{-(1+\delta)}\left(\sum_{t=2}^{m_T}\sum_{j=1}^{t-1}\psi_j^2\varepsilon_{t-j}^2 + \sum_{t=m_T+1}^{T}\sum_{j=1}^{m_T}\psi_j^2\varepsilon_{t-j}^2\right)$$
$$+2T^{-(1+\delta)}\left(\sum_{t=3}^{m_T}\sum_{j=2}^{t-1}\sum_{i=1}^{j-1}\psi_i\psi_j\varepsilon_{t-i}\varepsilon_{t-j} + \sum_{t=m_T+1}^{T}\sum_{j=2}^{m_T}\sum_{i=1}^{j-1}\psi_i\psi_j\varepsilon_{t-i}\varepsilon_{t-j}\right).$$

The first term is $O_p(1)$. Next, we want to show that the second term is negligible asymptotically. In the second term, straightforward algebra shows that

$$\sum_{t=3}^{m_T}\sum_{j=2}^{t-1}\sum_{i=1}^{j-1}\psi_i\psi_j\varepsilon_{t-i}\varepsilon_{t-j} = O_p\left(m_T^2\right) = O_p\left(T^{2\delta}\right).$$

Therefore,

$$T^{-(1+\delta)}\sum_{t=3}^{m_T}\sum_{j=2}^{t-1}\sum_{i=1}^{j-1}\psi_i\psi_j\varepsilon_{t-i}\varepsilon_{t-j} = O_p\left(T^{-(1-\delta)}\right).$$

The expression $\sum_{t=m_T+1}^{T}\sum_{j=2}^{m_T}\sum_{i=1}^{j-1}\psi_i\psi_j\varepsilon_{t-i}\varepsilon_{t-j}$ contains $\frac{1}{2}(T-m_T)m_T(m_T-1)$ terms with the maximum $m_T - 1$ repetitions for each of the cross product terms $\varepsilon_i\varepsilon_j$ (with different coefficients). Assumption 1 ensures that all the cross product terms are uncorrelated unless they are the same, and that the variance of $\varepsilon_i\varepsilon_j$ is finite for all i and j. Then, it follows that

$$\sum_{t=m_T+1}^{T}\sum_{j=2}^{m_T}\sum_{i=1}^{j-1}\psi_i\psi_j\varepsilon_{t-i}\varepsilon_{t-j} = O_p\left(T^{\delta+\frac{1}{2}(1+\delta)}\right),$$

and

$$T^{-(1+\delta)}\sum_{t=m_T+1}^{T}\sum_{j=2}^{m_T}\sum_{i=1}^{j-1}\psi_i\psi_j\varepsilon_{t-i}\varepsilon_{t-j}=O_p\left(T^{-\frac{1}{2}(1-\delta)}\right).$$

The proof is complete if we show

$$T^{-(1+\delta)}\left(\sum_{t=2}^{m_T}\sum_{j=1}^{t-1}\psi_j^2\varepsilon_{t-j}^2+\sum_{t=m_T+1}^{T}\sum_{j=1}^{m_T}\psi_j^2\varepsilon_{t-j}^2\right)\xrightarrow{p}\bar{\psi}_2\sigma^2.$$

For this, we have

$$\begin{aligned}
& T^{-(1+\delta)}\left(\sum_{t=2}^{m_T}\sum_{j=1}^{t-1}\psi_j^2\varepsilon_{t-j}^2+\sum_{t=m_T+1}^{T}\sum_{j=1}^{m_T}\psi_j^2\varepsilon_{t-j}^2\right)\\
= \; & T^{-(1+\delta)}\left(\sum_{j=1}^{m_T}\psi_j^2\sum_{t=1}^{T-m_T}\varepsilon_t^2+\sum_{j=1}^{m_T-1}\psi_j^2\varepsilon_{T-j}^2\right)\\
= \; & T^{-(1+\delta)}\sum_{j=1}^{m_T}\psi_j^2\sum_{t=1}^{T}\varepsilon_t^2+O_p(T^{-(1-\delta)})\xrightarrow{p}\bar{\psi}_2\sigma^2.
\end{aligned}$$

The last equality is due to the martingale laws of large numbers.

Proof of Theorem 1. Under the null hypothesis, we have

$$t_{IV}=\frac{\sum_{t=1}^{T}w_{t-1}\Delta y_t}{\hat{\sigma}\sqrt{\sum_{t=1}^{T}w_{t-1}^2}}=\frac{\sum_{t=2}^{T}\xi_{t-1}\varepsilon_t}{\hat{\sigma}\sqrt{\sum_{t=2}^{T}\xi_{t-1}^2}},$$

where ξ_{t-1} is defined in Lemma 1. Since $\left\{\xi_{t-1}\varepsilon_t\right\}_{t=2}^{\infty}$ is a martingale difference sequence, and

$$Var\left(T^{-(1+\delta)/2}\sum_{t=2}^{T}\xi_{t-1}\varepsilon_t\right)=\bar{\psi}_2\sigma^4+o\left(1\right),$$

which is consistently estimated by $\left(T^{-(1+\delta)}\sum_{t=2}^{T}\xi_{t-1}^2\right)\hat{\sigma}^2$ as shown in Lemma 1, the martingale CLT applies to prove that t_{IV} converges to a standard normal distribution; see, for example, White (1999, Corollary 5.26).

Proof of Theorem 2. The t-statistic, t_{IV}, is invariant to the value of γ. We assume without loss of generality $\gamma=0$. Therefore, $\Delta\tilde{y}_t=\Delta x_t-\Delta z_{*t}\hat{\gamma}_*$, and $\hat{\gamma}_* = O_p(T^{-1/2})$. The bias term can be ignored asymptotically since $bias = O(m_T) = O(T^\delta)$. Let $\Delta\mathbf{z}_* = (\Delta z_{*1}, \Delta z_{*2}, ..., \Delta z_{*T})'$. Under the null hypothesis

$$\begin{aligned} t_{IV} &= \frac{\sum_{t=2}^T \tilde{w}_{t-1}\Delta\tilde{y}_t - bias}{\hat{\sigma}\sqrt{\sum_{t=2}^T \tilde{w}_{t-1}^2}} = \frac{\sum_{t=2}^T \tilde{w}_{t-1}\Delta\tilde{y}_t}{\hat{\sigma}\sqrt{\sum_{t=2}^T \tilde{w}_{t-1}^2}} + o_p(1) \\ &= \frac{\left(\boldsymbol{\xi}_{-1} - \Gamma_T \Delta\mathbf{z}_*\hat{\gamma}_*\right)'(\boldsymbol{\varepsilon} - \Delta\mathbf{z}_*\hat{\gamma}_*)}{\hat{\sigma}\sqrt{\left(\boldsymbol{\xi}_{-1} - \Gamma_T\Delta\mathbf{z}_*\hat{\gamma}_*\right)'\left(\boldsymbol{\xi}_{-1} - \Gamma_T\Delta\mathbf{z}_*\hat{\gamma}_*\right)}} + o_p(1). \end{aligned}$$

Dividing the both numerator and the denominator by $\sqrt{\boldsymbol{\xi}_{-1}'\boldsymbol{\xi}_{-1}}$, we have for the numerator

$$\frac{\boldsymbol{\xi}_{-1}'\boldsymbol{\varepsilon}}{\sqrt{\boldsymbol{\xi}_{-1}'\boldsymbol{\xi}_{-1}}} - \frac{\boldsymbol{\xi}_{-1}'\Gamma_T\Delta\mathbf{z}_*\hat{\gamma}_*}{\sqrt{\boldsymbol{\xi}_{-1}'\boldsymbol{\xi}_{-1}}} - \frac{\hat{\gamma}_*'\Delta\mathbf{z}_*'\boldsymbol{\varepsilon}}{\sqrt{\boldsymbol{\xi}_{-1}'\boldsymbol{\xi}_{-1}}} + \frac{\hat{\gamma}_*'\Delta\mathbf{z}_*'\Gamma_T'\Delta\mathbf{z}_*\hat{\gamma}_*}{\sqrt{\boldsymbol{\xi}_{-1}'\boldsymbol{\xi}_{-1}}}.$$

The first term is $O_p(1)$. Note that $\Delta\mathbf{z}_{*t} = (1, D_t, \Delta D_t)$. Thus, the first column of $\boldsymbol{\xi}_{-1}'\Gamma_T\Delta\mathbf{z}_*$ takes the highest order term. Letting ι_T be the $T \times 1$ vector, we have $\boldsymbol{\xi}_{-1}'\Gamma_T\iota_T = O_p(T^{\frac{1+\delta}{2}})$. Hence, $\boldsymbol{\xi}_{-1}'\Gamma_T\Delta\mathbf{z}_*\hat{\gamma}_* = O_p(T^{\frac{\delta}{2}})$, and the second term is $O_p(T^{-1/2})$. Similarly, it is straightforward to show that the third and fourth terms are $o_p(1)$. The terms inside of the square root of the denominator becomes

$$1 - 2\frac{\boldsymbol{\xi}_{-1}'\Gamma_T\Delta\mathbf{z}_*\hat{\gamma}_*}{\boldsymbol{\xi}_{-1}'\boldsymbol{\xi}_{-1}} + \frac{\hat{\gamma}_*'\Delta\mathbf{z}_*'\Gamma_T'\Gamma_T\Delta\mathbf{z}_*\hat{\gamma}_*}{\boldsymbol{\xi}_{-1}'\boldsymbol{\xi}_{-1}}.$$

Following similar analysis, we can have $\frac{\boldsymbol{\xi}_{-1}'\Gamma_T\Delta\mathbf{z}_*\hat{\gamma}_*}{\boldsymbol{\xi}_{-1}'\boldsymbol{\xi}_{-1}} = o_p(1)$ and $\frac{\hat{\gamma}_*'\Delta\mathbf{z}_*'\Gamma_T'\Gamma_T\Delta\mathbf{z}_*\hat{\gamma}_*}{\boldsymbol{\xi}_{-1}'\boldsymbol{\xi}_{-1}} = o_p(1)$. The desired result follows in view of Theorem 1.

Bias Term in Equation (16): We let $\boldsymbol{\varepsilon}_{-1} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_{T-1})'$, and $\boldsymbol{\varepsilon} = (\varepsilon_2, \varepsilon_3, ..., \varepsilon_T)'$; they are defined in the proof of Lemma 1. Let $\bar{\varepsilon} = \frac{1}{T}\sum_{t=1}^T \varepsilon_t$ and ι_{T-1} be the $(T-1) \times 1$ column vector of ones. We then have, under the null hypothesis

$$\begin{aligned} & E\left(\sum_{t=2}^T \tilde{w}_{t-1}\Delta\tilde{y}_t\right) = E\left[(\boldsymbol{\varepsilon} - \bar{\varepsilon}\iota_{T-1})'\,\Gamma_T'\,(\boldsymbol{\varepsilon}_{-1} - \bar{\varepsilon}\iota_{T-1})\right] \\ =\ & E\left[\boldsymbol{\varepsilon}'\Gamma_T'\boldsymbol{\varepsilon}_{-1} - \bar{\varepsilon}\left(\boldsymbol{\varepsilon}'\Gamma_T'\iota_{T-1}\right) - \bar{\varepsilon}\left(\iota_{T-1}'\Gamma_T'\boldsymbol{\varepsilon}\right) + \bar{\varepsilon}^2\iota_{T-1}\Gamma_T'\iota_{T-1}\right] \\ =\ & -\frac{\sigma_1^2}{T}\iota_{T-1}\Gamma_T'\iota_{T-1}. \end{aligned}$$

Straightforward algebra provides the result in (16).

Bias Term in Equation (21): Let Γ_{1T} be the T_B dimensional Γ_T, Γ_{2T} be the $T - T_B - 1$ dimensional Γ_T. Let $\boldsymbol{\varepsilon}_{1,-1} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_{T_B-1})'$, $\boldsymbol{\varepsilon}_1 = (\varepsilon_2, \varepsilon_3, ..., \varepsilon_{T_B})'$, and $\bar{\varepsilon}_1 = \frac{1}{T_B}\sum_{t=1}^{T_B}\varepsilon_t$. Let $\boldsymbol{\varepsilon}_{2,-1} = (\varepsilon_{T_{B+1}}, \varepsilon_{T_{B+2}}, ..., \varepsilon_{T-1})'$, $\boldsymbol{\varepsilon}_2 = (\varepsilon_{T_{B+2}}, \varepsilon_{T_{B+2}}, ..., \varepsilon_T)'$, and $\bar{\varepsilon}_2 = \frac{1}{T-T_B}\sum_{t=T_B+1}^{T}\varepsilon_t$. In the presence of a structural break, we have

$$\begin{aligned} & E\left(\sum_{t=2}^{T}\tilde{w}_{t-1}\Delta\tilde{y}_t\right) \\ = \; & E\left[\begin{array}{c} (\boldsymbol{\varepsilon}_1 - \bar{\varepsilon}_1\iota_{T_B-1})'\,\Gamma_{1T}'\,(\boldsymbol{\varepsilon}_{1,-1} - \bar{\varepsilon}_1\iota_{T_B-1}) + \\ (\boldsymbol{\varepsilon}_2 - \bar{\varepsilon}_1\iota_{T-T_B-1})'\,\Gamma_{2T}'\,(\boldsymbol{\varepsilon}_{2,-1} - \bar{\varepsilon}_2\iota_{T-T_B-1}) \end{array}\right]. \end{aligned}$$

Equation (21) follows in view of the result for equation (16).

Table 1. Size and Power of IV tests
(Drift Model, $d_t = \gamma_0$)

T	ϕ	DF	GLS	IV-Uniform Weight			IV-Bartlett Weight		
				δ=0.7	δ=0.8	δ=0.9	δ=0.7	δ=0.8	δ=0.9
50	1	0.057	0.110	0.078	0.082	0.085	0.081	0.085	0.087
	0.9	0.147	0.383	0.245	0.257	0.268	0.274	0.300	0.317
		(0.133)	(0.197)	(0.164)	(0.171)	(0.172)	(0.184)	(0.195)	(0.195)
	0.95	0.089	0.216	0.143	0.151	0.162	0.152	0.163	0.173
		(0.080)	(0.099)	(0.093)	(0.096)	(0.096)	(0.094)	(0.097)	(0.101)
100	1	0.054	0.077	0.075	0.080	0.086	0.076	0.082	0.086
	0.9	0.359	0.622	0.468	0.479	0.498	0.548	0.623	0.665
		(0.341)	(0.495)	(0.362)	(0.361)	(0.362)	(0.438)	(0.486)	(0.512)
	0.95	0.145	0.284	0.234	0.253	0.269	0.250	0.284	0.313
		(0.135)	(0.198)	(0.166)	(0.171)	(0.168)	(0.177)	(0.192)	(0.198)
200	1	0.051	0.064	0.073	0.080	0.086	0.074	0.080	0.085
	0.9	0.879	0.871	0.803	0.794	0.777	0.929	0.966	0.973
		(0.877)	(0.840)	(0.728)	(0.698)	(0.665)	(0.881)	(0.930)	(0.936)
	0.95	0.339	0.532	0.448	0.480	0.489	0.499	0.599	0.661
		(0.336)	(0.467)	(0.357)	(0.360)	(0.347)	(0.405)	(0.471)	(0.507)
	0.99	0.071	0.112	0.116	0.130	0.142	0.117	0.130	0.142
		(0.069)	(0.087)	(0.081)	(0.083)	(0.084)	(0.081)	(0.085)	(0.089)
300	1	0.051	0.058	0.070	0.080	0.086	0.071	0.078	0.085
	0.9	0.996	0.940	0.943	0.913	0.883	0.996	0.999	0.998
		(0.996)	(0.930)	(0.909)	(0.862)	(0.820)	(0.992)	(0.996)	(0.995)
	0.95	0.637	0.722	0.633	0.661	0.653	0.729	0.840	0.891
		(0.633)	(0.687)	(0.541)	(0.542)	(0.527)	(0.649)	(0.744)	(0.786)
	0.99	0.082	0.135	0.143	0.164	0.182	0.141	0.168	0.184
		(0.081)	(0.118)	(0.103)	(0.107)	(0.110)	(0.099)	(0.113)	(0.113)
500	1	0.050	0.055	0.070	0.075	0.084	0.071	0.076	0.083
	0.9	1.000	0.985	0.994	0.982	0.963	1.000	1.000	1.000
		(1.000)	(0.983)	(0.989)	(0.967)	(0.934)	(1.000)	(1.000)	(1.000)
	0.95	0.970	0.875	0.885	0.866	0.833	0.968	0.992	0.994
		(0.969)	(0.864)	(0.837)	(0.804)	(0.746)	(0.947)	(0.981)	(0.982)
	0.99	0.127	0.201	0.197	0.227	0.255	0.194	0.250	0.290
		(0.126)	(0.186)	(0.146)	(0.156)	(0.165)	(0.147)	(0.174)	(0.185)
1000	1	0.050	0.055	0.068	0.076	0.085	0.069	0.075	0.084
	0.9	1.000	0.999	1.000	0.999	0.995	1.000	1.000	1.000
		(1.000)	(0.999)	(1.000)	(0.998)	(0.991)	(1.000)	(1.000)	(1.000)
	0.95	1.000	0.982	0.994	0.983	0.962	1.000	1.000	1.000
		(1.000)	(0.979)	(0.990)	(0.969)	(0.931)	(1.000)	(1.000)	(1.000)
	0.99	0.333	0.451	0.380	0.447	0.476	0.390	0.527	0.632
		(0.332)	(0.431)	(0.311)	(0.345)	(0.337)	(0.315)	(0.418)	(0.483)

Note: All of the tests are conducted at the 5% level. The figures in parentheses are size-adjusted power. For the IV tests, the asymptotic one-tailed critical value (-1.645) of the standard normal distribution was used in all cases.

Table 2. Size and Power of IV tests
(Linear Trend Model, $d_t = \gamma_0 + \gamma_1 t$)

T	ϕ	DF	LM	GLS	IV-Uniform Weight			IV-Bartlett Weight		
					δ=0.6	δ=0.7	δ=0.8	δ=0.6	δ=0.7	δ=0.8
50	1	0.061	0.057	0.054	0.057	0.036	0.012	0.068	0.057	0.038
	0.9	0.106	0.115	0.111	0.102	0.060	0.022	0.136	0.115	0.078
		(0.087)	(0.102)	(0.103)	(0.093)	(0.083)	(0.085)	(0.101)	(0.104)	(0.104)
	0.95	0.076	0.073	0.071	0.071	0.044	0.016	0.086	0.073	0.049
		(0.065)	(0.065)	(0.065)	(0.063)	(0.061)	(0.059)	(0.064)	(0.064)	(0.066)
	0.99	0.059	0.056	0.056	0.057	0.036	0.013	0.071	0.057	0.038
		(0.048)	(0.050)	(0.052)	(0.049)	(0.049)	(0.054)	(0.050)	(0.050)	(0.050)
100	1	0.056	0.054	0.052	0.055	0.036	0.013	0.062	0.053	0.033
	0.9	0.223	0.253	0.269	0.209	0.137	0.046	0.277	0.264	0.190
		(0.204)	(0.241)	(0.260)	(0.197)	(0.176)	(0.138)	(0.237)	(0.253)	(0.255)
	0.95	0.095	0.102	0.105	0.099	0.068	0.022	0.117	0.108	0.071
		(0.086)	(0.097)	(0.102)	(0.094)	(0.090)	(0.078)	(0.096)	(0.102)	(0.101)
	0.99	0.059	0.058	0.056	0.061	0.039	0.013	0.066	0.057	0.037
		(0.053)	(0.055)	(0.055)	(0.056)	(0.052)	(0.051)	(0.054)	(0.055)	(0.054)
200	1	0.054	0.050	0.052	0.054	0.043	0.017	0.060	0.053	0.034
	0.9	0.656	0.715	0.746	0.543	0.409	0.142	0.689	0.733	0.650
		(0.637)	(0.716)	(0.734)	(0.527)	(0.448)	(0.304)	(0.649)	(0.721)	(0.742)
	0.95	0.206	0.242	0.265	0.211	0.164	0.052	0.251	0.259	0.190
		(0.192)	(0.243)	(0.255)	(0.199)	(0.186)	(0.134)	(0.218)	(0.248)	(0.257)
	0.99	0.061	0.058	0.061	0.064	0.052	0.020	0.070	0.065	0.041
		(0.057)	(0.059)	(0.057)	(0.059)	(0.059)	(0.059)	(0.059)	(0.061)	(0.061)
300	1	0.053	0.051	0.054	0.053	0.044	0.021	0.059	0.054	0.038
	0.9	0.956	0.946	0.952	0.821	0.685	0.310	0.929	0.966	0.951
		(0.951)	(0.945)	(0.948)	(0.809)	(0.718)	(0.506)	(0.913)	(0.959)	(0.972)
	0.95	0.398	0.472	0.502	0.361	0.277	0.103	0.424	0.480	0.411
		(0.382)	(0.469)	(0.479)	(0.345)	(0.304)	(0.211)	(0.388)	(0.454)	(0.486)
	0.99	0.069	0.069	0.077	0.070	0.056	0.029	0.078	0.075	0.052
		(0.065)	(0.068)	(0.070)	(0.065)	(0.064)	(0.066)	(0.067)	(0.068)	(0.070)
500	1	0.053	0.049	0.050	0.056	0.047	0.025	0.058	0.055	0.038
	0.9	1.000	0.999	0.997	0.988	0.958	0.694	0.999	1.000	1.000
		(1.000)	(0.999)	(0.997)	(0.985)	(0.963)	(0.821)	(0.999)	(1.000)	(1.000)
	0.95	0.848	0.865	0.865	0.695	0.620	0.277	0.773	0.875	0.857
		(0.836)	(0.868)	(0.866)	(0.667)	(0.638)	(0.426)	(0.742)	(0.863)	(0.896)
	0.99	0.083	0.089	0.096	0.100	0.083	0.041	0.101	0.103	0.076
		(0.078)	(0.091)	(0.096)	(0.090)	(0.089)	(0.080)	(0.087)	(0.097)	(0.098)
1000	1	0.051	0.049	0.046	0.056	0.051	0.031	0.060	0.057	0.043
	0.9	1.000	1.000	1.000	1.000	1.000	0.992	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(0.998)	(1.000)	(1.000)	(1.000)
	0.95	1.000	0.999	0.995	0.988	0.971	0.767	0.998	1.000	1.000
		(1.000)	(0.999)	(0.995)	(0.986)	(0.970)	(0.854)	(0.996)	(1.000)	(1.000)
	0.99	0.190	0.232	0.238	0.201	0.189	0.103	0.193	0.233	0.217
		(0.187)	(0.237)	(0.254)	(0.185)	(0.187)	(0.156)	(0.167)	(0.210)	(0.242)

See Table 1 footnotes.

Table 3-1. Size and Power of IV tests
(Level and Trend Shift with T_B = 0.5T)
($d_t = \gamma_0 + \gamma_1 t + \gamma_2 D_t + \gamma_3 (D_t \times t)$)

T	ϕ	DF	LM	GLS	IV-Uniform Weighting			IV-Bartlett Weighting		
					δ=0.6	δ=0.7	δ=0.8	δ=0.6	δ=0.7	δ=0.8
100	1	-	-	-	0.054	0.037	0.016	0.064	0.055	0.038
	0.9	-	-	-	0.145	0.091	0.035	0.187	0.173	0.123
		(0.116)	(0.153)	(0.160)	(0.136)	(0.117)	(0.103)	(0.151)	(0.159)	(0.159)
	0.95	-	-	-	0.077	0.053	0.022	0.094	0.083	0.058
		(0.071)	(0.075)	(0.077)	(0.072)	(0.069)	(0.067)	(0.074)	(0.074)	(0.077)
	0.99	-	-	-	0.056	0.041	0.017	0.069	0.061	0.042
		(0.055)	(0.054)	(0.053)	(0.052)	(0.053)	(0.052)	(0.054)	(0.054)	(0.054)
200	1	-	-	-	0.051	0.037	0.017	0.059	0.051	0.033
	0.9	-	-	-	0.384	0.269	0.097	0.471	0.493	0.430
		(0.371)	(0.491)	(0.537)	(0.379)	(0.335)	(0.232)	(0.430)	(0.489)	(0.526)
	0.95	-	-	-	0.141	0.099	0.041	0.159	0.159	0.114
		(0.114)	(0.155)	(0.167)	(0.138)	(0.132)	(0.112)	(0.136)	(0.156)	(0.162)
	0.99	-	-	-	0.056	0.040	0.020	0.066	0.055	0.038
		(0.052)	(0.055)	(0.059)	(0.054)	(0.054)	(0.055)	(0.055)	(0.054)	(0.057)
300	1	-	-	-	0.052	0.039	0.019	0.056	0.050	0.033
	0.9	-	-	-	0.667	0.533	0.193	0.783	0.841	0.802
		(0.741)	(0.811)	(0.875)	(0.658)	(0.597)	(0.378)	(0.759)	(0.841)	(0.874)
	0.95	-	-	-	0.235	0.191	0.065	0.271	0.286	0.230
		(0.212)	(0.297)	(0.316)	(0.229)	(0.229)	(0.163)	(0.249)	(0.286)	(0.307)
	0.99	-	-	-	0.059	0.045	0.022	0.068	0.058	0.040
		(0.057)	(0.063)	(0.064)	(0.057)	(0.058)	(0.060)	(0.059)	(0.058)	(0.062)
500	1	-	-	-	0.051	0.043	0.022	0.056	0.050	0.035
	0.9	-	-	-	0.961	0.899	0.521	0.990	0.998	0.998
		(0.996)	(0.989)	(0.997)	(0.960)	(0.918)	(0.712)	(0.988)	(0.998)	(0.999)
	0.95	-	-	-	0.506	0.453	0.186	0.562	0.655	0.641
		(0.534)	(0.678)	(0.729)	(0.502)	(0.492)	(0.338)	(0.535)	(0.653)	(0.721)
	0.99	-	-	-	0.070	0.064	0.027	0.073	0.071	0.052
		(0.065)	(0.072)	(0.073)	(0.069)	(0.074)	(0.066)	(0.066)	(0.071)	(0.073)
1000	1	-	-	-	0.054	0.045	0.027	0.056	0.053	0.039
	0.9	-	-	-	1.000	1.000	0.969	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(0.989)	(1.000)	(1.000)	(1.000)
	0.95	-	-	-	0.951	0.931	0.618	0.970	0.996	0.999
		(0.996)	(0.986)	(0.995)	(0.946)	(0.940)	(0.759)	(0.964)	(0.995)	(0.999)
	0.99	-	-	-	0.128	0.121	0.065	0.132	0.147	0.127
		(0.115)	(0.156)	(0.162)	(0.120)	(0.132)	(0.117)	(0.121)	(0.142)	(0.153)

See Table 1 footnotes.

Table 3-2. Size and Power of IV tests
(Level and Trend Shift with T_B = 0.2T)
$$\left(d_t = \gamma_0 + \gamma_1 t + \gamma_2 D_t + \gamma_3 \left(D_t \times t \right) \right)$$

T	ϕ	DF	LM	GLS	IV-Uniform Weighting			IV-Bartlett Weighting		
					δ=0.6	δ=0.7	δ=0.8	δ=0.6	δ=0.7	δ=0.8
100	1	-	-	-	0.049	0.033	0.025	0.057	0.047	0.030
	0.9	-	-	-	0.148	0.086	0.064	0.208	0.179	0.127
		(0.154)	(0.189)	(0.192)	(0.151)	(0.130)	(0.112)	(0.188)	(0.190)	(0.191)
	0.95	-	-	-	0.075	0.053	0.039	0.093	0.079	0.052
		(0.077)	(0.083)	(0.086)	(0.076)	(0.079)	(0.070)	(0.082)	(0.085)	(0.085)
	0.99	-	-	-	0.049	0.032	0.025	0.061	0.050	0.033
		(0.054)	(0.050)	(0.052)	(0.050)	(0.050)	(0.051)	(0.054)	(0.054)	(0.054)
200	1	-	-	-	0.046	0.033	0.020	0.054	0.043	0.026
	0.9	-	-	-	0.372	0.242	0.114	0.535	0.542	0.438
		(0.472)	(0.570)	(0.611)	(0.391)	(0.321)	(0.235)	(0.518)	(0.575)	(0.594)
	0.95	-	-	-	0.142	0.100	0.049	0.178	0.165	0.108
		(0.142)	(0.184)	(0.192)	(0.150)	(0.141)	(0.110)	(0.166)	(0.184)	(0.188)
	0.99	-	-	-	0.052	0.038	0.021	0.059	0.049	0.030
		(0.056)	(0.054)	(0.055)	(0.055)	(0.058)	(0.052)	(0.055)	(0.055)	(0.056)
300	1	-	-	-	0.046	0.035	0.019	0.051	0.043	0.025
	0.9	-	-	-	0.672	0.447	0.182	0.840	0.880	0.814
		(0.838)	(0.883)	(0.933)	(0.688)	(0.540)	(0.347)	(0.836)	(0.899)	(0.913)
	0.95	-	-	-	0.249	0.167	0.070	0.311	0.317	0.226
		(0.277)	(0.361)	(0.373)	(0.266)	(0.227)	(0.161)	(0.305)	(0.350)	(0.363)
	0.99	-	-	-	0.055	0.041	0.025	0.064	0.052	0.033
		(0.060)	(0.061)	(0.062)	(0.059)	(0.061)	(0.060)	(0.061)	(0.061)	(0.064)
500	1	-	-	-	0.049	0.037	0.020	0.053	0.046	0.026
	0.9	-	-	-	0.953	0.823	0.420	0.996	0.999	0.998
		(0.999)	(0.994)	(1.000)	(0.955)	(0.870)	(0.634)	(0.995)	(0.999)	(1.000)
	0.95	-	-	-	0.541	0.401	0.144	0.638	0.712	0.643
		(0.668)	(0.758)	(0.809)	(0.548)	(0.474)	(0.286)	(0.620)	(0.730)	(0.790)
	0.99	-	-	-	0.070	0.053	0.028	0.081	0.071	0.041
		(0.069)	(0.081)	(0.078)	(0.072)	(0.071)	(0.072)	(0.075)	(0.078)	(0.082)
1000	1	-	-	-	0.049	0.039	0.022	0.056	0.049	0.030
	0.9	-	-	-	1.000	0.999	0.904	1.000	1.000	1.000
		(1.000)	(1.000)	(1.000)	(1.000)	(0.999)	(0.967)	(1.000)	(1.000)	(1.000)
	0.95	-	-	-	0.960	0.878	0.452	0.989	0.998	0.999
		(0.999)	(0.995)	(1.000)	(0.961)	(0.909)	(0.650)	(0.986)	(0.999)	(1.000)
	0.99	-	-	-	0.145	0.116	0.055	0.149	0.160	0.118
		(0.143)	(0.181)	(0.185)	(0.145)	(0.144)	(0.121)	(0.138)	(0.164)	(0.176)

See Table 1 footnotes.

Table 4. Size and Power of IV tests
(Serially Correlated Errors, Linear Trend Model)

DGP			$T = 100$			$T = 200$			$T = 500$		
AR coeff			# of lags used			# of lags used			# of lags used		
a_1	a_2	ϕ	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
AR(1) Error											
0.8		1.0	0.036	0.030	0.031	0.045	0.038	0.036	0.049	0.047	0.046
		0.95	0.068	0.058	0.047	0.178	0.159	0.139	0.704	0.681	0.647
		0.90	0.124	0.094	0.081	0.406	0.360	0.320	0.985	0.975	0.961
0.5		1.0	0.037	0.033	0.033	0.046	0.045	0.045	0.048	0.045	0.045
		0.95	0.077	0.060	0.057	0.217	0.193	0.176	0.817	0.782	0.759
		0.90	0.171	0.131	0.110	0.589	0.526	0.463	0.999	0.999	0.997
0.0		1.0	0.038	0.031	0.029	0.045	0.044	0.044	0.048	0.046	0.046
		0.95	0.077	0.053	0.053	0.210	0.189	0.183	0.838	0.806	0.777
		0.90	0.186	0.137	0.117	0.640	0.571	0.514	0.999	0.999	0.999
-0.5		1.0	0.038	0.027	0.027	0.049	0.044	0.043	0.049	0.049	0.045
		0.95	0.066	0.052	0.044	0.180	0.169	0.157	0.775	0.773	0.733
		0.90	0.152	0.128	0.099	0.569	0.549	0.479	0.995	0.998	0.995
-0.8		1.0	0.035	0.025	0.026	0.047	0.040	0.042	0.050	0.048	0.047
		0.95	0.046	0.041	0.030	0.125	0.134	0.114	0.607	0.680	0.586
		0.90	0.103	0.102	0.069	0.425	0.470	0.363	0.929	0.988	0.938
AR(2) Error											
-0.3	0.4	1.0	0.000	0.032	0.029	0.000	0.044	0.044	0.000	0.049	0.046
		0.95	0.001	0.063	0.055	0.002	0.185	0.171	0.035	0.784	0.747
		0.90	0.003	0.140	0.109	0.022	0.537	0.473	0.614	0.999	0.997
0.3	0.4	1.0	0.001	0.032	0.030	0.000	0.042	0.037	0.000	0.047	0.048
		0.95	0.001	0.058	0.051	0.003	0.168	0.150	0.027	0.692	0.671
		0.90	0.003	0.103	0.092	0.015	0.378	0.345	0.337	0.981	0.971
0.8	-0.15	1.0	0.100	0.033	0.035	0.139	0.046	0.043	0.169	0.047	0.047
		0.95	0.183	0.061	0.058	0.442	0.188	0.177	0.961	0.769	0.754
		0.90	0.337	0.129	0.114	0.804	0.500	0.453	1.000	0.998	0.996
-0.8	0.15	1.0	0.006	0.026	0.024	0.009	0.040	0.040	0.010	0.048	0.046
		0.95	0.010	0.038	0.030	0.036	0.119	0.111	0.334	0.615	0.577
		0.90	0.029	0.085	0.064	0.190	0.390	0.342	0.853	0.950	0.930

Note: For the DGP with AR(2) errors, the coefficients are given from $a_1 = c_1+c_2$, $a_2 = -c_1c_2$, where c_1 and c_2 are the roots of $(\lambda-c_1)(\lambda-c_2)=0$. We use the values of $(c_1, c_2) = (-0.5, -0.8), (-0.5, 0.8), (0.5, 0.3)$ and $(-0.5,-0.3)$, respectively. For all cases, the IV tests with the Bartlett windows and $\delta=0.7$ are used. The asymptotic one-tailed critical value (-1.645) of the standard normal distribution was used in all cases.

Table 5. Power Dependence on x_0, the Initial Values of the Stochastic Process

T	ϕ	Std. Dev.	x_0	Drift			Linear Trend			
				DF	GLS	IV-Bartlett δ=0.9	DF	LM	GLS	IV-Bartlett δ=0.7
100	0.9	0	0	0.319	0.736	0.659	0.196	0.274	0.315	0.294
		1	2.23	0.339	0.398	0.479	0.199	0.238	0.244	0.246
		2	4.59	0.412	0.025	0.102	0.223	0.154	0.118	0.136
	0.95	0	0	0.121	0.315	0.287	0.083	0.107	0.113	0.110
		1	3.20	0.127	0.140	0.162	0.082	0.092	0.098	0.099
		2	6.41	0.156	0.010	0.022	0.082	0.072	0.063	0.067
	0.99	0	0	0.057	0.078	0.075	0.054	0.055	0.054	0.055
		1	7.09	0.056	0.062	0.062	0.053	0.056	0.055	0.056
		2	14.18	0.058	0.032	0.033	0.052	0.055	0.054	0.054
200	0.9	0	0	0.862	0.997	0.987	0.619	0.774	0.861	0.775
		1	2.23	0.877	0.904	0.962	0.640	0.712	0.728	0.724
		2	4.59	0.917	0.082	0.705	0.685	0.533	0.340	0.549
	0.95	0	0	0.315	0.757	0.651	0.186	0.270	0.314	0.283
		1	3.20	0.335	0.329	0.473	0.195	0.237	0.234	0.241
		2	6.41	0.410	0.009	0.103	0.216	0.151	0.099	0.141
	0.99	0	0	0.066	0.119	0.114	0.056	0.061	0.060	0.063
		1	7.09	0.069	0.079	0.080	0.057	0.060	0.059	0.061
		2	14.18	0.075	0.020	0.026	0.055	0.053	0.055	0.055
300	0.9	0	0	0.995	1.000	1.000	0.947	0.977	0.996	0.975
		1	2.23	0.996	0.998	1.000	0.952	0.950	0.975	0.964
		2	4.59	0.998	0.358	0.989	0.964	0.845	0.706	0.909
	0.95	0	0	0.604	0.966	0.902	0.371	0.513	0.601	0.504
		1	3.20	0.629	0.611	0.800	0.384	0.457	0.450	0.451
		2	6.41	0.707	0.015	0.362	0.425	0.306	0.165	0.299
	0.99	0	0	0.080	0.171	0.155	0.065	0.070	0.073	0.070
		1	7.09	0.085	0.088	0.100	0.063	0.068	0.070	0.067
		2	14.18	0.093	0.012	0.024	0.061	0.055	0.053	0.056
500	0.95	0	0	0.965	1.000	0.999	0.824	0.915	0.970	0.895
		1	3.20	0.970	0.965	0.995	0.839	0.874	0.878	0.869
		2	6.41	0.982	0.094	0.921	0.868	0.729	0.452	0.752
	0.99	0	0	0.117	0.314	0.266	0.077	0.101	0.113	0.104
		1	7.09	0.127	0.128	0.160	0.078	0.095	0.096	0.095
		2	14.18	0.153	0.007	0.027	0.080	0.071	0.062	0.071
1000	0.99	0	0	0.314	0.744	0.609	0.184	0.275	0.320	0.236
		1	7.09	0.333	0.263	0.469	0.187	0.235	0.234	0.208
		2	14.18	0.399	0.004	0.121	0.207	0.148	0.089	0.144

Note: All the figures are size-adjusted power at the 5% significance level. Initial values of the stochastic process are 0, 1, and 2 standard deviations. For example, when ϕ = 0.99, x_0 at the two standard deviations is $x_0 = 2\sigma / \sqrt{1-\phi^2} = 14.18$.

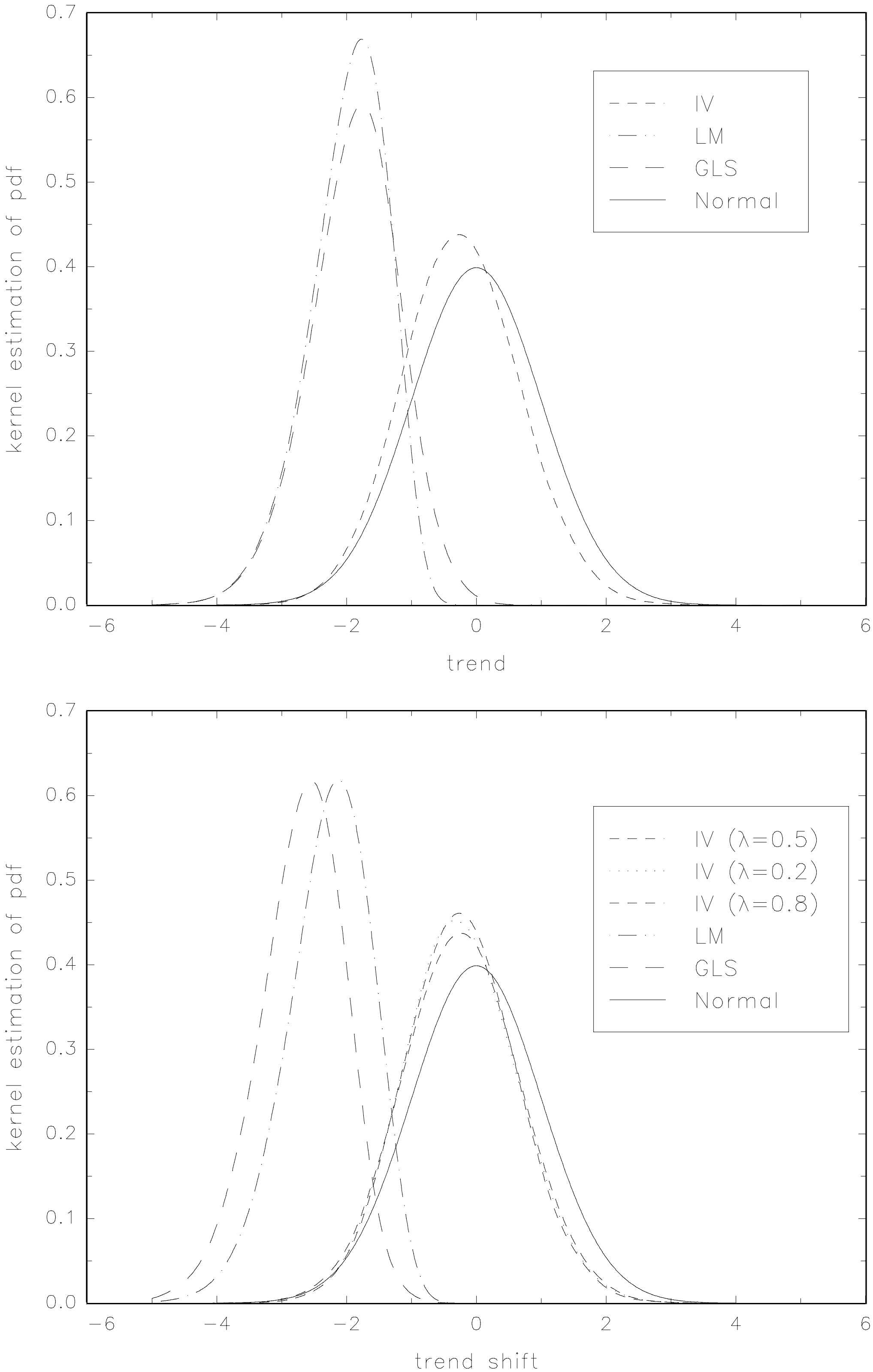

Figure 1. Estimated pdf (T = 100)

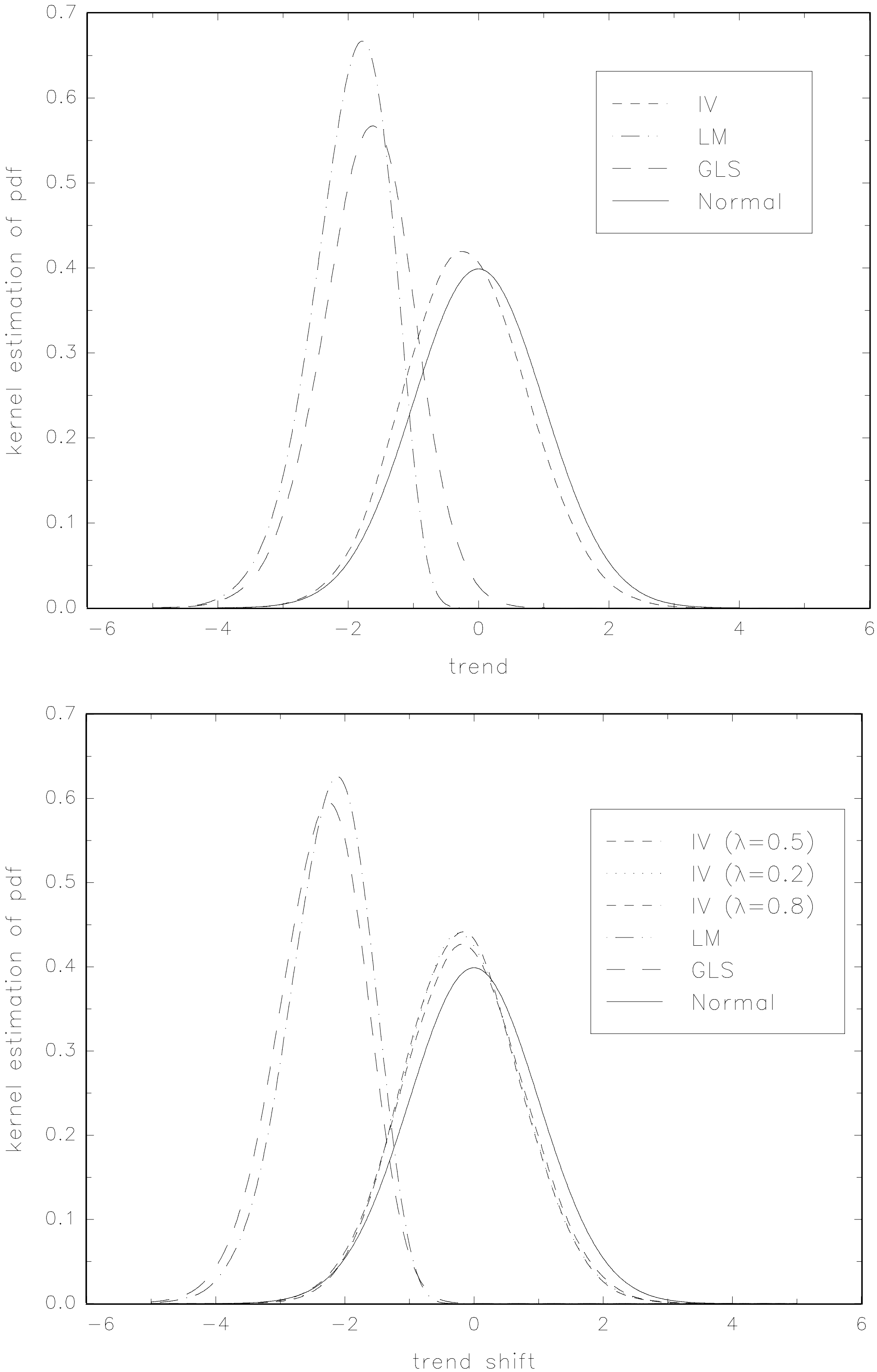

Figure 2. Estimated pdf (T = 500)