# THREE ESSAYS ON

# MORE POWERFUL COINTEGRATION TESTS

by

HYEJIN LEE

JUNSOO LEE, COMMITTEE CHAIR

WALTER ENDERS
JUN MA
J. BRIAN GRAY
SHAWN MOBBS

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Economics,
Finance, and Legal Studies
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2012

# ABSTRACT

The main focus of this dissertation is to find ways to improve the power in cointegration tests. This dissertation consists of three essays. In the first essay, a modified testing procedure for the Engle and Granger (1987; EG) cointegration test is suggested. Specifically, we suggest augmenting the usual EG testing regression with the first difference of the integrated regressors. The limiting distribution of this modified EG test under the null hypothesis will depend on the nuisance parameter, which reflects the signal-to-noise ratio. This essay shows that the nuisance parameter issue can be resolved when we follow the asymptotic distribution of the modified EG test, and use the relevant new sets of critical values corresponding to the estimated value of the nuisance parameter. It is found that the size and power properties of the modified EG test are fairly good. The modified EG test gains improved power rather than losing power as the signal-to-noise ratio increases.

In the second essay, we examine whether non-linear unit root tests is robust with non-normal errors, which provides a motivation for the third essay. Especially, the second essay demonstrates how popular nonlinear unit root tests perform in the presence of non-normal errors. Non-normal errors normally do not pose a problem in usual linear unit root tests since the least squares estimator will still be the most efficient under certain ideal conditions regardless of normal or non-normal errors. The asymptotic properties of the popular linear Dickey-Fuller tests, for example, will be unaffected by non-normal errors. As such, the literature has not paid much attention to this issue. Nevertheless, whether similar results will carry over to nonlinear unit root

tests with non-normal errors is a question that merits examination. To our surprise, the extant literature on nonlinear unit root tests has not examined this important question. We find that, in general, nonlinear unit root tests will suffer a loss of power in the presence of non-normal errors. In this regard, this essay brings out the neglected point that the obvious analogies of linear processes do not necessarily hold for nonlinear models.

The third essay suggests new cointegration tests that are more powerful in the presence of non-normal errors. We use a two-step procedure based on the “residual augmented least squares” (RALS) method to make use of nonlinear moment conditions driven by non-normal errors. By utilizing this neglected information, we can make the existing tests more powerful. The suggested testing procedure is easy to implement. The underlying idea is similar to adding stationary covariates to improve the power of the test, but the suggested procedure does not require any new covariates outside the system. Instead, we can exploit the information on the non-normal error distribution that is already available but ignored in the usual cointegration tests. Our simulation results show significant power gains over existing cointegration tests.

## DEDICATION

This dissertation is dedicated to those who wholeheartedly supported and guided me, especially to my parents and my husband.

## ACKNOWLEDGMENTS

I express my deep gratitude to my Heavenly Father. It is his indescribable grace to enable me to complete this work.

I would like to express my sincere appreciation to my advisor, Dr. Junsoo Lee. His endless enthusiasm for research and commitment to high standards inspired and motivated me throughout all my PhD years. I am also thankful to all of my committee members, Dr. Walter Enders, Dr. Jun Ma, Dr. Brian Gray, and Dr. Shawn Mobbs, for their invaluable comments, inspiring questions, and kind support for my dissertation.

I am deeply indebted to Dr. Jong-Won Lee. Without his wise advice and constructive support, I would not have had the opportunity to study here. I am also grateful to Dr. Juheon Park for enlightening me by giving me the first glance of Econometrics in my life.

Lastly, and importantly, I wish to give my thanks to my husband Dong-Yop who has supported me with sincere prayer throughout all my hard times, and to my lovely daughter Hannah.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

# INTRODUCTION

The main focus of this dissertation is to find ways to improve the power in cointegration tests. The concept of cointegration is well defined in the literature. Cointegration concerns the long-run equilibrium relationship between two or more nonstationary variables. If two nonstationary variables are cointegrated, they tend to move together. How can we capture this "co-movement" among the series and test it? Engle and Granger (1987; EG) is one of the pioneering papers about this issue. Their intuition is simple. If cointegration holds, the deviation from a long-run equilibrium relationship should be temporary. As such, EG suggest testing whether the residual from the long-run equilibrium regression has a mean reversion tendency. Many empirical studies have examined cointegration. For example, the law of one price and the associated purchasing power parity hypothesis, the term structure in interest theory, the Fisher effect, and price transmission of agricultural products or other commodities are leading examples topics in cointegration research. However, it is widely known that the EG test and other existing cointegration tests often fail to detect cointegration. One potential issue is poor power of the existing cointegration tests.

In the first essay, we propose an improved testing procedure for the EG. The EG cointegration test has been used the most widely in the literature since it has a good size property. In particular, the EG test is free of any nuisance parameters under the null. However, the power property of the EG test has been an issue. To our surprise, we can show that the power of the modified EG test improves significantly when we adopt a simple modification procedure. Specifically, we augment the usual EG testing regression with the first difference of the integrated regressors in a single equation framework where weak exogeneity of the regressors is assumed. Although adding the additional terms will induce a nuisance parameter, we can use the relevant new sets of critical values following the new limiting distribution of the modified EG test. The size property is fairly good. More importantly, the modified EG test gains significantly improved power as the signal-to-noise ratio increases, while the usual EG test tends to lose power under the same situation. As such, the signal-to-noise ratio becomes a useful factor to improve the power for the modified EG cointegration test without affecting the performance under the null. This outcome is opposite to the usual EG test, where increasing the signal-to-noise ratio was regarded as a source of power loss.

In the second essay, we examine an issue that provides a motivation for the third essay. The third essay tries to seek more powerful cointegration tests by utilizing information on non-normal errors. One related issue is whether non-linear unit root tests will remain robust with non-normal errors. As such, the second essay examines the performance of non-linear unit root tests in the presence of non-normal errors. In testing for unit roots or cointegration, it has been common to derive the critical values under the assumption of normal errors. Perhaps, this common practice relies on the presumption that the asymptotic distributions in these tests will be unaffected if the error terms are non-normal. It will be interesting to examine whether the usual tests will be affected much if the error term follows a non-normal distribution. Actually, this presumption holds trivially for the usual linear unit root and cointegration tests. Indeed, in light of the Gauss-Markov theorem and the functional central limit theorem, non-normal errors will not pose a problem, at least asymptotically, for linear based tests. The least squares estimators from linear models are the most efficient and the resulting test statistics have a stable distribution regardless of whether the distribution of the error term is normal. Nevertheless, whether similar results will carry over to nonlinear unit root tests with non-normal errors is a question that merits examination. Thus, our focus in the present essay is to examine how nonlinear unit root tests perform in the presence of non-normal errors. We note that nonlinear cointegration tests will exhibit the same property as in nonlinear unit root tests. We find that non-linear unit root tests will lose power in the presence of non-normal errors. This result makes a sharp contrast with the linear based tests. Thus, we conclude that it is necessary to specify correctly the distribution of the error term for nonlinear unit root and cointegration tests. Otherwise, they will suffer from loss of power. Nonlinearity and non-normal errors are often hard to distinguish, and the findings in this essay indicate that using the information non-normal errors can be more promising, as we demonstrate in the third essay.

In the third essay, we develop extended cointegration tests that can be more powerful when utilizing the information on non-normal errors. The relatively low power of the usual cointegration tests is well recognized in the literature. Indeed, the task of seeking more powerful tests is not a trivial concern. In this essay, we show how existing cointegration tests can become more powerful by utilizing information that the errors are non-normal. It seems clear that existing cointegration tests ignore information about non-normal errors. However, if we can possibly utilize this information in the cointegration test, the resulting estimator can be more efficient and the test will become more powerful. The question is how to utilize information that the errors are non-normal. If we know the true distribution of the error term, it might be possible to employ the maximum likelihood estimators (MLE) using the known density function. However, we seldom know the true density function of a non-normal error, and it is extremely difficult to identify the underlying distribution. Moreover, if we use incorrect information on the error distribution or the functional form of the relationship, the consequence could be even worse.

The new cointegration tests that we propose in this essay will utilize information on the non-normal distribution of the error term without pre-specifying a particular density function or functional form. In doing this, we extend the work of Im, Lee and Tieslau (2011), who develop unit root tests with non-normal errors. The underlying idea is to adopt a two-step procedure following the "residual augmented least squares" (RALS) method of Im and Schmidt (2008), which can make use of nonlinear moment conditions driven by non-normal errors. If the errors are non-normal, the higher moments of the error term (or residuals) will contain information on the nature of the non-normal errors. If we can utilize this information in the higher moments of the residuals, then we can potentially obtain more powerful cointegration tests. The suggested RALS testing procedure is easy to implement because it does not require non-linear estimation techniques, even though we utilize nonlinear moment conditions associated with the non-normal errors. We base our suggested tests on the usual least square estimation and adopt a linearized RALS procedure to utilize the nonlinear moment conditions from higher moments of the residuals. The procedure is justified by the GMM approach and the RALS estimator is as efficient as the GMM estimator. We compare the consistency and efficiency of the proposed test with four different types of single equation based cointegration tests. The Monte Carlo simulation results demonstrate that the proposed tests obtain significant power gains over the existing cointegration tests.

CHAPTER 1

## MORE POWERFUL ENGLE-GRANGER COINTEGRATION TESTS

This essay proposes an improved testing procedure for the Engle and Granger (1987; EG) cointegration test. The residual based EG test has been used widely in the literature. It is quite intuitive, easy to implement, and has good size property; see, for example, Haug (1993, 1996) and Pesavento (2004). In particular, the EG test shows a good size property, and the test is free of any nuisance parameters under the null. Meanwhile, the power property of the EG test can be an issue. For example, Pesavento (2004) finds that the EG cointegration test is less powerful relative to other popular cointegration tests, and Kremers, Ericsson, and Dolado (1992) note that the EG test can yield loss of power when the signal-to-noise ratio increases.

This essay proposes a simple modification procedure to improve the power of the EG test in such cases. Specifically, we suggest augmenting the usual EG testing regression with the first difference of the integrated regressors. To the best of our knowledge, this simple modification has not been considered in the literature. This may be due to the fact that adding these terms will induce a nuisance parameter problem for the resulting cointegration test. Although the usual EG test does not depend on the nuisance parameter reflecting the signal-to-noise ratio under the null, the limiting distribution of the modified EG test will depend on the nuisance parameter induced by the added terms under the null hypothesis. If this happens, the critical values of the EG test cannot be used. However, the nuisance parameter issue can be resolved when we follow the asymptotic distribution of the modified EG cointegration test and use the relevant new sets of critical values corresponding to the estimated value of the nuisance parameter. Indeed, this essay shows that the size and power properties of the modified EG test are fairly good when we use the new sets of critical values. The nuisance parameter dependency of the modified EG test is clearly resolved; we do not observe any significant size distortions under the null in the baseline cases. This outcome is due to the property of the usual EG cointegration test which is free of the nuisance parameter under the null even when the signal-to-noise ratio varies. This feature can be useful for our modified EG test when we explicitly include the term capturing the signal-to-noise ratio. Although the resulting test depends on the nuisance parameter, this dependency does not pose a problem when it is based on the EG type regression.[1] More importantly, the modified EG test tends to gain improved power as the signal-to-noise ratio increases, while the usual EG test tends to lose power. As such, the signal-to-noise ratio becomes a useful factor to improve the power for the modified EG cointegration test without affecting the performance under the null, even though it was regarded as the source of power loss in the usual EG test.

One interesting point is that the asymptotic distribution of the modified EG test is similar to that of the unit root test with stationary covariates as initially advocated by Hansen (1995). The proposed stationary covariates will induce the nuisance parameter, but this can be estimated by a nonparametric approach. Thus, the underlying concept of the present essay is similar to that of Hansen (1995). The augmented terms work like stationary covariates and can play an important role to improve the power of the modified EG test. However, in this essay, we consider a similar approach in a cointegration system and do not need to seek outside stationary covariates, which are required in the other stationary covariate literature.

[1] This feature makes a sharp contrast from the cointegration test based on the error correction model (ECM); see, for example, Zivot (2000). Unlike the EG type test, the usual ECM based test depends on the nuisance parameter under the null when the signal-to-noise ratio changes. In this case, the modified ECM test which includes a similar term to capture the signal-to-noise ratio does not have a good size property under the null; see Lee (2011) for details.

The remainder of the essay will proceed as follows. In section 1.1, we discuss the testing model. Section 1.2 provides the limiting distribution for the proposed test. We examine the power gain issue for the proposed test in section 1.3 using Monte Carlo simulations. Finally, concluding remarks are found in section 1.4.

## 1.1. Model and Test

The cointegration test of Engle and Granger (1987; EG) is based on a single equation model. If we have a system of equations and the system is correctly specified, one may use the test using the full information maximum likelihood (FIML) estimation. However, as will be discussed more in detail in Essay 3, in certain cases the single equation model is preferred because it provides a parsimonious representation of the underlying model. The usual EG test is based on the following regression:

$$\Delta\hat{\mu}_t = \delta\hat{\mu}_{t-1} + v_t \tag{1.1.1}$$

where $\hat{\mu}_t$ is the residual from the long-run equilibrium relationship

$$x_{1t} = d_t + \beta' x_{2t} + \mu_t, \tag{1.1.2}$$

and $x_{2t}$ is a $(n-1)\times 1$ vector in the system of equations with $x_t = (x_{1t}, x'_{2t})'$ and $d_t$ denotes the deterministic term. We follow Harbo, Johansen, Nielsen, and Behbek (1998) and consider the conditional error correction model (ECM) for $\Delta x_{1t}$ under the assumption that $x_{2t}$ is weakly exogenous,

$$\Delta x_{1t} = c + \delta z_{t-1} + \phi' \Delta x_{2t} + e_t \tag{1.1.3}$$

where $z_{t-1} = x_{1,t-1} - \beta' x_{2,t-1}$ is the error correction term. The lagged terms of $\Delta x_{1t}$ and $\Delta x_{2t}$ are omitted to simplify the notation. It is important to note that the contemporaneous terms of

$\Delta x_{2t}$ appear in the above equation. We may subtract $\beta' \Delta x_{2t}$ from the above ECM in (1.1.3) and re-arrange it as follows:

$$\Delta(x_{1t} - \beta' x_{2t}) = c + \delta(x_{1,t-1} - \beta' x_{2,t-1}) + (\phi - \beta)' \Delta x_{2t} + e_t. \tag{1.1.4}$$

By relating the above equation to (1.1.1) along with $\Delta\mu_t = \Delta(x_{1t} - \beta' x_{2t})$ and $\mu_t = x_{1t} - \beta' x_{2t}$, we can observe that the signal-to-noise ratio term $(\phi - \beta)' \Delta x_{2t}$ is assumed be absent in (1.1.1). Here, $\phi$ is the short-run parameter in the relationship between $\Delta x_{1t}$ and $\Delta x_{2t}$, while $\beta$ is the long-run parameter in the relationship between $x_{1t}$ and $x_{2t}$. When these coefficients are the same ($\phi = \beta$), the signal-to-noise ratio term disappears and the usual EG equation (1.1.1) holds. In general, these coefficients are different, but the usual EG test is still valid since the null distribution does not depend on the nuisance parameter even when the true data generating process (DGP) involves the signal-to-noise ratio as in (1.1.3). However, there is room for improvement for the EG test since the test tends to lose power under the alternative hypothesis when the signal-to-noise ratio increases. This occurs when the short-run and the long-run coefficients are different ($\phi \neq \beta$), and the signal-to-noise ratio increases as the discrepancy between these coefficients gets large. We will examine this in more detail in the next section; see also see Kremers et al. (1992) and Ericsson and MacKinnon (2002).

In order to resolve this issue, we suggest a simple modification procedure. We suggest augmenting the usual testing regression with the first difference of the integrated regressors ( $\Delta x_{2t}$) such that the modified testing regression becomes

$$\Delta\hat{\mu}_t = c + \delta\hat{\mu}_{t-1} + \phi' \Delta x_{2t} + \epsilon_t \tag{1.1.5}$$

where a constant term is included. To match (1.1.1) with (1.1.5), we let

$$v_t = \phi' \Delta x_{2t} + \epsilon_t. \tag{1.1.6}$$

We wish to note that the variance of the new error term $\epsilon_t$ in (1.1.5) becomes smaller than the variance of the original error term of the EG regression $v_t$ in (1.1.1) because $var(\epsilon_t) = var(v_t - \phi\Delta x_{2t}) = var(v_t) - \frac{cov(v_t, \Delta x_2)^2}{var(\Delta x_{2t})}$. Thus, the regression parameter $\delta$ can be more precisely estimated from (1.1.5) than from (1.1.1). That is, confidence intervals will be smaller and the resulting test becomes more powerful.

## 1.2. Asymptotic Properties

Our main concern is to test if there is a cointegration relationship among the set of I(1) variables in $x_t = (x_{1t}, x'_{2t})'$. We consider the following null and alternative hypotheses

$$H_0 : \delta = 0 \text{ against } H_1 : \delta < 0.$$

As in Engle and Granger (1987), a t-statistic, which we denote as $t_{EG2}$, is obtained from (1.1.5) in order to test the above hypothesis

$$t_{EG2} = \frac{\hat{\delta}}{se(\hat{\delta})} \tag{1.2.1}$$

where $\hat{\delta}$ is the OLS estimator and $se(\hat{\delta})$ is the estimated standard error. We adopt the following assumption:

Assumption 1. *Suppose the following two conditions hold:*

*(1) $\{u_t\}$ and $\{\epsilon_t\}$ are i.i.d. processes with $E(u_t) = E(\epsilon_t) = 0$, $0 < var(u_t) < \infty$, and $0 < var(\epsilon_t) < \infty$.*

*(2) $x_{1t} = x_{2t}\beta_0 + u_t$, where $x_{2t} = x_{2t-1} + \epsilon_t$, $t = 1, 2, \cdots, T$, and $x_{20} = 0$ with $\beta_0 \in \Re$.*

Thorem 1. *Let $\hat{\mu}_t$ be the estimated residual generated by* (1.1.2). *Suppose $\hat{\hat{\delta}}$ is an estimator from equation* (1.1.5) *based on $\hat{\mu}_t$. Then, $\hat{\delta} \xrightarrow{p} \delta$.*

PROOF:

From the discussion above, we can relate the EG and modified EG test as

$$\Delta\hat{\mu}_t = \delta\hat{\mu}_{t-1} + \phi'\Delta x_{2t} + \epsilon_t$$

$$v_t = \phi'\Delta x_{2t} + \epsilon_t.$$

Suppose that Assumption 1 holds. The OLS estimator for $\delta$ in (1.1.5) is characterized by:

$$T(\hat{\delta} - \delta) = \frac{\frac{1}{T}\sum_{t=2}^{T}\hat{\mu}_{t-1}\epsilon_t}{\frac{1}{T^2}\sum_{t=2}^{T}\hat{\mu}_{t-1}^2} + o_p(1) \qquad (1.2.2)$$

Using the asymptotic result for the usual residual based test, we have

$$\frac{1}{T^2}\sum_{t=2}^{T}\hat{\mu}_{t-1}^2 \Rightarrow \sigma_v^2\int_0^1 (w_1^c)^2 dw_1. \qquad (1.2.3)$$

where $w_1^c$ is a standard Brownian motion of the *N(0,1)* random variable. Additionally, from the lemma in Hansen (1995),

$$\frac{1}{T}\sum_{t=2}^{T}\hat{\mu}_{t-1}\epsilon_t \to a(1)^{-1}\sigma_v\sigma_\epsilon\left(\rho\int_0^1 w_1^c dw_1 + (1-\rho^2)^{(1/2)}\int_0^1 w_1^c dw_2\right). \qquad (1.2.4)$$

where $\rho = \dfrac{\sigma_{v\epsilon}}{\sigma_v\sigma_\epsilon}$.

Plugging in (1.2.3) and (1.2.4) to (1.2.2) gives,

$$T(\hat{\delta} - \delta) = a(1)R\left(\rho\frac{\int_0^1 w_1^c dw_1}{\int_0^1 (w_1^c)^2} + (1-\rho^2)^{(1/2)}\frac{\int_0^1 w_1^c dw_2}{\int_0^1 (w_1^c)^2}\right) \qquad (1.2.5)$$

where $R = \frac{\sigma_\epsilon}{\sigma_\nu}$. From (1.2.5), we have $T(\hat{\delta} - \delta) = o_p(1)$, such that $\hat{\hat{\delta}} \xrightarrow{p} \delta$.

□

The above theorem implies that the estimator of $\delta$ is still consistent even when we use the estimated cointegrating vector $\hat{\beta}$ instead of a pre-specified vector $\beta$. The limiting distribution of the proposed modified EG test under the null hypothesis of no cointegration is represented as follows:

Theorem 2. *Let $t_{EG2}$ be the t-statistic for $\delta = 0$ in the testing regression* (1.1.5). *Then, under the null of no-cointegration $H_0 : \delta = 0$, we have*

$$t_{EG2} = \rho t_{EG} + \sqrt{(1-\rho^2)}N(0,1) \tag{1.2.6}$$

*where $t_{EG}$ is the usual EG t-statistic on $\delta = 0$ in the testing regression* (1.1.1)*, and $\rho^2$ is the long-run squared correlation between $v_t$ and $\epsilon_t$, as defined*

$$\rho^2 = \frac{\sigma_{v\epsilon}^2}{\sigma_v^2 \sigma_\epsilon^2} \tag{1.2.7}$$

PROOF:

From the result in theorem 1, the *t*-statistic under the null of $\delta = 0$ can be obtained as:

$$t(\hat{\delta}) = \hat{\delta} \cdot \sigma_\epsilon^{-1} \cdot T \left( \frac{1}{T^2} \sum_{t=2}^T \hat{\mu}_{t-1}^2 - \frac{1}{T^2} \sum_{t=2}^T \hat{\mu}_{t-1} \Delta x_{2t}' (\sum_{t=2}^T \Delta x_{2t} \Delta x_{2t}')^{-1} \sum_{t=2}^T \Delta x_{2t} \hat{\mu}_{t-1} \right)^{\frac{1}{2}}$$

$$= \sigma_\epsilon^{-1} \left( \frac{1}{T^2} \sum_{t=2}^T \hat{\mu}_{t-1}^2 \right)^{1/2} \cdot T\hat{\delta} + o_p(1).$$

Using $T\delta = -ca(1)$, test statistics under the null of $\delta = 0$ can be obtained as

$$t(\hat{\delta}) = \sigma_\epsilon^{-1} \left( a(1)^{-2} \sigma_\nu^2 \int_0^1 (w_1^c)^2 \right)^{\frac{1}{2}} \times \left[ -ca(1) + a(1)R \left( \rho \frac{\int_0^1 w_1^c dw_1}{\int_0^1 (w_1^c)^2} + (1-\rho^2)^{\frac{1}{2}} \frac{\int_0^1 w_1^c dw_2}{\int_0^1 (w_1^c)^2} \right) \right]$$

$$= -\frac{c}{R} \left( \int_0^1 (w_1^c)^2 \right)^{(1/2)} + \rho \frac{\int_0^1 w_1^c dw_1}{\left( \int_0^1 (w_1^c)^2 \right)^{(1/2)}} + (1-\rho^2) \frac{\int_0^1 w_1^c dw_2}{\left( \int_0^1 (w_1^c)^2 \right)^{(1/2)}} \tag{1.2.8}$$

Here, the null of no cointegration holds when $c = 0$. Therefore,

$$t(\hat{\delta}) = \rho t_{EG} + \sqrt{(1-\rho^2)}N(0,1).$$

Since $w_2(r)$ and $w_1^c(r)$ are independent Brownian motions, the ratio in (1.2.8) is normally distributed: see Phillips and Park (1988). □

Theorem 2 shows that the asymptotic distribution of $t_{EG2}$ is a linear combination of the distribution of the usual EG test $t_{EG}$ and a standard normal random variable, while the weight is the nuisance parameter $\rho^2$. As noted above, adding $\Delta x_{2t}$ as in (1.1.5) will induce a nuisance parameter issue; the limiting distribution of $t_{EG2}$ depends on the nuisance parameter reflecting the signal-to-noise ratio under the null hypothesis. Specifically, the limiting distribution of the modified EG test depends on the relative contribution of the variance of $v_t(= \phi\Delta x_{2t} + \epsilon_t)$ to the variance of $\epsilon_t$. Note that the above asymptotic distribution of the modified EG test is similar to that of the unit root test with stationary covariates as advocated by Hansen (1995). The underlying concept is the same. If stationary covariates are available, including them in the unit root testing equation is an effective way to improve the power of the unit root test without affecting its consistency.

When $\rho^2$ = 1, the distribution collapses to the usual EG distribution. This happens when the two error terms $v_t$ and $\epsilon_t$ are perfectly correlated such that $\Delta x_{2t}$ is not used to explain $v_t$ at all. Then, two error terms, $v_t$ and $\epsilon_t$, have the same long run variance. We also note that $cov(v_t\Delta x_{2t}) = 0$ is the condition for $\rho^2$ = 1. This implies there is no significant difference between the two testing procedures. If this is the case, the performance of the proposed test will be similar to that of the EG test. Another extreme case occurs when $\rho^2$ = 0. This happens when the covariance between $v_t$ and $\epsilon_t$ becomes zero. However, this situation may not occur. Since $cov(v_t\epsilon_t) = \phi cov(\Delta x_{2t}\epsilon_t) + var(\epsilon_t)$, this happens only when $\phi cov(\Delta x_{2t}\epsilon_t) = -var(\epsilon_t)$, which is rare. Because $\Delta x_{2t}$ is a regressor in (1.1.5) and $\epsilon_t$ is the error term, their covariance from the regression should be close to 0 if the model is correctly specified. However, a random variable $\epsilon_t$ does not have zero variance. For these reasons, we exclude the case of $\rho^2 = 0$.

The long-run squared correlation parameter $\rho^2$ can be obtained from a long run covariance matrix between $v_t$ and $\epsilon_t$

$$\Omega = \begin{pmatrix} \sigma_\epsilon^2 & \sigma_{\epsilon v} \\ \sigma_{v\epsilon} & \sigma_v^2 \end{pmatrix} \tag{1.2.9}$$

This matrix can be estimated by using a nonparametric estimation technique.

$$\hat{\Omega} = \begin{pmatrix} \hat{\sigma}_\epsilon^2 & \hat{\sigma}_{\epsilon v} \\ \hat{\sigma}_{v\epsilon} & \hat{\sigma}_v^2 \end{pmatrix} = \sum_{k=-\xi}^{\xi} W\left(\frac{k}{\xi}\right) \frac{1}{T} \sum_{t=1}^{T} \hat{V}_t \hat{V}_{t-k}' \tag{1.2.10}$$

where $W(\cdot)$ is a Bartlett or Parzen kernel weight function, $\xi$ is a bandwidth parameter and $\hat{V}_t = (\hat{\epsilon}_t, \hat{v}_t)'$ is constructed from the residuals from the regression of (1.1.5) and (1.1.1), respectively.

The distribution of $t_{EG2}$ will be affected by the different deterministic terms $(d_t)$ in (1.1.2). In this essay, we have considered two cases where $d_t = [1]$ for the model with "constant only" and $d_t = [1, t]'$ for the model with "constant and trend" in (1.1.2), while we consider a constant term only in (1.1.5), $c = [1]$, to take into account the possible non-zero constant of the newly added first differenced term. For a practical matter, we have obtained the critical values of the modified EG test in the following procedures. First, we run the long run equilibrium in (1.1.2) and obtain the residual $\{\hat{\mu}_t\}$, then we estimate the usual EG regression for the empirical distribution of $t_{EG}$. Then, for a given value of $\rho^2$ = 0.0, 0.1, 0.2, $\cdots$, 0.9 and 1.0, we obtain the empirical distribution of $t_{EG2}$ by using (1.2.6) with the weighted average of $t_{EG}$ and the simulated random variable following the standard normal distribution. Table 1.1 gives the empirical critical values of $t_{EG2}$ for the model with a constant $(d_t = [1])$, for different values of $\rho^2$, when $T$ = 100 and $T$ = 1000. Here, $k$ denotes for the number of cointegrating vectors. Table

1.2 gives the critical values for the model with a constant and a trend $(d_t = [1, t]')$, which we obtained through 50,000 replications.

The nuisance parameter issue is resolved by utilizing the estimated value of $\rho^2$. Thus, we can use the relevant critical values corresponding to the estimated value of the nuisance parameter, following the asymptotic distribution of the above modified EG cointegration test ( $t_{EG2}$).

## 1.3. Monte Carlo Experiment

The finite sample performance of the modified EG test ($t_{EG2}$) is examined and compared to the performances of the standard EG test ($t_{EG}$) via Monte Carlo simulations. To examine the size and power properties of the test statistics, we use the following data generating process (DGP)

$$\Delta x_{1t} = \delta_1(x_{1,t-1} - \beta x_{2,t-1}) + \phi_1 \Delta x_{2t} + \epsilon_{1t} \tag{1.3.1}$$

$$\Delta x_{2t} = \delta_2(x_{1,t-1} - \beta x_{2,t-1}) + \phi_2 \Delta x_{1t} + \psi \Delta x_{2,t-1} + \epsilon_{2t} \tag{1.3.2}$$

$$\Omega = E(\epsilon_t' \epsilon_t) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

where $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t})$, $\sigma_1^2 = \sigma_{\Delta x_{1t}}^2$, and $\sigma_2^2 = \sigma_{\Delta x_{2t}}^2$. We let $\beta = 1$ and $\sigma_{12} = \theta$. The first equation (1.3.1) is the conditional expectation of $\Delta x_{1t}$ on $\Delta x_{2t}$ and the information set $I_{t-1}$, which is obtained from the partitioned vector error correction model with respect to $x_{1t}$ and $x_{2t}$. The second equation (1.3.2) is the conditional expectation of $\Delta x_{2t}$ on $\Delta x_{1t}$, and $I_{t-1}$. Since our modified EG test works only for the single-equation model where weak exogeneity of $x_{2t}$ is assumed, we set $\delta_2 = 0$ and $\phi_2 = 0$.

Several Previous studies adopted similar DGP; see Banerjee et al. (1986), Boswijk and Frances (1992), Kremers et al. (1992), Boswijk (1994), Banerjee et al. (1998), Enders et al. (2010), Li and Lee (2010). However, we generalized DGP in the above studies to examine the performance of the proposed test from various perspectives.

(i) We consider the case where the error processes $\epsilon_{1t}$ and $\epsilon_{2t}$ are possibly correlated with the correlation coefficient, $\theta = \{0.0, 0.3, 0.7\}$.

(ii) We also consider the case where $\Delta x_{2t}$ is persistent and examine how the persistence of $\Delta x_{2t}$ affects the performance of the proposed test. As such, we allow $\Delta x_{2t}$ to have an AR(1) process with the persistence parameter $\psi = \{0.0, 0.6, 0.9\}$.

(iii) We let $\sigma_1^2 = 1$, but we let the ratio of the two variances of $\Delta x_{1t}$ and $\Delta x_{2t}$ vary along with $\sigma_2^2 = \{1, 6, 16\}$. As we will discuss in greater detail later, one source of power improvement for the modified EG test is the ratio of the two variances, which is a major factor of the signal-to-noise ratio.

(iv) We examine the effects of another source of the signal-to-noise ratio. As noted previously, we relax the assumption implied in the usual EG test that the long-run ($\beta$) coefficient in the relationship between $x_{1t}$ and $x_{2t}$ is the same as the short-run ($\phi_1$) coefficient in the relationship between $\Delta x_{1t}$ and $\Delta x_{2t}$. As we let $\beta = 1$, this implied assumption does not hold when $\phi_1 \neq 1$. Thus, we consider two cases: (i) the long-run parameter is the same as the short-run parameter such that $\phi_1 = 1.0 = \beta$, and (ii) the long-run parameter is different from the short-run parameter such that $\phi_1 = 0.5 \neq \beta$. The cases for $\phi_1 = 0$ or 2.0 have been also considered, but their results (not reported) show similar patterns as when $\phi_1 = 0.5$.

All simulation results reported in this essay were performed by using 20,000 replications. To begin with, we consider the size of the two tests by assuming $\delta_1 = 0$ in the DGP, so that the null of no cointegration is true. Table 1.3 reports the 5% rejection frequencies of the standard EG test ($t_{EG}$) and our proposed modified EG test ($t_{EG2}$). We begin by examining the baseline case where the covariate term is a white noise with $\psi = 0$, and it is not correlated with the regression error so that $\theta = 0$. The results in Table 1.3 show that both $t_{EG}$ and $t_{EG2}$ have fairly correct size properties in all cases that we examined. The sizes of both tests are close to the nominal value 5%, regardless of using different values of $\sigma_2^2 = \{1, 6, 16\}$ and different values of $\phi_1 = \{1.0, 0.5\}$, which are two major sources of making the signal-to-noise ratio vary. The results imply that the effects of the different variance ratios and the discrepancy of the short-run and long-run coefficients are effectively controlled for in the modified EG test. Also, these results show that the null distribution of the usual EG test is not affected by these factors. The usual EG test is still valid under the null when the DGP implies different signal-to-noise ratios. Similar results (not reported here) are obtained for the cases with larger sample sizes, such as $T$ = 500 or 1,000.

Next, we allow for more persistent covariates with $\psi \neq 0$, while assuming no correlation between errors ($\theta = 0$). The results in Table 1.4 show that the net effects of persistent covariates are at the minimum, and the sizes of both tests are not affected much. Notice that the covariate term $\Delta x_{2t}$ is still stationary as long as $\psi$ < 1. The last issue in examining the size property is the case where the error terms of $\Delta x_{1t}$ and $\Delta x_{2t}$ are correlated such that $\theta \neq 0$. From Table 1.5, we note that size properties are not affected by the covariance structure when the covariate is a white noise (i.e.,$\psi = 0$), even when errors are correlated ($\theta \neq 0$). Thus, both $t_{EG}$ and $t_{EG2}$ are not affected much by correlated errors under the null. The size properties are still as good as in the baseline case, regardless of different values of $\sigma_2^2$ and $\phi_1$. However, we observe minor size distortions for the modified EG test in the presence of combined effects where $\Delta x_{2t}$ is not a white noise but is persistent (for example, $\psi = 0.5$) and the errors are correlated (for example, $\theta = 0.5$). Nonetheless, we do not observe any severe size distortions from the usual EG test even in such cases. One possible conjecture for these results is that the nuisance parameter $\rho^2$ in (1.2.9) can hardly be estimated precisely by using the nonparametric estimation scheme in (1.2.10) when these combined effects exist. The modified EG test can be affected accordingly, while the usual EG test is less affected since it does not require the estimate of the nuisance parameter. However, we expect that when $\Delta x_{2t}$ is highly persistent and errors are also highly correlated, many, if not all, of the usual cointegration tests will be affected accordingly. Despite this, it is encouraging that the size distortions tend to be reduced significantly even in the presence of strongly combined effects when the ratio of the error variance is large (for example, $\sigma_2^2$ = 16).

Now, we turn to the power properties of the modified EG test. Empirical rejection rates for the false null are calculated by using the data generated with $\delta_1 = -0.1$. For all results, we report the size-adjusted power at the 5% level. Table 1.6 displays the results in the baseline case with $\psi = 0$ and $\theta = 0$. When $\phi_1$ = 1, the power gain of the modified EG test ($t_{EG2}$) over the usual EG test ($t_{EG}$) is not significant in all cases using different models or different values of $\sigma_2^2$. This result is as expected. However, when the long-run parameter differs from the short-run parameter (for example, $\phi_1$ = 0.5), we observe quite different power properties. As the variance of $\Delta x_{2t}$ increases ($\sigma_2^2$ = 1, 6, 16), the power of $t_{EG2}$ increases drastically while the power of $t_{EG}$ decreases significantly. Thus, the power gain is greater when the variance of $\Delta x_{2t}$ gets larger. When $\sigma_2^2$ = 16, the power of $t_{EG2}$ increases up to almost 1 even when *T* = 100. The power loss issue of $t_{EG}$ was initially noted in Kremer et al. (1992), and our results confirm this finding. However, the results in Table 1.6 show clearly that the power loss of $t_{EG}$ is resolved in the suggested modified EG test $t_{EG2}$. Instead of losing power, $t_{EG2}$ gains power as the signal-to-noise ratio increases. Thus, the signal-to-noise ratio affects the power of $t_{EG2}$ in the opposite way.

The effect of more persistent covariates on power is examined in Table 1.7. There seems no clear pattern to explain all results clearly under different magnitudes of the persistency measure and different variance ratios. In some cases, increasing persistency in the covariate helps to increase the testing efficiency up to a certain level of $\psi$, but the power of both tests tends to decrease, when the covariate term is highly persistent ($\psi$ = 0.9, for example). In the other cases, power goes down directly as the persistency goes up. One thing that we would like to point out is that the improved power of the modified EG tests becomes more evident when $\phi$ differs from the long-run coefficient value of $\beta$ = 1.

In Table 1.8, we examine the power of the two tests when the errors are correlated ($\theta \neq 0$ ). We are also interested in examining the effects on the power in the presence of combined effects where $\Delta x_{2t}$ is persistent and the errors are correlated as well. First, when $\phi = 1$, neither test is much affected by the combined factors of persistent covariates and correlated errors. These results are as expected. Second, when $\phi_1$ = 0.5, both tests show quite different behaviors. Looking at the net effect of correlated errors when $\psi$ = 0, we find that $t_{EG2}$ tends to gain power and $t_{EG}$ loses power, as $\sigma_2^2$ increases (thereby, the signal-to-noise ratio also increases). We still observe the familiar pattern that as $\sigma_2^2$ increases, the power of $t_{EG2}$ tends to increase. However, the power of $t_{EG2}$ does not necessarily increase as $\psi$ increases when $\theta \neq 0$. The results are mixed and do not show a clear pattern of the direction of the power in the case where both $\psi \neq 0$ and $\theta \neq 0$ hold. In general, the power of $t_{EG2}$ is lower than in the cases of no correlations compared to the results in Table 1.7. Nevertheless, we observe that $t_{EG}$ suffers from loss of power by correlated errors much more seriously than $t_{EG2}$; the power of $t_{EG}$ is almost 0 under the same condition. However, when $T$ = 300, both $t_{EG}$ and $t_{EG2}$ tests gain considerably more power compared to the case with $T$ = 100 even in the presence of the combined effects of persistency and correlated errors.

To visually investigate the source of power improvement, we examine the power of these tests when some of the parameters change. Since $var(\epsilon_t) = var(v_t) - \phi_1^2 var(\Delta x_{2t})$[2], we can vary either $\phi_1$ (the coefficient of $\Delta x_{2t}$) or $\sigma_2^2$ (the variance of $\Delta x_{2t}$), while holding the other parameter constant as in the baseline case. In Figure 1.1 and 1.2, we show the plot of the power function when $\phi_1$ changes from 1.0 to 1.9 by increments of 0.1. It is clear that the power of $t_{EG2}$ increases monotonically as $\phi_1$ increases. On the other hand, the power of $t_{EG}$ decreases monotonically. In Figure 1.3 and 1.4, we provide the plot of the power function when $\sigma_2^2$ changes from 1 to 16. Again, we observe the same pattern that $t_{EG2}$ gains power but $t_{EG}$ loses power as $\sigma_2^2$ increases. In sum, magnitude of the variance of covariates and the covariate term itself work in the direction of improving the power for $t_{EG2}$, while the effect on $t_{EG}$ is the opposite.

## 1.4. Concluding Remarks

In this essay, we have proposed an improved testing procedure for the Engle and Granger (1987; EG) cointegration test. In particular, it is found that the size and power properties of the modified EG test are fairly good. First, the dependency of the modified EG test on the nuisance parameter can be resolved when the parameter can be estimated by using a nonparametric estimation procedure. Second, the modified EG test tends to gain improved power as the signal-

[2] Here, $var(\epsilon_t) = var(v_t) + \phi^2 var(\Delta x_{2t}) - 2cov(v_t, \phi\Delta x_{2t}) = var(v_t) - \phi^2 var(\Delta x_{2t})$, since $cov(v_t, \phi\Delta x_{2t}) = cov(\phi\Delta x_{2t} + \epsilon_t, \phi\Delta x_{2t}) = 2\phi^2 var(\Delta x_{2t})$ when $cov(\epsilon_t, \Delta x_{2t}) = 0$.

to-noise ratio increases, while the usual EG test loses power under the same situation. Thus, the signal-to-noise ratio affects the power of the modified EG tests in the opposite way. Clearly, the signal-to-noise ratio becomes a useful factor to improve the power of the modified EG cointegration test, despite the fact that it was regarded as the source of the problem of losing power in the usual EG test. In summary, it is encouraging that the improvement of the modified EG tests over the usual EG tests is highly significant in general.

Table 1.1 Critical values of the modified EG test ($t_{EG2}$)
(model with a constant)

| *k* | $\rho^2$ | *T* = 100 | | | *T* = 1,000 | | |
|---|---|---|---|---|---|---|---|
| | | 1% | 5% | 10% | 1% | 5% | 10% |
| 1 | 0.1 | -2.964 | -2.274 | -1.910 | -2.934 | -2.243 | -1.894 |
| | 0.2 | -3.155 | -2.509 | -2.160 | -3.163 | -2.504 | -2.168 |
| | 0.3 | -3.311 | -2.673 | -2.336 | -3.314 | -2.685 | -2.340 |
| | 0.4 | -3.465 | -2.834 | -2.495 | -3.448 | -2.818 | -2.465 |
| | 0.5 | -3.595 | -2.955 | -2.621 | -3.568 | -2.944 | -2.609 |
| | 0.6 | -3.673 | -3.066 | -2.736 | -3.646 | -3.016 | -2.697 |
| | 0.7 | -3.767 | -3.157 | -2.834 | -3.715 | -3.117 | -2.809 |
| | 0.8 | -3.862 | -3.246 | -2.920 | -3.818 | -3.201 | -2.895 |
| | 0.9 | -3.900 | -3.321 | -3.014 | -3.851 | -3.265 | -2.972 |
| | 1 | -4.023 | -3.394 | -3.084 | -3.884 | -3.349 | -3.052 |
| 2 | 0.1 | -3.069 | -2.401 | -2.036 | -3.030 | -2.392 | -2.042 |
| | 0.2 | -3.369 | -2.700 | -2.347 | -3.326 | -2.664 | -2.309 |
| | 0.3 | -3.584 | -2.914 | -2.571 | -3.550 | -2.902 | -2.557 |
| | 0.4 | -3.733 | -3.110 | -2.764 | -3.691 | -3.070 | -2.744 |
| | 0.5 | -3.881 | -3.257 | -2.912 | -3.820 | -3.218 | -2.888 |
| | 0.6 | -4.003 | -3.396 | -3.069 | -3.942 | -3.336 | -3.011 |
| | 0.7 | -4.133 | -3.510 | -3.191 | -4.046 | -3.473 | -3.160 |
| | 0.8 | -4.253 | -3.632 | -3.312 | -4.170 | -3.569 | -3.265 |
| | 0.9 | -4.360 | -3.723 | -3.424 | -4.256 | -3.669 | -3.373 |
| | 1 | -4.443 | -3.835 | -3.518 | -4.328 | -3.763 | -3.465 |
| 3 | 0.1 | -3.184 | -2.532 | -2.170 | -3.176 | -2.501 | -2.150 |
| | 0.2 | -3.493 | -2.851 | -2.509 | -3.478 | -2.837 | -2.498 |
| | 0.3 | -3.767 | -3.120 | -2.786 | -3.714 | -3.086 | -2.755 |
| | 0.4 | -3.997 | -3.342 | -3.001 | -3.905 | -3.294 | -2.966 |
| | 0.5 | -4.153 | -3.522 | -3.192 | -4.088 | -3.485 | -3.154 |
| | 0.6 | -4.326 | -3.707 | -3.371 | -4.234 | -3.630 | -3.316 |
| | 0.7 | -4.448 | -3.840 | -3.516 | -4.385 | -3.772 | -3.465 |
| | 0.8 | -4.587 | -3.968 | -3.647 | -4.452 | -3.891 | -3.580 |
| | 0.9 | -4.741 | -4.095 | -3.770 | -4.573 | -3.995 | -3.694 |
| | 1 | -4.799 | -4.203 | -3.883 | -4.661 | -4.104 | -3.816 |
| 4 | 0.1 | -3.307 | -2.627 | -2.270 | -3.306 | -2.611 | -2.247 |
| | 0.2 | -3.688 | -3.022 | -2.665 | -3.690 | -3.005 | -2.644 |
| | 0.3 | -3.956 | -3.307 | -2.960 | -3.917 | -3.286 | -2.946 |
| | 0.4 | -4.198 | -3.574 | -3.227 | -4.130 | -3.503 | -3.180 |
| | 0.5 | -4.418 | -3.766 | -3.425 | -4.324 | -3.711 | -3.383 |
| | 0.6 | -4.602 | -3.956 | -3.628 | -4.468 | -3.881 | -3.564 |
| | 0.7 | -4.715 | -4.113 | -3.790 | -4.639 | -4.040 | -3.732 |
| | 0.8 | -4.909 | -4.296 | -3.964 | -4.745 | -4.187 | -3.885 |
| | 0.9 | -5.041 | -4.431 | -4.103 | -4.880 | -4.305 | -4.006 |
| | 1 | -2.964 | -2.274 | -1.910 | -2.934 | -2.243 | -1.894 |

Table 1.2 Critical values of the modified EG test ($t_{EG2}$)
(model with a trend)

| $k$ | $\rho^2$ | $T$ = 100 | | | $T$ = 1,000 | | |
|---|---|---|---|---|---|---|---|
| | | 1% | 5% | 10% | 1% | 5% | 10% |
| 1 | 0.1 | -3.138 | -2.448 | -2.070 | -3.086 | -2.422 | -2.059 |
| | 0.2 | -3.373 | -2.711 | -2.382 | -3.343 | -2.696 | -2.355 |
| | 0.3 | -3.607 | -2.956 | -2.622 | -3.541 | -2.935 | -2.595 |
| | 0.4 | -3.790 | -3.146 | -2.807 | -3.736 | -3.116 | -2.779 |
| | 0.5 | -3.929 | -3.312 | -2.984 | -3.868 | -3.244 | -2.930 |
| | 0.6 | -4.040 | -3.413 | -3.096 | -4.008 | -3.397 | -3.074 |
| | 0.7 | -4.171 | -3.568 | -3.242 | -4.089 | -3.498 | -3.185 |
| | 0.8 | -4.288 | -3.667 | -3.356 | -4.194 | -3.607 | -3.307 |
| | 0.9 | -4.390 | -3.782 | -3.473 | -4.278 | -3.697 | -3.404 |
| | 1 | -4.500 | -3.881 | -3.573 | -4.325 | -3.788 | -3.502 |
| 2 | 0.1 | -3.208 | -2.546 | -2.185 | -3.186 | -2.526 | -2.171 |
| | 0.2 | -3.559 | -2.899 | -2.546 | -3.528 | -2.870 | -2.520 |
| | 0.3 | -3.803 | -3.155 | -2.806 | -3.761 | -3.125 | -2.773 |
| | 0.4 | -3.986 | -3.368 | -3.032 | -3.946 | -3.325 | -2.992 |
| | 0.5 | -4.230 | -3.568 | -3.223 | -4.089 | -3.499 | -3.167 |
| | 0.6 | -4.340 | -3.717 | -3.386 | -4.263 | -3.642 | -3.330 |
| | 0.7 | -4.497 | -3.854 | -3.539 | -4.382 | -3.780 | -3.482 |
| | 0.8 | -4.648 | -4.012 | -3.686 | -4.483 | -3.905 | -3.616 |
| | 0.9 | -4.746 | -4.125 | -3.805 | -4.568 | -4.026 | -3.727 |
| | 1 | -4.845 | -4.235 | -3.925 | -4.686 | -4.130 | -3.845 |
| 3 | 0.1 | -3.347 | -2.652 | -2.289 | -3.291 | -2.626 | -2.275 |
| | 0.2 | -3.674 | -3.042 | -2.681 | -3.638 | -2.977 | -2.635 |
| | 0.3 | -4.020 | -3.348 | -2.983 | -3.923 | -3.289 | -2.952 |
| | 0.4 | -4.211 | -3.577 | -3.239 | -4.163 | -3.517 | -3.186 |
| | 0.5 | -4.418 | -3.764 | -3.438 | -4.339 | -3.729 | -3.406 |
| | 0.6 | -4.607 | -3.971 | -3.639 | -4.457 | -3.881 | -3.570 |
| | 0.7 | -4.782 | -4.140 | -3.815 | -4.639 | -4.046 | -3.736 |
| | 0.8 | -4.937 | -4.313 | -3.991 | -4.794 | -4.198 | -3.898 |
| | 0.9 | -5.058 | -4.443 | -4.126 | -4.908 | -4.340 | -4.038 |
| | 1 | -5.192 | -4.571 | -4.264 | -4.991 | -4.452 | -4.165 |
| 4 | 0.1 | -3.388 | -2.739 | -2.380 | -3.383 | -2.710 | -2.359 |
| | 0.2 | -3.860 | -3.154 | -2.817 | -3.781 | -3.137 | -2.790 |
| | 0.3 | -4.160 | -3.491 | -3.148 | -4.061 | -3.447 | -3.110 |
| | 0.4 | -4.387 | -3.770 | -3.428 | -4.321 | -3.703 | -3.381 |
| | 0.5 | -4.640 | -4.002 | -3.664 | -4.507 | -3.919 | -3.593 |
| | 0.6 | -4.845 | -4.218 | -3.888 | -4.726 | -4.123 | -3.804 |
| | 0.7 | -5.034 | -4.409 | -4.080 | -4.882 | -4.295 | -3.993 |
| | 0.8 | -5.180 | -4.565 | -4.244 | -5.027 | -4.463 | -4.159 |
| | 0.9 | -5.397 | -4.746 | -4.432 | -5.137 | -4.588 | -4.294 |
| | 1 | -5.546 | -4.902 | -4.581 | -5.277 | -4.734 | -4.441 |

Table 1.3 Size Property: Baseline case $(\psi = \theta = 0)$

| Models | $\phi_1$ | $\sigma_2^2$ | $T=100$ $t_{EG}$ | $T=100$ $t_{EG2}$ | $T=300$ $t_{EG}$ | $T=300$ $t_{EG2}$ |
|---|---|---|---|---|---|---|
| constant | 1 | 1 | 0.051 | 0.050 | 0.051 | 0.050 |
| | | 6 | 0.052 | 0.053 | 0.052 | 0.052 |
| | | 16 | 0.052 | 0.053 | 0.050 | 0.049 |
| | 0.5 | 1 | 0.050 | 0.053 | 0.050 | 0.050 |
| | | 6 | 0.052 | 0.049 | 0.050 | 0.048 |
| | | 16 | 0.051 | 0.051 | 0.051 | 0.050 |
| constant and trend | 1 | 1 | 0.049 | 0.054 | 0.052 | 0.055 |
| | | 6 | 0.048 | 0.049 | 0.050 | 0.054 |
| | | 16 | 0.049 | 0.052 | 0.049 | 0.053 |
| | 0.5 | 1 | 0.047 | 0.049 | 0.051 | 0.051 |
| | | 6 | 0.052 | 0.055 | 0.049 | 0.052 |
| | | 16 | 0.047 | 0.052 | 0.051 | 0.052 |

Table 1.4 Size Property: Effects of Persistent Covariates

| $\phi_1$ | $\psi$ | $\sigma_2^2$ | $T = 100$ | | $T = 300$ | |
|---|---|---|---|---|---|---|
| | | | $t_{EG}$ | $t_{EG2}$ | $t_{EG}$ | $t_{EG2}$ |
| 1 | 0.0 | 1 | 0.049 | 0.051 | 0.051 | 0.050 |
| | | 6 | 0.052 | 0.053 | 0.052 | 0.052 |
| | | 16 | 0.052 | 0.053 | 0.050 | 0.049 |
| | 0.6 | 1 | 0.051 | 0.052 | 0.033 | 0.047 |
| | | 6 | 0.036 | 0.051 | 0.034 | 0.047 |
| | | 16 | 0.040 | 0.054 | 0.035 | 0.047 |
| | 0.9 | 1 | 0.050 | 0.050 | 0.034 | 0.047 |
| | | 6 | 0.038 | 0.059 | 0.037 | 0.052 |
| | | 16 | 0.043 | 0.064 | 0.036 | 0.050 |
| 0.5 | 0.0 | 1 | 0.049 | 0.052 | 0.050 | 0.050 |
| | | 6 | 0.052 | 0.049 | 0.050 | 0.048 |
| | | 16 | 0.051 | 0.051 | 0.051 | 0.050 |
| | 0.6 | 1 | 0.052 | 0.051 | 0.032 | 0.046 |
| | | 6 | 0.037 | 0.056 | 0.033 | 0.046 |
| | | 16 | 0.037 | 0.051 | 0.035 | 0.047 |
| | 0.9 | 1 | 0.049 | 0.050 | 0.036 | 0.052 |
| | | 6 | 0.044 | 0.066 | 0.037 | 0.051 |
| | | 16 | 0.045 | 0.064 | 0.039 | 0.056 |

Note: These results are obtained for the model with a constant. The results for the model with a trend show similar patterns, and these results are omitted in this table and Table 5 to save space.

Table 1.5 Size Property: Combined Effects of Persistent Covariates and Correlated Errors

| $\phi_1$ | $\psi$ | $\theta$ | $\sigma_2^2$ | $T = 100$ | | $T = 300$ | |
|---|---|---|---|---|---|---|---|
| | | | | $t_{EG}$ | $t_{EG2}$ | $t_{EG}$ | $t_{EG2}$ |
| 1 | 0.0 | 0.3 | 1 | 0.054 | 0.054 | 0.051 | 0.051 |
| | | | 6 | 0.050 | 0.051 | 0.050 | 0.050 |
| | | | 16 | 0.049 | 0.049 | 0.050 | 0.051 |
| | | 0.5 | 1 | 0.052 | 0.053 | 0.049 | 0.049 |
| | | | 6 | 0.051 | 0.050 | 0.048 | 0.048 |
| | | | 16 | 0.051 | 0.050 | 0.051 | 0.050 |
| | 0.5 | 0.3 | 1 | 0.041 | 0.071 | 0.039 | 0.069 |
| | | | 6 | 0.039 | 0.057 | 0.036 | 0.052 |
| | | | 16 | 0.037 | 0.054 | 0.038 | 0.049 |
| | | 0.5 | 1 | 0.048 | 0.123 | 0.046 | 0.117 |
| | | | 6 | 0.042 | 0.062 | 0.036 | 0.058 |
| | | | 16 | 0.036 | 0.051 | 0.036 | 0.052 |
| 0.5 | 0.0 | 0.3 | 1 | 0.051 | 0.052 | 0.051 | 0.051 |
| | | | 6 | 0.051 | 0.050 | 0.049 | 0.047 |
| | | | 16 | 0.049 | 0.051 | 0.050 | 0.050 |
| | | 0.5 | 1 | 0.052 | 0.053 | 0.052 | 0.050 |
| | | | 6 | 0.049 | 0.050 | 0.053 | 0.053 |
| | | | 16 | 0.049 | 0.050 | 0.050 | 0.052 |
| | 0.5 | 0.3 | 1 | 0.038 | 0.068 | 0.038 | 0.069 |
| | | | 6 | 0.038 | 0.057 | 0.037 | 0.052 |
| | | | 16 | 0.037 | 0.056 | 0.035 | 0.049 |
| | | 0.5 | 1 | 0.045 | 0.117 | 0.050 | 0.124 |
| | | | 6 | 0.039 | 0.059 | 0.035 | 0.054 |
| | | | 16 | 0.038 | 0.055 | 0.037 | 0.053 |

Table 1.6 Power Property: Baseline case $(\psi = \theta = 0)$

| Models | $\phi_1$ | $\sigma_2^2$ | $T = 100$ | | $T = 300$ | |
|---|---|---|---|---|---|---|
| | | | $t_{EG}$ | $t_{EG2}$ | $t_{EG}$ | $t_{EG2}$ |
| constant | 1 | 1 | 0.215 | 0.257 | 0.968 | 0.980 |
| | | 6 | 0.231 | 0.261 | 0.967 | 0.980 |
| | | 16 | 0.218 | 0.253 | 0.965 | 0.978 |
| | 0.5 | 1 | 0.204 | 0.311 | 0.956 | 0.995 |
| | | 6 | 0.070 | 0.945 | 0.864 | 1.000 |
| | | 16 | 0.059 | 1.000 | 0.857 | 1.000 |
| constant and trend | 1 | 1 | 0.163 | 0.177 | 0.876 | 0.899 |
| | | 6 | 0.151 | 0.166 | 0.876 | 0.898 |
| | | 16 | 0.167 | 0.176 | 0.868 | 0.885 |
| | 0.5 | 1 | 0.128 | 0.180 | 0.827 | 0.952 |
| | | 6 | 0.015 | 0.695 | 0.487 | 1.000 |
| | | 16 | 0.010 | 0.982 | 0.409 | 1.000 |

Table 1.7 Power Property: Effects of Persistent Covariates

| $\phi_1$ | $\psi$ | $\sigma_2^2$ | $T = 100$ | | $T = 300$ | |
|---|---|---|---|---|---|---|
| | | | $t_{EG}$ | $t_{EG2}$ | $t_{EG}$ | $t_{EG2}$ |
| 1 | 0 | 1 | 0.215 | 0.257 | 0.968 | 0.980 |
| | | 6 | 0.231 | 0.261 | 0.967 | 0.980 |
| | | 16 | 0.218 | 0.253 | 0.965 | 0.978 |
| | 0.6 | 1 | 0.276 | 0.262 | 0.986 | 0.978 |
| | | 6 | 0.262 | 0.254 | 0.985 | 0.977 |
| | | 16 | 0.261 | 0.246 | 0.987 | 0.978 |
| | 0.9 | 1 | 0.243 | 0.247 | 0.982 | 0.967 |
| | | 6 | 0.241 | 0.238 | 0.982 | 0.967 |
| | | 16 | 0.248 | 0.236 | 0.982 | 0.970 |
| 0.5 | 0 | 1 | 0.204 | 0.311 | 0.956 | 0.995 |
| | | 6 | 0.070 | 0.945 | 0.864 | 1.000 |
| | | 16 | 0.059 | 1.000 | 0.857 | 1.000 |
| | 0.6 | 1 | 0.067 | 0.453 | 0.540 | 1.000 |
| | | 6 | 0.012 | 0.995 | 0.008 | 1.000 |
| | | 16 | 0.013 | 0.999 | 0.008 | 1.000 |
| | 0.9 | 1 | 0.012 | 0.582 | 0.005 | 1.000 |
| | | 6 | 0.033 | 0.851 | 0.007 | 1.000 |
| | | 16 | 0.037 | 0.853 | 0.009 | 1.000 |

Note: The results for the model with a trend show similar patterns, and these results are omitted in this table and Table 8 to save space.

Table 1.8 Power Property: Combined Effects of Persistent Covariates and Correlated Errors

| $\phi_1$ | $\psi$ | $\theta$ | $\sigma_2^2$ | $T=100$ | | $T=300$ | |
|---|---|---|---|---|---|---|---|
| | | | | $t_{EG}$ | $t_{EG2}$ | $t_{EG}$ | $t_{EG2}$ |
| 1 | 0.0 | 0.3 | 1 | 0.197 | 0.272 | 0.959 | 0.986 |
| | | | 6 | 0.222 | 0.259 | 0.963 | 0.978 |
| | | | 16 | 0.235 | 0.271 | 0.967 | 0.978 |
| | | 0.5 | 1 | 0.182 | 0.317 | 0.950 | 0.995 |
| | | | 6 | 0.218 | 0.271 | 0.967 | 0.984 |
| | | | 16 | 0.221 | 0.273 | 0.964 | 0.978 |
| | 0.5 | 0.3 | 1 | 0.249 | 0.332 | 0.976 | 0.989 |
| | | | 6 | 0.261 | 0.272 | 0.982 | 0.980 |
| | | | 16 | 0.259 | 0.258 | 0.981 | 0.976 |
| | | 0.5 | 1 | 0.216 | 0.463 | 0.964 | 0.998 |
| | | | 6 | 0.248 | 0.288 | 0.983 | 0.986 |
| | | | 16 | 0.277 | 0.286 | 0.982 | 0.979 |
| 0.5 | 0.0 | 0.3 | 1 | 0.217 | 0.257 | 0.964 | 0.983 |
| | | | 6 | 0.133 | 0.510 | 0.924 | 1.000 |
| | | | 16 | 0.092 | 0.800 | 0.889 | 1.000 |
| | | 0.5 | 1 | 0.221 | 0.259 | 0.962 | 0.976 |
| | | | 6 | 0.144 | 0.488 | 0.914 | 1.000 |
| | | | 16 | 0.093 | 0.782 | 0.888 | 1.000 |
| | 0.5 | 0.3 | 1 | 0.141 | 0.151 | 0.850 | 0.971 |
| | | | 6 | 0.017 | 0.724 | 0.137 | 1.000 |
| | | | 16 | 0.010 | 0.953 | 0.051 | 1.000 |
| | | 0.5 | 1 | 0.180 | 0.078 | 0.890 | 0.782 |
| | | | 6 | 0.018 | 0.657 | 0.137 | 1.000 |
| | | | 16 | 0.011 | 0.945 | 0.044 | 1.000 |

Figure 1.1 Power Function under Different Magnitudes of Coefficients ($\sigma_2^2 = 1$)

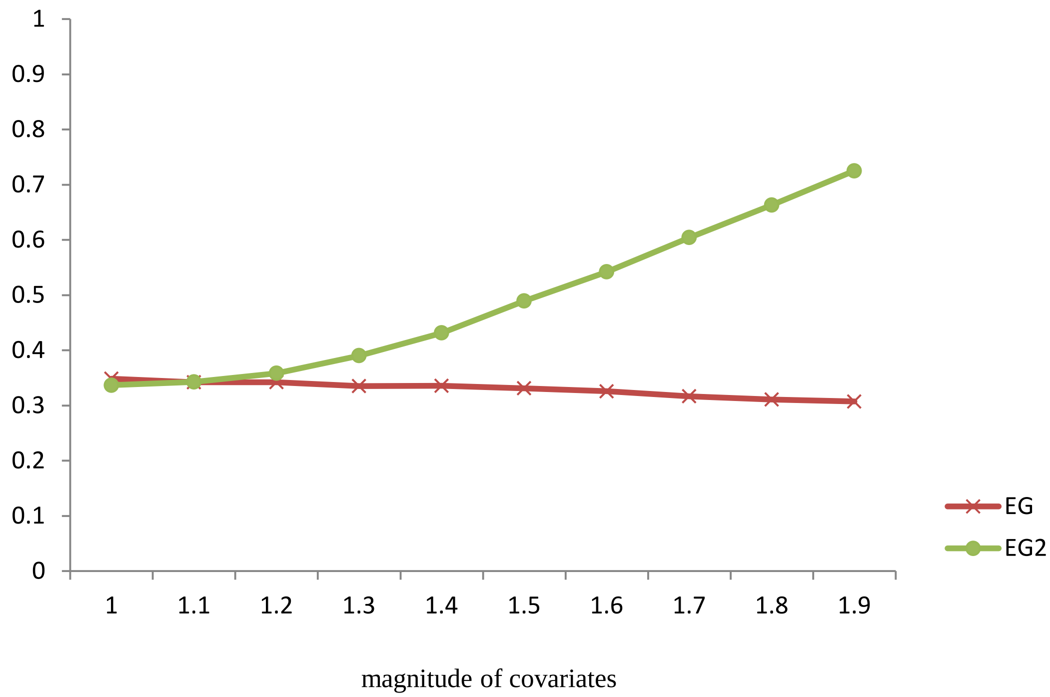


Figure 1.2 Power Function under Different Magnitudes of the Coefficients ($\sigma_2^2 = 6$)

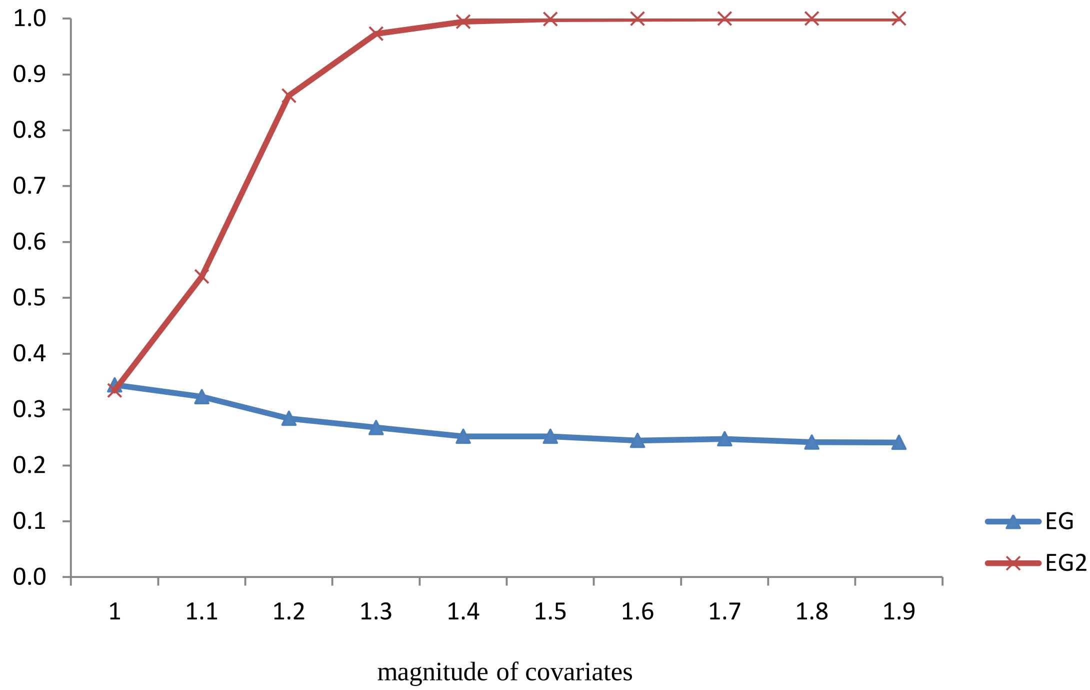


Figure 1.3 Power Function under Different Values of the Variance ($\phi_1 = 0.1$)

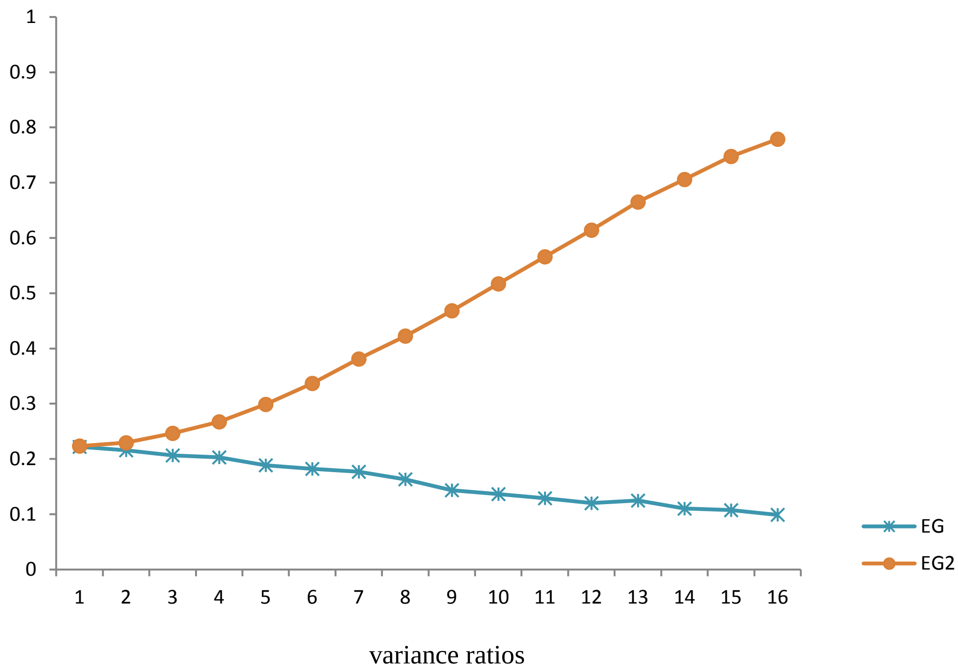
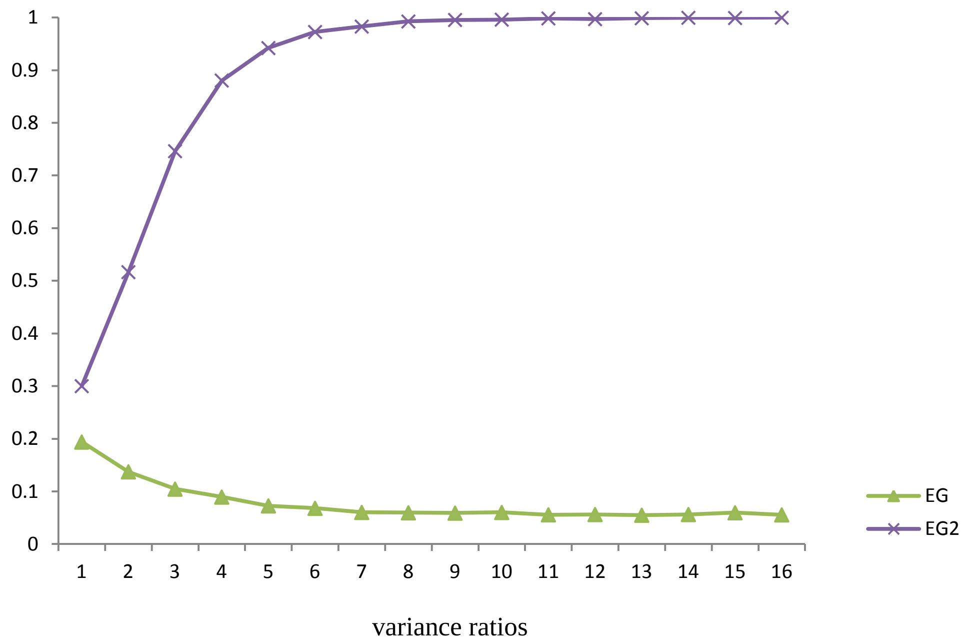


Figure 1.4. Power Function under Different Values of the Variance ($\phi_1 = 0.5$)



# CHAPTER 2

## HOW DO NONLINEAR UNIT ROOT TESTS PERFORM WITH NON-NORMAL ERRORS?

The usual unit root tests assume: (i) a linear model, (ii) the errors in the testing regression are white noise, and (iii) the errors are normally distributed. For example, the critical values of the pioneering Dickey-Fuller (DF, 1979) tests were derived under these assumptions. A relevant question is, "What will happen to the usual unit root tests if these conditions are not met?" While the first two assumptions have been widely investigated in the literature, to the best of our knowledge, there has been no examination of the consequences of the third assumption regarding non-normal errors.

Regarding the second assumption, many papers have already examined cases with non-white noise error term.[1] Our interest lies in the first and third assumptions. Regarding the first, widely known evidence demonstrates that the usual linear unit root tests will lose power in the presence of nonlinearity. Various nonlinear unit root tests suggest that while the usual linear unit root tests lose power when the underlying model is nonlinear, the suggested solutions can potentially recover this loss of power. These tests assume a particular nonlinear function typically with normal errors. However, whether nonlinear tests are robust to various model

[1] In particular, the popular literature documents how to control for serially correlated errors in unit root tests. The augmented Dickey-Fuller tests (1979) and the nonparametric tests of Phillips and Perron (1988) have been very popular in this regard. Subsequent studies examined the cases in which certain heterogeneity exists in the error term. The usual unit root tests, if properly modified, can be valid asymptotically if certain regularity conditions are satisfied. Hall and Heyde (1980) note the required condition that the errors need to be martingale differences, under which the martingale central limit theorem will apply. In this regard, Kim and Schmidt (1995) examined the performance of the DF tests with IGARCH errors.

specifications is a question. For example, one may wonder if the nonlinear tests will maintain power in the reverse case in which the underlying model is linear. In such a case, our examination shows that nonlinear unit root tests can be generally less powerful than linear unit root tests.

In this essay, however, we are primarily interested in the third assumption of normal errors. The literature has been quiet regarding how the usual unit root tests perform when the errors are non-normal. Still, the common practice is to tabulate critical values for nonstandard distributions in different types of linear and nonlinear unit root tests, while assuming a normal distribution in the error term. Perhaps, this common practice relies on the presumption that the asymptotic distributions in these tests will be unaffected if the error terms are non-normal. Actually, this presumption holds trivially for the usual linear unit root tests. In particular, certain regularity conditions for the validity of linear unit root tests do not involve the requirement that the error should follow a normal distribution; see Phillips and Perron (1988) and Hall and Heyde (1980). Indeed, in light of the Gauss-Markov theorem and the functional central limit theorem, non-normal errors will not pose a problem, at least asymptotically, for linear unit root tests. The least squares estimators from linear models are the most efficient and the resulting test statistics have a stable distribution regardless of whether the distribution of the error term is normal. Thus, the usual linear unit root tests will be unaffected even when the errors are non-normal. We confirm this result later in this essay. Nevertheless, whether similar results will carry over to nonlinear unit root tests with non-normal errors is a question that merits examination. To our surprise, however, the extant literature on nonlinear unit root tests has not examined this important question. Thus, our focus in the present essay is to examine how nonlinear unit root tests perform in the presence of non-normal errors.

## 2.1. Nonlinear Unit Root Tests

Our goal in this essay is to examine the performance of unit root tests in the presence of non-normal errors. We pay special attention to nonlinear unit root tests, and we also examine the linear DF unit root tests as a reference point. For this purpose, we consider three popular nonlinear unit root tests: (i) the exponential smooth transition autoregressive (ESTAR) nonlinear unit root tests of Kapetanios, Shin and Snell (2003); (ii) the *sign* test of So and Shin (2001); and (iii) the *inf-t* test of Park and Shintani (2005). Choi and Moh (2007) previously examined these three nonlinear unit root tests under various nonlinear model specifications.

All these nonlinear unit root tests assume a linear model under the null of a unit root and a specific nonlinear model under the alternative hypothesis. However, nonlinearity can arise both under the null hypothesis of a unit root and under the alternative hypothesis of stationarity. In this essay, we consider non-normal errors under both the null and alternative hypotheses.

We will briefly explain the three nonlinear unit root tests employed in our analysis. The nonlinear unit root test of Kapetanios *et al*. (2003) considers the linear unit root hypothesis against the nonlinear alternative of an exponentially smooth transition autoregressive (ESTAR) model

$$\Delta y_t = \gamma y_{t-1}[1 - exp(-\theta y_{t-1}^2)] + u_t. \tag{2.1.1}$$

Testing on $\theta = 0$ amounts to testing linearity versus nonlinearity, where we have $\gamma = 0$ under the null hypothesis of a unit root. Furthermore, testing on either $\theta = 0$ or $\gamma = 0$ involves a nuisance parameter problem in which the other parameter is present only under the alternative of each test. Thus, Kapetanios et al. (2003) assume the null of a linear unit root and utilize the following regression from a first-order Taylor expansion

$$\Delta y_t = \delta y_{t-1}^3 + e_t. \tag{2.1.2}$$

They consider the null of the unit root as $\delta = 0$ and tabulate critical values of the corresponding t-statistic, which we denote as *KSS*.

The *sign* test of So and Shin (2001) is a nonparametric test for the unit-root null against the alternative of a general linear and non-linear stationary AR process. Suppose that $y_t$, $t = 1, .., T$, denotes an observed time series following a possibly nonlinear process. The *sign* test is given as

$$S_T(1) = \sum_{t=1}^{T} sign(y_t - y_{t-1}) sign(y_{t-1} - \hat{m}_{t-1}) \tag{2.1.3}$$

where $\hat{m}_{t-1}$ is the median of $\{y_i\}_{i=1}^{t-1}$. The null of a unit root hypothesis is rejected if $S_T(1) \leq 2B_T(\alpha) - T$ where $2B_T(\alpha)$ is the lower $\alpha_{th}$ quantile of the binomial distribution, $binomial(T, 0.5)$.

Park and Shintani (2005) consider general transitional AR models covering a wide range of threshold, discrete and smooth transition dynamics. We consider the *inf-t* test which is based on the following model

$$\Delta y_t = \lambda(\theta) y_{t-1} \pi(y_{t-d}, \theta) + c_1 \Delta y_{t-1} + u_t, \tag{2.1.4}$$

where $y_{t-d}$ is the transition variable with a delay parameter $d$, and $\theta$ is the parameter describing a real valued transition function $\pi$. The null hypothesis implies that $\lambda = 0$. Here, the expression $\lambda(\theta)$ is employed to denote that $\lambda$ is a function of the nuisance parameter $\theta$, which can be identified only under the alternative hypothesis of stationarity. Again, the null implies a linear model, and nonlinearity exists only under the alternative hypothesis. We consider the ESTAR model for the transition function $\pi = [1 - exp(-\theta y_{t-1}^2)]$, as described in (2.1.1), but other nonlinear models may also be considered. Park and Shintani (2005) consider the following *t*-statistic on $\lambda = 0$ against $\lambda < 0$

$$t_n(\theta) = \frac{\hat{\lambda}_n(\theta)}{s(\hat{\lambda}_n(\theta))} \tag{2.1.5}$$

where $s(\hat{\lambda}_n(\theta))$ is the standard error of the estimate $\hat{\lambda}_n(\theta)$. Then, the *inf-t* test is obtained as the minimum of the above *t*-statistics over a wide range of values for the transition parameter $\theta$ via a grid-search

$$t_n = \inf_{\theta \in \theta_n} t_n(\theta), \tag{2.1.6}$$

where the limiting distribution of the *inf-t* test ($t_n$) depends on the transition function.

## 2.2. Monte Carlo Experiment

To conduct our Monte Carlo simulations, we consider the data generating process (DGP)

$$y_t = z_t'\delta + e_t, \tag{2.2.1}$$

$$e_t = \beta e_{t-1} + u_t,$$

where $z_t = [1]$ in the model with a constant, and $z_t = [1, t]'$ in the model with a constant and trend. Any valid tests should be invariant to the coefficients $\delta$ in the corresponding DGP. Whether these tests are invariant $\delta$ is not the focus of the present essay, and we choose $\delta = 0$ in the DGP and assume a linear model under the null. We examine the size properties under the null hypothesis, $\beta = 1.0$, and use $\beta = 0.9$ to examine the power of the tests.

Here, $u_t$ is allowed to follow various types of distributions. Specifically, we consider

(i) standard normal distribution

(ii) – (v) chi-square distributions with df = 1, 2, 3, and 4

(vi) – (ix) t-distributions with df = 2, 3, 4, and 8

(x) double exponential distribution with df = 1, symmetric around 0.5

(xi) beta distribution, B(2, 2).

We focus on examining the performance in small samples with $T$ = 50 and 100, and report the results in Tables 2.1 and 2.2. However, we also consider larger samples with $T$ = 300 and 500 to examine the asymptotic behavior of the tests; these results are shown in Tables 2.3 and 2.4. We report the rejection frequencies of a one-sided test at the 5% significance level. All of the results are produced with 20,000 replications. Since we do not allow for serial correlations, we do not include augmented lagged terms in the testing regression. However, one augmentation lag is included only for the *inf-t* test, following the suggestion of Park and Shintani (2005); see equation (2.1.4).

Table 2.1 reports the results for the model with a constant. The results for the model with a trend are displayed in Table 2.2. We first examine the size properties. The 5% rejection rates (sizes) under the null are reported in Panel A in Tables 2.1 and 2.2. When the error term $u_t$ follows the standard normal distribution, all tests show correct sizes close to 5%. This is as expected. We next examine the cases with non-normal errors. It is clear that the linear DF tests have correct sizes regardless of whether $u_t$ follows various types of non-normal distributions. This result is also as expected, since the linear tests will remain robust under non-normal errors in light of the Gauss-Markov theorem and the functional central limit theorem. The *KSS* nonlinear unit root tests also exhibit good size properties. This is so, because they assume linearity under the null, as in the DGP in our simulations. However, we observe non-negligible size distortions in the *sign* test of So and Shin (2001). For example, when the error term follows the chi-square distribution with df = 1, the sizes of the *sign* test are 0.071 and 0.136, when $T$ = 50 and 100, respectively, for the model with a constant. The same pattern occurs in Table 2.2 for the model with trend. Moreover, the size distortions do not disappear when the sample size increases; see the additional results with $T$ = 300 and 500 in Tables 2.3 and 2.4. (The same result holds in unreported simulation results with *T* = 1,000.) A mild size distortion is observed in the *inf-t* test of Park and Shintani (2005) for the model with a constant when the error follows the chi-square distribution (with df = 1 and 2) and t-distribution (df = 2 and 3). The problem tends to disappear when *T* increases. The same patterns can be found for the model with a trend, while the size distortion problem for the t-distribution with df = 2 remains in large samples; see Table 2.4.

The most important implication of non-normal errors, however, lies in their power properties. Panel B in Tables 2.1 and 2.2 report the power of the tests when $\beta = 0.9$ in the DGP. Given the observed size distortions, it would be fair to examine the size-adjusted power of the tests, for which we report the results in Panel C. We observe that the power of the linear DF tests does not vary much over different non-normal distributions of the error term. This result is again as expected. The power of the DF test under non-normal errors is comparable to the power under the normal distribution, although we observe a negligible loss of power in a few cases.

Our main question of interest is the power properties of the nonlinear unit root tests. First, we examine the benchmark case when the error term follows a normal distribution. Looking at the size-adjusted powers, we note that the linear DF tests are generally more powerful than the nonlinear unit root tests. This finding may appear to contrast with the perception that nonlinear tests are more powerful than the linear DF tests. However, previous studies might have emphasized only one side of the story. Although linear tests lose power in the presence of nonlinearity, the converse also holds. Unnecessary use of nonlinear unit root tests can lead to loss of power. This loss of power is more evident in finite samples. For example, when *T* = 100, the size-adjusted powers of the *KSS* test, the sign test, and the *inf-t* test are 0.221, 0.257 and 0.235, respectively, in the model with a constant, while the power of the DF test is 0.325; see

Panel C of Table 2.1. The same result is shown in Table 2.2 for the model with a trend, where the size-adjusted powers of these tests are 0.140, 0.106 and 0.165 while the power of the DF test is 0.188. The same pattern is observed in sample sizes using $T$ = 50, 300, and 500. In sum, the linear DF tests are generally more powerful than these nonlinear tests in the benchmark case under a normal distribution.

Second, when the error term follows a non-normal distribution, it is clear that both the *KSS* test and the *inf-t* test lose power, compared to the linear DF test. The loss of power is much more evident when the distribution is skewed with lower degrees of freedom (chi-square distribution with df = 1 or 2). Thus, unlike in the linear based tests, when the errors are non-normal, all of the nonlinear unit root tests are adversely affected. These results indicates that correct specification of the error term distribution is necessary when applying nonlinear unit root tests, or else loss of power will result. This finding seems to have been neglected in the literature.

The result of power loss in the nonlinear unit root tests can be intuitively explained. As noted in Park and Phillips (1999), the convergence rates of sample functions are path dependent for nonlinear transformations of integrated processes. However, analysis using a specific nonlinear transformation is not an easy job, although it is possible to consider various types of unknown nonlinear models; see So and Shin (2001). Clearly, there is no theory to support the notion that nonlinear tests are robust to non-normal errors, although theory predicts that linear tests are unaffected by non-normal errors. Thus, the obvious analogies of linear processes do not carry over for nonlinear unit root tests.[2]

We do not find a clear pattern for the power property in the *sign* test of So and Shin (2001), except that the *sign* test loses power compared to the DF tests when the distribution of

[2] It is well known that the usual maximum likelihood estimators in cross-sectional nonlinear models, such as in Tobit or Probit models, can be inconsistent when the error term is non-normal. This paper may be the first to demonstrate that the same analogy holds for nonlinear unit root tests.

the error term is not symmetric. For example, consider the case where $u_t$ follows the chi-square distribution with df = 1. The power of the *sign* test is 0.103 (0.025) in the model with a constant (a trend), while the corresponding power of the DF test is 0.312 (0.187). As So and Shin (2001) note, the *sign* test loses power in cases with an asymmetric distribution with a median bias. Furthermore, it seems that the power loss problem is still evident even in large samples in such cases. For example, the size-adjusted power of the *sign* test with *T* = 500 are 0.031, 0.188, 0.349 and 0.504 for the model with a constant, if the error term has the chi-square distribution with df = 1, 2, 3, and 4, respectively; see Table 2.3. The power of the DF test is 1.00 in all of the corresponding cases. Note that the sign test already has size distortions when the distribution is not symmetric.

However, the *sign* test is more powerful than the linear DF test in some cases when the distribution is symmetric with heavy-tailed errors. For example, the size-adjusted power of the *sign* test is 0.506 (0.173) when $u_t$ follows the t-distribution with df = 2 in the model with a constant (a trend), while the power of the DF test is 0.275 (0.149); see Panel C in Tables 2.1 and 2.2. Thus, the power of the *sign* test depends on whether and how the information on non-normal errors can be utilized. The *sign* test can gain or lose power if it utilizes the additional information or fails to capture properly the information of the different types of non-normal errors with asymmetric distributions.

A related issue is the effect of mis-specified models when using nonlinear unit root tests. It can be reasonably argued that nonlinear tests will lose power when the underlying model is not correctly specified. Indeed, these nonlinear unit root tests are expected to be less powerful in other model specifications that differ from the particular nonlinear model designed for each test. For example, the ESTAR unit root tests can be more powerful than the DF tests when the DGP follows an ESTAR model, but less powerful than the DF tests or other nonlinear tests when different nonlinear models are more appropriate; see also Choi and Moh (2007). Our findings in this essay show a similar outcome for nonlinear unit root tests when the error term has non-normal distribution.

### 2.3. Concluding Remarks

In this essay, we have examined the performance of several nonlinear unit root tests when non-normal errors are present. We find that nonlinear unit root tests lose power when the error term follows different types of non-normal distributions. In this regard, this essay brings out the neglected point that the obvious analogies of linear processes do not necessarily hold for nonlinear models.

However, it is extremely difficult to identify a correct nonlinear model with a correct distribution of the error term. Furthermore, it seems infeasible in most cases to develop a test to distinguish different nonlinear model specifications or error distributions. There is also a sense that nonlinearity and non-normal errors are mixed in a time series. If so, it might be more desirable to consider adopting a linear model while attempting to employ the information that the errors are non-normal errors. Such an approach could lead to tests that are robust and more powerful. In particular, Essay 3 of this dissertation adopts the approach, and considers new cointegration tests that can be more powerful in the presence of non-normal errors.

Table 2.1 Size and Power of Various Tests
(Model with a Constant)

| Distribution | A. Size | | | | B. Power | | | | C. Size-adjusted Power | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *DF* | *KSS* | *sign* | *Inf-t* | *DF* | *KSS* | *sign* | *Inf-t* | *DF* | *KSS* | *sign* | *Inf-t* |
| | | | | | | $T$ = 50 | | | | | | |
| Normal | 0.052 | 0.055 | 0.058 | 0.058 | 0.122 | 0.092 | 0.116 | 0.106 | 0.119 | 0.084 | 0.116 | 0.092 |
| Chi-square df = 1 | 0.047 | 0.078 | 0.071 | 0.086 | 0.106 | 0.083 | 0.113 | 0.094 | 0.112 | 0.054 | 0.113 | 0.051 |
| Chi-square df = 2 | 0.054 | 0.068 | 0.067 | 0.076 | 0.117 | 0.080 | 0.108 | 0.101 | 0.109 | 0.057 | 0.108 | 0.061 |
| Chi-square df = 3 | 0.051 | 0.062 | 0.060 | 0.070 | 0.116 | 0.083 | 0.107 | 0.098 | 0.114 | 0.067 | 0.107 | 0.069 |
| Chi-square df = 4 | 0.052 | 0.060 | 0.058 | 0.066 | 0.120 | 0.083 | 0.111 | 0.100 | 0.115 | 0.068 | 0.111 | 0.076 |
| Student t df = 2 | 0.058 | 0.060 | 0.057 | 0.099 | 0.101 | 0.062 | 0.214 | 0.107 | 0.083 | 0.053 | 0.214 | 0.044 |
| Student t df = 3 | 0.056 | 0.056 | 0.057 | 0.082 | 0.112 | 0.073 | 0.163 | 0.107 | 0.101 | 0.066 | 0.163 | 0.062 |
| Student t df = 4 | 0.054 | 0.054 | 0.058 | 0.076 | 0.116 | 0.078 | 0.147 | 0.107 | 0.107 | 0.072 | 0.147 | 0.066 |
| Student t df = 8 | 0.057 | 0.052 | 0.061 | 0.063 | 0.122 | 0.087 | 0.131 | 0.107 | 0.109 | 0.083 | 0.131 | 0.084 |
| Double exponential | 0.055 | 0.054 | 0.058 | 0.071 | 0.118 | 0.082 | 0.169 | 0.107 | 0.108 | 0.075 | 0.169 | 0.072 |
| Beta (2,2) | 0.055 | 0.049 | 0.057 | 0.045 | 0.129 | 0.099 | 0.095 | 0.098 | 0.116 | 0.100 | 0.095 | 0.106 |
| | | | | | | $T$ = 100 | | | | | | |
| Normal | 0.047 | 0.049 | 0.061 | 0.046 | 0.313 | 0.218 | 0.257 | 0.220 | 0.325 | 0.221 | 0.257 | 0.235 |
| Chi-square df = 1 | 0.049 | 0.067 | 0.136 | 0.068 | 0.306 | 0.133 | 0.236 | 0.186 | 0.312 | 0.093 | 0.103 | 0.123 |
| Chi-square df = 2 | 0.049 | 0.059 | 0.108 | 0.060 | 0.306 | 0.163 | 0.237 | 0.193 | 0.312 | 0.134 | 0.130 | 0.164 |
| Chi-square df = 3 | 0.048 | 0.054 | 0.090 | 0.056 | 0.305 | 0.178 | 0.236 | 0.197 | 0.311 | 0.164 | 0.179 | 0.180 |
| Chi-square df = 4 | 0.048 | 0.053 | 0.085 | 0.052 | 0.307 | 0.184 | 0.239 | 0.202 | 0.319 | 0.175 | 0.180 | 0.195 |
| Student t df = 2 | 0.052 | 0.052 | 0.062 | 0.089 | 0.282 | 0.108 | 0.506 | 0.181 | 0.275 | 0.103 | 0.506 | 0.084 |
| Student t df = 3 | 0.050 | 0.054 | 0.066 | 0.071 | 0.303 | 0.154 | 0.384 | 0.205 | 0.301 | 0.145 | 0.384 | 0.136 |
| Student t df = 4 | 0.050 | 0.054 | 0.065 | 0.065 | 0.310 | 0.176 | 0.341 | 0.213 | 0.311 | 0.163 | 0.341 | 0.166 |
| Student t df = 8 | 0.051 | 0.050 | 0.064 | 0.056 | 0.316 | 0.200 | 0.289 | 0.214 | 0.309 | 0.202 | 0.289 | 0.197 |
| Double exponential | 0.049 | 0.051 | 0.062 | 0.059 | 0.311 | 0.186 | 0.412 | 0.213 | 0.317 | 0.181 | 0.412 | 0.185 |
| Beta (2,2) | 0.050 | 0.049 | 0.061 | 0.043 | 0.311 | 0.236 | 0.182 | 0.219 | 0.309 | 0.239 | 0.182 | 0.249 |

Note: The 5% significance level was used.

Table 2.2 Size and Power of Various Tests
(Model with a Trend)

| Distribution | A. Size | | | | B. Power | | | | C. Size-adjusted Power | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *DF* | *KSS* | *sign* | *Inf-t* | *DF* | *KSS* | *sign* | *Inf-t* | *DF* | *KSS* | *sign* | *Inf-t* |
| | | | | | | $T = 50$ | | | | | | |
| Normal | 0.056 | 0.041 | 0.040 | 0.060 | 0.091 | 0.062 | 0.042 | 0.095 | 0.079 | 0.074 | 0.082 | 0.081 |
| Chi-square df = 1 | 0.050 | 0.063 | 0.128 | 0.133 | 0.075 | 0.051 | 0.089 | 0.114 | 0.075 | 0.039 | 0.030 | 0.032 |
| Chi-square df = 2 | 0.056 | 0.053 | 0.082 | 0.105 | 0.084 | 0.060 | 0.061 | 0.111 | 0.075 | 0.056 | 0.031 | 0.046 |
| Chi-square df = 3 | 0.052 | 0.053 | 0.067 | 0.092 | 0.085 | 0.058 | 0.056 | 0.108 | 0.081 | 0.056 | 0.056 | 0.055 |
| Chi-square df = 4 | 0.056 | 0.050 | 0.061 | 0.085 | 0.087 | 0.056 | 0.053 | 0.103 | 0.075 | 0.056 | 0.053 | 0.057 |
| Student t df = 2 | 0.060 | 0.046 | 0.040 | 0.147 | 0.084 | 0.042 | 0.058 | 0.122 | 0.068 | 0.046 | 0.100 | 0.036 |
| Student t df = 3 | 0.057 | 0.046 | 0.040 | 0.111 | 0.086 | 0.050 | 0.050 | 0.116 | 0.073 | 0.055 | 0.089 | 0.047 |
| Student t df = 4 | 0.054 | 0.041 | 0.039 | 0.093 | 0.085 | 0.052 | 0.049 | 0.108 | 0.081 | 0.061 | 0.088 | 0.053 |
| Student t df = 8 | 0.055 | 0.044 | 0.041 | 0.072 | 0.094 | 0.057 | 0.047 | 0.106 | 0.085 | 0.064 | 0.085 | 0.074 |
| Double exponential | 0.060 | 0.046 | 0.039 | 0.090 | 0.088 | 0.055 | 0.050 | 0.105 | 0.075 | 0.061 | 0.091 | 0.057 |
| Beta (2,2) | 0.059 | 0.041 | 0.040 | 0.052 | 0.093 | 0.060 | 0.041 | 0.088 | 0.080 | 0.073 | 0.077 | 0.086 |
| | | | | | | $T = 100$ | | | | | | |
| Normal | 0.050 | 0.041 | 0.051 | 0.049 | 0.189 | 0.117 | 0.106 | 0.161 | 0.188 | 0.140 | 0.106 | 0.165 |
| Chi-square df = 1 | 0.048 | 0.050 | 0.159 | 0.097 | 0.180 | 0.072 | 0.154 | 0.153 | 0.187 | 0.073 | 0.025 | 0.060 |
| Chi-square df = 2 | 0.052 | 0.046 | 0.112 | 0.078 | 0.179 | 0.085 | 0.129 | 0.151 | 0.175 | 0.092 | 0.060 | 0.094 |
| Chi-square df = 3 | 0.050 | 0.044 | 0.089 | 0.067 | 0.183 | 0.093 | 0.120 | 0.152 | 0.182 | 0.107 | 0.081 | 0.110 |
| Chi-square df = 4 | 0.052 | 0.045 | 0.082 | 0.067 | 0.186 | 0.099 | 0.121 | 0.157 | 0.178 | 0.109 | 0.082 | 0.118 |
| Student t df = 2 | 0.054 | 0.047 | 0.053 | 0.132 | 0.164 | 0.059 | 0.173 | 0.152 | 0.149 | 0.063 | 0.173 | 0.038 |
| Student t df = 3 | 0.057 | 0.045 | 0.053 | 0.097 | 0.177 | 0.083 | 0.149 | 0.156 | 0.156 | 0.094 | 0.149 | 0.069 |
| Student t df = 4 | 0.049 | 0.043 | 0.050 | 0.078 | 0.180 | 0.093 | 0.132 | 0.156 | 0.183 | 0.109 | 0.132 | 0.094 |
| Student t df = 8 | 0.050 | 0.039 | 0.050 | 0.057 | 0.187 | 0.111 | 0.121 | 0.161 | 0.187 | 0.137 | 0.121 | 0.145 |
| Double exponential | 0.052 | 0.043 | 0.051 | 0.068 | 0.187 | 0.100 | 0.161 | 0.160 | 0.182 | 0.114 | 0.161 | 0.119 |
| Beta (2,2) | 0.052 | 0.039 | 0.055 | 0.043 | 0.192 | 0.132 | 0.090 | 0.158 | 0.187 | 0.161 | 0.090 | 0.177 |

Table 2.3 Size and Power of Various Tests in Large Samples
(Model with a Constant)

| Distribution | A. Size | | | | B. Power | | | | C. Size-adjusted Power | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *DF* | *KSS* | *sign* | *Inf-t* | *DF* | *KSS* | *sign* | *Inf-t* | *DF* | *KSS* | *sign* | *Inf-t* |
| | | | | | | $T$ = 300 | | | | | | |
| Normal | 0.052 | 0.051 | 0.061 | 0.048 | 0.995 | 0.806 | 0.715 | 0.953 | 0.994 | 0.804 | 0.666 | 0.956 |
| Chi-square df = 1 | 0.047 | 0.053 | 0.254 | 0.058 | 0.995 | 0.764 | 0.538 | 0.957 | 0.995 | 0.751 | 0.054 | 0.942 |
| Chi-square df = 2 | 0.051 | 0.050 | 0.212 | 0.054 | 0.995 | 0.780 | 0.604 | 0.956 | 0.994 | 0.777 | 0.189 | 0.949 |
| Chi-square df = 3 | 0.049 | 0.050 | 0.170 | 0.050 | 0.994 | 0.796 | 0.637 | 0.954 | 0.995 | 0.795 | 0.307 | 0.955 |
| Chi-square df = 4 | 0.051 | 0.053 | 0.149 | 0.052 | 0.996 | 0.794 | 0.654 | 0.956 | 0.995 | 0.787 | 0.363 | 0.952 |
| Student t df = 2 | 0.056 | 0.056 | 0.059 | 0.090 | 0.995 | 0.625 | 0.976 | 0.961 | 0.993 | 0.581 | 0.976 | 0.817 |
| Student t df = 3 | 0.052 | 0.055 | 0.059 | 0.068 | 0.996 | 0.697 | 0.929 | 0.959 | 0.996 | 0.674 | 0.929 | 0.919 |
| Student t df = 4 | 0.052 | 0.054 | 0.058 | 0.061 | 0.996 | 0.741 | 0.876 | 0.956 | 0.996 | 0.727 | 0.876 | 0.935 |
| Student t df = 8 | 0.051 | 0.050 | 0.057 | 0.053 | 0.996 | 0.796 | 0.789 | 0.955 | 0.996 | 0.796 | 0.789 | 0.950 |
| Double exponential | 0.051 | 0.051 | 0.061 | 0.054 | 0.996 | 0.782 | 0.949 | 0.953 | 0.995 | 0.779 | 0.949 | 0.945 |
| Beta (2,2) | 0.050 | 0.047 | 0.060 | 0.045 | 0.995 | 0.824 | 0.478 | 0.953 | 0.995 | 0.832 | 0.478 | 0.960 |
| | | | | | | $T$ = 500 | | | | | | |
| Normal | 0.054 | 0.053 | 0.060 | 0.050 | 1.000 | 0.969 | 0.935 | 1.000 | 1.000 | 0.966 | 0.935 | 1.000 |
| Chi-square df = 1 | 0.048 | 0.053 | 0.317 | 0.053 | 1.000 | 0.939 | 0.746 | 1.000 | 1.000 | 0.936 | 0.031 | 1.000 |
| Chi-square df = 2 | 0.045 | 0.046 | 0.265 | 0.047 | 1.000 | 0.946 | 0.835 | 1.000 | 1.000 | 0.952 | 0.188 | 1.000 |
| Chi-square df = 3 | 0.048 | 0.049 | 0.229 | 0.049 | 1.000 | 0.953 | 0.872 | 1.000 | 1.000 | 0.954 | 0.349 | 1.000 |
| Chi-square df = 4 | 0.049 | 0.052 | 0.201 | 0.050 | 1.000 | 0.954 | 0.882 | 1.000 | 1.000 | 0.952 | 0.504 | 1.000 |
| Student t df = 2 | 0.055 | 0.054 | 0.056 | 0.087 | 1.000 | 0.939 | 0.999 | 0.999 | 0.999 | 0.928 | 0.999 | 0.998 |
| Student t df = 3 | 0.052 | 0.052 | 0.056 | 0.067 | 1.000 | 0.926 | 0.997 | 1.000 | 1.000 | 0.922 | 0.997 | 1.000 |
| Student t df = 4 | 0.053 | 0.055 | 0.057 | 0.058 | 1.000 | 0.940 | 0.991 | 1.000 | 1.000 | 0.934 | 0.991 | 1.000 |
| Student t df = 8 | 0.047 | 0.047 | 0.058 | 0.049 | 1.000 | 0.963 | 0.968 | 1.000 | 1.000 | 0.966 | 0.968 | 1.000 |
| Double exponential | 0.051 | 0.052 | 0.055 | 0.052 | 1.000 | 0.960 | 0.999 | 1.000 | 1.000 | 0.957 | 0.999 | 1.000 |
| Beta (2,2) | 0.046 | 0.047 | 0.056 | 0.043 | 1.000 | 0.974 | 0.723 | 1.000 | 1.000 | 0.975 | 0.723 | 1.000 |

Table 2.4 Size and Power of Various Tests in Large Samples
(Model with a Trend)

| Distribution | A. Size | | | | B. Power | | | | C. Size-adjusted Power | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *DF* | *KSS* | *sign* | *Inf-t* | *DF* | *KSS* | *sign* | *Inf-t* | *DF* | *KSS* | *sign* | *Inf-t* |
| | | | | | | $T$ = 300 | | | | | | |
| Normal | 0.051 | 0.041 | 0.053 | 0.047 | 0.948 | 0.624 | 0.484 | 0.847 | 0.947 | 0.667 | 0.484 | 0.857 |
| Chi-square df = 1 | 0.049 | 0.044 | 0.222 | 0.068 | 0.948 | 0.514 | 0.392 | 0.851 | 0.950 | 0.550 | 0.042 | 0.773 |
| Chi-square df = 2 | 0.049 | 0.041 | 0.162 | 0.059 | 0.949 | 0.566 | 0.424 | 0.845 | 0.950 | 0.609 | 0.129 | 0.812 |
| Chi-square df = 3 | 0.050 | 0.042 | 0.132 | 0.058 | 0.948 | 0.579 | 0.443 | 0.839 | 0.948 | 0.627 | 0.185 | 0.813 |
| Chi-square df = 4 | 0.050 | 0.038 | 0.116 | 0.056 | 0.945 | 0.588 | 0.444 | 0.841 | 0.945 | 0.641 | 0.265 | 0.814 |
| Student t df = 2 | 0.053 | 0.044 | 0.053 | 0.124 | 0.955 | 0.314 | 0.808 | 0.849 | 0.950 | 0.365 | 0.808 | 0.343 |
| Student t df = 3 | 0.050 | 0.040 | 0.052 | 0.081 | 0.951 | 0.448 | 0.710 | 0.838 | 0.950 | 0.500 | 0.710 | 0.684 |
| Student t df = 4 | 0.050 | 0.042 | 0.051 | 0.063 | 0.946 | 0.517 | 0.646 | 0.838 | 0.946 | 0.555 | 0.646 | 0.792 |
| Student t df = 8 | 0.051 | 0.042 | 0.052 | 0.052 | 0.948 | 0.605 | 0.557 | 0.847 | 0.945 | 0.639 | 0.557 | 0.838 |
| Double exponential | 0.050 | 0.040 | 0.051 | 0.053 | 0.949 | 0.573 | 0.757 | 0.843 | 0.949 | 0.622 | 0.757 | 0.827 |
| Beta (2,2) | 0.052 | 0.038 | 0.056 | 0.044 | 0.949 | 0.649 | 0.310 | 0.847 | 0.947 | 0.698 | 0.310 | 0.867 |
| | | | | | | $T$ = 500 | | | | | | |
| Normal | 0.049 | 0.040 | 0.056 | 0.042 | 1.000 | 0.915 | 0.813 | 0.999 | 1.000 | 0.932 | 0.813 | 0.999 |
| Chi-square df = 1 | 0.050 | 0.043 | 0.265 | 0.060 | 1.000 | 0.835 | 0.612 | 0.999 | 1.000 | 0.853 | 0.046 | 0.999 |
| Chi-square df = 2 | 0.051 | 0.043 | 0.206 | 0.054 | 1.000 | 0.863 | 0.703 | 0.999 | 1.000 | 0.880 | 0.182 | 0.999 |
| Chi-square df = 3 | 0.049 | 0.040 | 0.169 | 0.051 | 1.000 | 0.874 | 0.737 | 0.999 | 1.000 | 0.895 | 0.339 | 0.999 |
| Chi-square df = 4 | 0.047 | 0.039 | 0.141 | 0.050 | 1.000 | 0.883 | 0.757 | 0.999 | 1.000 | 0.903 | 0.468 | 0.999 |
| Student t df = 2 | 0.054 | 0.045 | 0.055 | 0.120 | 0.999 | 0.767 | 0.983 | 0.998 | 0.999 | 0.795 | 0.983 | 0.957 |
| Student t df = 3 | 0.048 | 0.037 | 0.050 | 0.074 | 1.000 | 0.792 | 0.966 | 0.999 | 1.000 | 0.833 | 0.972 | 0.995 |
| Student t df = 4 | 0.047 | 0.038 | 0.052 | 0.058 | 1.000 | 0.847 | 0.940 | 0.999 | 1.000 | 0.882 | 0.940 | 0.998 |
| Student t df = 8 | 0.049 | 0.037 | 0.050 | 0.044 | 1.000 | 0.899 | 0.881 | 1.000 | 1.000 | 0.921 | 0.881 | 1.000 |
| Double exponential | 0.050 | 0.041 | 0.051 | 0.051 | 1.000 | 0.883 | 0.976 | 0.999 | 1.000 | 0.906 | 0.976 | 0.999 |
| Beta (2,2) | 0.048 | 0.040 | 0.054 | 0.044 | 1.000 | 0.915 | 0.566 | 0.999 | 1.000 | 0.932 | 0.566 | 0.999 |

CHAPTER 3

# MORE POWERFUL COINTEGRATION TESTS WITH NON-NORMAL ERRORS

This essay proposes new cointegration tests that can be more powerful when utilizing the information of non-normal errors. The relatively low power of the usual cointegration tests is well recognized in the literature. Indeed, the task of seeking more powerful tests is not a trivial concern. In this essay, we demonstrate how existing cointegration tests can become more powerful by utilizing information that the errors are non-normal.

It seems clear that existing cointegration tests ignore information about non-normal errors. That is, the usual cointegration tests assume normal errors. For example, to derive critical values in the usual cointegration tests, the models utilize normal errors. This does not pose a problem even when the error terms are non-normal, since the limiting distribution of the test statistic is unaffected and the parameter estimates from the testing regression are still consistent. This outcome is due to the central limit theorem and the functional central limit theorem, and the same result still holds for the usual unit root tests and cointegration tests regardless of the distribution of the error term. However, we wish to note that this outcome does not necessarily mean that we should ignore information that the errors are non-normal. If we can possibly utilize this information in the cointegration test, the resulting estimator can be more efficient and the test will become more powerful.

The question is how to utilize information that the errors are non-normal. If we know the true distribution of the error term, it might be possible to employ the maximum likelihood estimators (MLE) using the known density function. However, we seldom know the true density function of a non-normal error, and it is extremely difficult to gauge the underlying distribution. If we use incorrect information on the error distribution or the functional form of the relationship, the consequence could be even worse.

The new cointegration tests we propose in this essay will utilize information on the non-normal distribution of the error term without pre-specifying a particular density function or a functional form. In doing this, we extend the work of Im, Lee and Tieslau (2011), who develop unit root tests with non-normal errors. The underlying idea is to adopt a two-step procedure following the “residual augmented least squares” (RALS) method of Im and Schmidt (2008), which can make use of nonlinear moment conditions driven by non-normal errors. If the errors are non-normal, the higher moments of the error term (or residuals) will contain information on the nature of the non-normal errors. If we can utilize this information in the higher moments of the residuals, then we can potentially obtain more powerful cointegration tests. The suggested RALS testing procedure is easy to implement because it does not require non-linear estimation techniques even though we utilize nonlinear moment conditions associated with the non-normal errors. We base our suggested tests on the usual least square estimation, and we adopt a linearized RALS procedure to utilize the nonlinear moment conditions from higher moments of the residuals. The procedure is justified by the GMM approach and the RALS estimator is as efficient as the GMM estimator.

The use of non-normal errors may have broader implications because they can reflect neglected nonlinearity of the relationships between variables of interest. Many macroeconomic models often entail nonlinear or asymmetric relationships among macro and financial time series variables, but, widely known evidence demonstrates that the usual linear tests will lose power in the presence of nonlinearity. Although several nonlinear unit root and cointegration tests have been suggested to detect nonlinearity, these tests can be vulnerable if the specified nonlinear form differs from the true nonlinear form. However, in terms of a modeling purpose, there have been some attempts to capture nonlinearity with non-normal errors. For example, so-called leptokurtic distributions are modeled as the *t*-distribution where the degree of freedom takes a low value close to 1, which is the case with the Cauchy distribution in the extreme. Still, developing cointegration tests using non-normal errors might have been difficult since the true density function is unknown, although there is a sense that non-normal errors can mimic various forms of nonlinearity. Furthermore, it seems important to note that as shown in Essay 3 of this dissertation, nonlinear unit root tests suffer from loss of power in the presence of non-normal errors. We expect that the same problem can be found in nonlinear cointegration tests.

The rest of this essay is organized as follows. In section 3.1, we formulate the single equation cointegration model from a Vector Auto Regressions (VAR) and Error Correction (ECM) model. In section 3.2, we explain the RALS cointegration testing procedures. Section 3.3 provides asymptotic results for the proposed tests. In addition, section 3.4 provides simulation results to see how the suggested tests perform under various scenarios. Concluding remarks are found in section 3.5.

## 3.1 Cointegration Models

In this essay, we consider three types of cointegration tests in the single equation model. Presumably, the tests using the full information maximum likelihood (FIML) estimation and system equations are more efficient if the system is correctly specified. However, there are certain cases in which the single equation model is good enough or preferred over the system equation model since the single equation model provides a parsimonious representation of the underlying model. [1] For example, one can correctly specify the Taylor rule model, the monetary model of real exchange rates, and the purchasing power parity model, among others. In this essay, we focus on single-equation based cointegration tests.[2]

To explain the motivation of the cointegration tests based on the single equation model, we first consider a VAR (*p*) process with an $n$-dimensional nonstationary variable $y_t$

$$\Phi(L)y_t = d_t + \epsilon_t \tag{3.1.1}$$

where $\epsilon_t \sim iid(0, \Omega)$, and $E(\epsilon_t \epsilon_\tau') = \begin{cases} \Omega & \text{for } t = \tau \\ 0 & otherwise \end{cases}$ .

Here $\Phi(L) = I_n - \Phi_1 L^1 - \Phi_2 L^2 - \cdots - \Phi_p L^p$, and $d_t$ denotes the deterministic terms. Note that normality of the error term is often assumed for convenience in the usual tests, but the asymptotic results do not hinge on this assumption in the linear models. If the linear combination of $y_t$ becomes I(0), one may consider the following vector error correction model (VECM):

[1] See Ericsson and MacKinnon (2002) for more details on the advantages and disadvantages of tests using a single equation versus a system of equations.
[2] It is possible to develop new cointegration tests with non-normal errors in the framework of a system of equations but this issue is relegated to future research.

$$\Delta y_t = d_t + \Gamma_0 y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \epsilon_t \tag{3.1.2}$$

where $\Gamma_0 = -\Phi(1) = \sum_{i=1}^{p} \Phi_i - I_n, \Gamma_i = -[\Phi_{s+1} + \Phi_{s+2} + \cdots + \Phi_p], i = 1, \cdots, p-1$, and

$$\Gamma_0 = BA' = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \begin{pmatrix} 1 & -\beta' \end{pmatrix}. \tag{3.1.3}$$

Suppose $y_t$ is decomposed into a scalar $y_{1t}$ and a vector $y_{2t}$ of dimension $n-1$. Thus, $b_1$ is a scalar and $b_2$ and $\beta$ are $(n-1) \times 1$ vectors. Now, we assume the vector $y_{2t}$ is weakly exogenous. From Harbo, Johansen, Nielsen, and Behbek (1998), the conditional model for $\Delta y_{1t}$ is given by

$$\Delta y_{1t} = d_{1t} + \delta_1 (y_{1,t-1} - \beta' y_{2,t-1}) + \phi' \Delta y_{2t} + e_t \tag{3.1.4}$$

where the lagged terms of $\Delta y_{1t}$ and $\Delta y_{2t}$ are omitted to simplify the notation. Assuming a prespecified cointegrating vector and weakly exogenous $y_{2t}$, $(\delta_1, \beta')$ can be efficiently estimated from the single conditional ECM (3.1.4). It is important to note that the contemporaneous terms of $\Delta y_{2t}$ appear in the above single equation model. When the cointegrating vector $\beta$ is not known, one can replace $y_{1,t-1} - \beta' y_{2,t-1}$ with $\hat{z}_{t-1}$, where $\hat{z}_{t-1} = y_{1,t-1} - \hat{\beta}' y_{2,t-1}$ is the estimated error correction term. Then, we have

$$\Delta y_{1t} = d_{1t} + \delta_1 \hat{z}_{t-1} + \phi' \Delta y_{2t} + e_t \tag{3.1.5}$$

For the RALS based ECM test, we use ECM model with a prespecified cointegrating vectror in (3.1.4). One can test the null hypothesis of no cointegration against the alternative hypothesis of cointegration

$$H_0 : \delta_1 = 0 \text{ against } H_1 : \delta_1 < 0. \tag{3.1.6}$$

The ECM cointegration test employs OLS to obtain the estimate of $\delta_1$ and to construct a t-statistic for $\delta_1 = 0$. One important issue is that the distribution of the ECM test depends on the nuisance parameter, since $\Delta y_{2t}$ in (3.1.4) works like a stationary covariate; see Zivot (2000). The limiting distribution depends on the relative contribution of the variance of $e_t$ to the variance of $\epsilon_t = \phi' \Delta y_{2t} + e_t$.

The conditional ECM in (3.1.4) can be expressed as a conditional autoregressive distributed lag (ADL) model

$$\Delta y_{1t} = d_{1t} + \delta_1 y_{1,t-1} + \gamma' y_{2,t-1} + \phi' \Delta y_{2t} + e_t \tag{3.1.7}$$

where we can relate (3.1.4) to (3.1.7) with $-\delta_1 \beta' = \gamma'$. The ADL test, which was initially suggested by Banerjee, Dolado, and Mestre (1998) is based on the *t*-test on $\delta_1 = 0$ from (3.1.7). Unlike the ECM based test, the stationary covariate term $\Delta y_{2t}$ does not yield a nuisance parameter in the ADL test. It is also possible to consider a F-test on the joint restriction, $\delta_1 = \gamma = 0$, but we focus on the *t*-test on $\delta_1 = 0$ in this essay.

The pioneering work of Engle and Granger (1987; EG) is based on a two-step procedure. In the first step, the residuals $\hat{u}_t$ are obtained from the long-run equilibrium regression of $y_{1t}$ on $y_{2t}$ along with a constant and/or a trend function. The EG cointegration test is based on the *t*-statistic on $\delta_1 = 0$ from

$$\Delta \hat{u}_t = \delta_1 \hat{u}_{t-1} + e_t \, . \tag{3.1.8}$$

To seek room for improving the power of the EG test, one can subtract $\Delta y_{2t}$ from the usual error correction model in (3.1.4) and re-arrange as follows:

$$\Delta(y_{1t} - \beta' y_{2t}) = d_{1t} + \delta_1 (y_{1,t-1} - \beta' y_{2,t-1}) + (\phi - \beta)' \Delta y_{2t} + e_t.$$

By relating the above equation to (3.1.8), we can observe that the signal-to-noise ratio term $(\phi - \beta)'\Delta y_{2t}$ is missing in (3.1.8). This does not pose a problem when $\phi = \beta$. Here, $\phi$ is the short-run parameter in the relationship between $\Delta y_{1t}$ and $\Delta y_{2t}$, while $\beta$ is the long-run parameter in the relationship between $y_{1t}$ and $y_{2t}$. Thus, the EG equation in (3.1.8) emerges only when $\phi = \beta$. In general, these coefficients $(\phi, \beta)$ are different from each other, and the EG tests tend to lose power under the alternative hypothesis as the difference $|\phi - \beta|$ becomes large, which is one source of the increase of the signal-to-noise ratio. In order to resolve this issue, we may consider the following regression for the extended EG test as considered in Essay 1.

$$\Delta \hat{u}_t = c_0 + \delta_1 \hat{u}_{t-1} + c'\Delta y_{2t} + e_t^* \qquad (3.1.9)$$

for which we relate the error term in (3.1.8) with $e_t = c'\Delta y_{2t} + e_t^*$. Then, we can still use the *t*-statistic on $\delta_1 = 0$ from (3.1.9). We call this test the modified EG cointegration test and denote as EG2. Note that the added term $\Delta y_{2t}$ will incur a nuisance parameter in the modified EG test under the null hypothesis, as in the ECM based test in (3.1.4). However, this nuisance parameter issue will be resolved using proper asymptotic results in the next section.

### 3.2. RALS Cointegration Tests

Next, we explain how to utilize the information on non-normality of the error term. We follow Im et al. (2011) to adopt the "residual augmented least squares" (RALS) method, which was initially suggested by Im and Schmidt (2008). To begin with, we let $x_t$ contain all regressors, excluding $\Delta y_{2t}$, in each of the testing regressions for the ECM test in (3.1.4), the ADL test in (3.1.7), the EG test in (3.1.8), and the EG2 test in (3.1.9). When the lagged variables of $\Delta y_{1t}$, $\Delta y_{2t}$ or $\Delta \hat{u}_t$ are included to capture autocorrelations in each of the testing regressions, $x_t$ will include these terms.

We let $e_t$ denote the error term for each of the usual testing regressions in (3.1.4), (3.1.7), (3.1.8), and (3.1.9). Following the RALS procedure, we consider the subsequent moment conditions

$$E[g(e_t) \otimes x_t] = 0, t = 1, 2, .., T \tag{3.2.1}$$

where $g(e_t)$ is a vector that satisfies the following assumption.

***Assumption* 1**. *$g(e_t)$ is differentiable and satisfies the first-order Lipshitz condition $\left|g_j'(x) - g_j'(y)\right| < M\,|x - y|$ for some constant $M$ for all $j$, where $g_j(\cdot)$ is the $j$-th element of $g(\cdot)$. Also, we have that $E\left[g(e_t)\right] = 0,$ the second moment of $g(e_t)$ exists, and $E\left[g'(e_t)\right] < \infty$.*

Then, we let $g(e_t) = (e_t, [h(e_t) - K]')'$, where $h(e_t)$ is a nonlinear function of the error term $e_t$, and $K = E(e_t)$. Thus, we have two different moment conditions

$$E(e_t \otimes x_t) = 0 \tag{3.2.2}$$

$$E[(h(e_t) - K) \otimes x_t] = 0. \tag{3.2.3}$$

The first is the usual moment condition from the ordinary least squares estimation and the second involves the moment conditions based on the nonlinear function of $e_t$. We define $C = E\left[g(e_t)g(e_t)'\right]$, $D = E\left[g'(e_t)\right]$, and $\psi(e_t) = D'C^{-1}g(e_t)$. Specifically, we have

$$C = \begin{bmatrix} \sigma_e^2 & C_{21}' \\ C_{21} & C_{22} \end{bmatrix}, \text{ and } D = \begin{bmatrix} 1 \\ D_2 \end{bmatrix} \tag{3.2.4}$$

where $C_{21} = E\left[e_t h(e_t)\right]$, $C_{22} = E\left[h(e_t)h(e_t)'\right]$, and $D_2 = E\left[h'\left(e_t\right)\right]$. Then, we can consider the following term

$$w_t = h(e_t) - K - e_t D_2\,. \tag{3.2.5}$$

The RALS procedure requires to augment $w_t$ to the relevant cointegration testing regressions. To implement the RALS procedure to test for cointegration, we consider $h(e_t)$ = $[e_t^2, e_t^3]'$. The motivation of the choice of this function is that higher moments of $e_t$ will contain information on non-normal errors of the error term. However, it is possible to consider more moment conditions. Since $e_t$ is not observable, we use the residuals of each of the cointegration testing regressions in (3.1.4), (3.1.7), (3.1.8), or (3.1.9) for different cointegration tests. We denote $\hat{e}_t$ as the residuals from each of the testing regression model and define

$$\hat{w}_t = h(\hat{e}_t) - \hat{K} - \hat{e}_t \hat{D}_2 \tag{3.2.6}$$

where $h(\hat{e}_t) = [\hat{e}_t^2, \hat{e}_t^3]'$, $\hat{K} = \frac{1}{T}\sum_{t=1}^{T} h(\hat{e}_t)$, and $\hat{D}_2 = \frac{1}{T}\sum_{t=1}^{T} h'(\hat{e}_t)$. Specifically,

$$\hat{w}_t = \left[\hat{e}_t^2 - m_2,\ \hat{e}_t^3 - m_3 - 3m_2\hat{e}_t\right]' \tag{3.2.7}$$

where $m_j = T^{-1}\sum_{t=1}^{T} \hat{e}_t^j$. The first term in $\hat{w}_t$ is associated with the moment condition $E[(e_t^2 - \sigma_e^2)x_t] = 0$, which implies no heteroskedasticity. In the absence of heteroskedasticity, this information does not induce efficiency gains. However, in the presence of heteroskedasticity, the estimator using this condition becomes more efficient. The second term in $\hat{w}_t$ is associated with the redundancy condition $\mu_{j+1} = j\sigma^2\mu_{j-1}$ with $\mu_j = E(e_t^j)$, which holds under the normal distribution. This condition is not useful only for the normal distribution. However, under a non-normal distribution, the estimator utilizing this condition becomes more efficient, and tests based on this information become more powerful.

The RALS cointegration testing regressions are given by augmenting $\hat{w}_t$ to each of the relevant cointegration testing regressions.

ECM: $$\Delta y_{1t} = d_{1t} + \delta_1 z_{t-1} + \phi'\Delta y_{2t} + \hat{w}_t'\gamma + \nu_t \tag{3.2.8}$$

ADL: $$\Delta y_{1t} = d_{1t} + \delta_1 y_{1,t-1} + \gamma' y_{2,t-1} + \phi'\Delta y_{2t} + \hat{w}_t'\gamma + \nu_t \tag{3.2.9}$$

EG: $$\Delta \hat{u}_t = d_{1t} + \delta_1 \hat{u}_{t-1} + \hat{w}_t' \gamma + \nu_t \tag{3.2.10}$$

EG2: $$\Delta \hat{u}_t = d_{1t} + \delta_1 \hat{u}_{t-1} + \phi' \Delta y_{2t} + \hat{w}_t' \gamma + \nu_t. \tag{3.2.11}$$

For the deterministic term $d_{1t}$, we consider a constant and a trend function. [3] For ease of exposition, all lag terms are suppressed.

### 3.3. Asymptotic Properties

For each of the RALS cointegration tests, we consider the usual *t*-statistic on $\delta_1 = 0$. We denote each test as $t_i^*$, *i* = ECM, ADL, EG, or EG2. In addition, we let $t_i$, *i* = ECM, ADL, EG, or EG2, denote the usual *t*-statistic from the usual (non-RALS) cointegration testing regression using (3.1.4), (3.1.7), (3.1.8), and (3.1.9), respectively. Then, it can be shown that the asymptotic distribution of the RALS cointegration test is given as follows:

**Theorem 1**. *Suppose that Assumption 1 holds, and we consider the usual t-statistic on* $\delta_1 = 0$ *in each of the testing regressions* (3.2.8), (3.2.9), (3.2.10), *and* (3.2.11). *Under the null hypothesis of no-cointegration, we have*

$$t_i^* \to \rho \, t_i + \sqrt{1-\rho^2} \, Z \tag{3.3.1}$$

*where i* = ECM, ADL, EG, or EG2; *Z follows the standard normal distribution; and* $\rho$ *is the long-run correlation between* $e_t$ *in* (3.1.4), (3.1.7), (3.1.8), (3.1.9), *and* $\nu_t$ *in* (3.2.8), (3.2.9), (3.2.10), *and* (3.2.11), *respectively*

$$\rho^2 = \frac{\sigma_{\nu e}^2}{\sigma_\nu^2 \sigma_e^2}. \tag{3.3.2}$$

PROOF:

---

[3] Note that both EG and EG2 tests include a constant term in their short run dynamics relationship. Also, a constant and/or a trend term is included in the long-run regression, $y_{1t} = c_0 + \beta y_{2t} + u_t$.

First, we consider the limiting distribution of RALS ADL test. To begin with, notice the relationship given as $e_t = \hat{w}_t'\gamma + \nu_t$ by relating RALS-ADL test ($t^*_{ADL}$) in (3.2.9) and the ADL test in (3.1.7). We let $Q$ denote a set of variables including stationary covariate terms which appear on the right hand side, $Q_t = (y_{2,t-1}', \Delta y_{2t}', \hat{w}_t')'$ and $M = I - Q(Q'Q)^{-1}Q'$. Here, the partitioned inverses is given as $M = M_1 - M_1 y_{1,t-1}(y_{1,t-1}' M_1 y_{1,t-1}) y_{1,t-1}' M_1$ where $M_1 = I - y_{1,t-1}(y_{1,t-1}' y_{1,t-1})^{-1} y_{1,t-1}'$. As suggested in Banerjee et al. (1998), we have for the OLS estimator of $\delta_1$

$$T(\hat{\delta}_1 - \delta_1) = \frac{\frac{1}{T} y_{1,t-1}\nu_t}{\frac{1}{T^2} y_{1,t-1} M y_{1,t-1}}. \tag{3.3.3}$$

Additionally, we adopt the following results:

$$T^{-2} y_{1,t-1}' M y_{1,t-1} = T^{-2} y_{1,t-1}' M_1 y_{1,t-1} + o_p(1) \Rightarrow a(1)^{-2}\sigma_e^2 \int_0^1 (w_1^c)^2 \tag{3.3.4}$$

$$\frac{1}{T}\sum_{t=2}^{T} y_{1,t-1}\nu_t \to a(1)^{-1}\sigma_e\sigma_\nu \left(\rho \int_0^1 w_1^c dw_1 + (1-\rho^2)^{(1/2)} \int_0^1 w_1^c dw_2\right) \tag{3.3.5}$$

where $\rho = \dfrac{\sigma_{v\epsilon}}{\sigma_v \sigma_\epsilon}$.

Here, (3.3.4) is the standard result for the squared sum of I(1) variable and (3.3.5) is given in Hansen (1995). We wish to note that the RALS term $\hat{w}_t'$ is a nonlinear function of stationary covariates. As such, including this term does not affect the above results. Then, we can show

$$T(\hat{\delta}_1 - \delta_1) = a(1)R\left(\rho \frac{\int_0^1 w_1^c dw_1}{\int_0^1 (w_1^c)^2} + (1-\rho^2)^{1/2} \frac{\int_0^1 w_1^c dw_2}{\int_0^1 (w_1^c)^2}\right),$$

where $R = \frac{\sigma_\nu}{\sigma_e}$. Furthermore, we have

$$t(\hat{\delta}_1) = \hat{\delta}_1 \cdot \sigma_\nu^{-1} \cdot T \left( \frac{1}{T^2} \sum_{t=2}^{T} y_{t-1}^2 - \frac{1}{T^2} \sum_{t=2}^{T} y_{t-1} Q_t' (\sum_{t=2}^{T} Q_t Q_t')^{-1} \sum_{t=2}^{T} Q_t y_{t-1} \right)^{\frac{1}{2}}$$

$$= T\hat{\delta}_1 \cdot \sigma_\nu^{-1} \left( \frac{1}{T^2} \sum_{t=2}^{T} y_{t-1}^2 - \frac{1}{T^2} \sum_{t=2}^{T} y_{t-1} \hat{w}_t' (\sum_{t=2}^{T} \hat{w}_t \hat{w}_t')^{-1} \sum_{t=2}^{T} \hat{w}_t y_{t-1} \right)^{\frac{1}{2}}$$

$$= \sigma_\nu^{-1} \left( \frac{1}{T^2} \sum_{t=2}^{T} y_{t-1}^2 \right)^{1/2} \cdot T\hat{\delta}_1 + o_p(1).$$

Using $T\delta = -ca(1)$, test statistics under the null of $\delta_1 = 0$ can be obtained as

$$t(\hat{\delta}_1) = \sigma_\nu^{-1} \left( a(1)^{-2} \sigma_\epsilon^2 \int_0^1 (w_1^c)^2 \right)^{\frac{1}{2}} \times \left[ -ca(1) + a(1) R \left( \rho \frac{\int_0^1 w_1^c dw_1}{\int_0^1 (w_1^c)^2} + (1-\rho^2)^{\frac{1}{2}} \frac{\int_0^1 w_1^c dw_2}{\int_0^1 (w_1^c)^2} \right) \right]$$

$$= -\frac{c}{R} \left( \int_0^1 (w_1^c)^2 \right)^{1/2} + \rho \frac{\int_0^1 w_1^c dw_1}{\left( \int_0^1 (w_1^c)^2 \right)^{1/2}} + (1-\rho^2) \frac{\int_0^1 w_1^c dw_2}{\left( \int_0^1 (w_1^c)^2 \right)^{1/2}}$$

where $w_2(r)$ is a standard Brownian motion independent of $w_1(r)$. Here, the null of no cointegration holds when $c = 0$. Also, we follow the result of Theorem 2 in Im, Lee and Tieslau (2010)), which showed that the GMM estimator based on the moment conditions in (3.2.2) and (3.2.3) has the same asymptotic distribution of the RALS based test. Moreover, using $w_t$ or $\hat{w}_t$ yields asymptotically the same estimator. Therefore, using the limiting distribution of the usual ADL test, as suggested in Banerjee (1998),

$$t(\hat{\delta}_{1,ADL}) = \frac{\int_0^1 w_1^c dw_1}{\left( \int_0^1 (w_1^c)^2 \right)^{(1/2)}}.$$

We have

$$t(\hat{\delta}_1) = \rho t_{\hat{\delta}_{1,ADL}} + \sqrt{(1-\rho^2)} N(0,1).$$

□

Second, we consider the modified RALS-EG test ($t^*_{EG2}$) for which the testing equation is given in (3.2.10).

A matrix $Q$ is similarly defined as in the above and includes all stationary covariate terms and their lagged values. By omitting the lagged terms, we have $Q_t = (\Delta y'_{2t}, \hat{w}'_t)'$. From the modified RALS-EG regression in (3.2.11) and the EG test in (3.1.8), we have the relationship of $e_t = \phi' \Delta y_{2t} + \hat{w}'_t \gamma + \nu_t$. Then, the OLS estimator for $\delta_1$ is characterized by

$$T(\hat{\delta}_1 - \delta_1) = \frac{\frac{1}{T}\sum_{t=2}^{T} \hat{u}_{t-1}\nu_t}{\frac{1}{T^2}\sum_{t=2}^{T} \hat{u}_{t-1}^2} + o_p(1) \tag{3.3.6}$$

Now, we adopt the following limiting distribution.

$$\frac{1}{T^2}\sum_{t=2}^{T} \hat{u}_{t-1}^2 \Rightarrow \xi' \int_0^1 (w_1^c)^2 dw_1 \xi \tag{3.3.7}$$

$$\frac{1}{T}\sum_{t=2}^{T} \hat{u}_{t-1}\nu_t \to a(1)^{-1}\sigma_\epsilon \sigma_\nu \xi' \left( \rho \int_0^1 w_1^c dw_1 + (1-\rho^2)^{(1/2)} \int_0^1 w_1^c dw_2 \right) \xi \tag{3.3.8}$$

with $\xi = \begin{bmatrix} 1 \\ -(\int_0^1 (w_2)^2 dr)^{-1} \int_0^1 w_2 w_1 dr \end{bmatrix}$.

Plugging in (3.3.7) and (3.3.8) to (3.3.6) gives

$$T(\hat{\delta}_1 - \delta_1) = a(1)R \left( \rho \frac{\xi' \int_0^1 w_1^c dw_1 \xi}{\xi' \int_0^1 (w_1^c)^2 \xi} + (1-\rho^2)^{(1/2)} \frac{\xi' \int_0^1 w_1^c dw_2 \xi}{\xi' \int_0^1 (w_1^c)^2 \xi} \right)$$

where $R = \frac{\sigma_\nu}{\sigma_e}$. Using the similar approach represented in the ADL model, the test statistics under the null of $\delta_1 = 0$ can be obtained as:

$$t(\hat{\delta}_1) = -\frac{c}{R}\left( \int_0^1 (w_1^c)^2 \right)^{1/2} + \rho \frac{\xi' \int_0^1 w_1^c dw_1 \xi}{\left( \xi' \int_0^1 (w_1^c)^2 \xi \right)^{1/2}} + (1-\rho^2) \frac{\int_0^1 w_1^c dw_2}{\left( \int_0^1 (w_1^c)^2 \right)^{1/2}}$$

Notice that with the relationship of $e_t = \phi' \Delta y_{2t} + \hat{w}_t' \gamma + \nu_t$, we are comparing the standard EG tests with our RALS based EG2 test. Since the limiting distribution of the usual EG test under the null of no cointegration is given by

$$t(\hat{\delta}_{1,EG}) = \frac{\xi' \int_0^1 w_1^c dw_1 \xi}{\left(\xi' \int_0^1 (w_1^c)^2 \xi\right)^{1/2}},$$

we have

$$t(\hat{\delta}_1) = \rho t_{\hat{\delta}_1, EG2} + \sqrt{(1-\rho^2)} N(0,1).$$

The procedure driving the limiting distribution for RALS EG tests is similar to that of the RALS EG2 tests, except for $Q_t$ includes only one stationary covariate term, $Q_t = (\hat{w}_t')'$. Therefore, we omit it here for simplicity.

□

Third, we examine the RALS-ECM test ($t^*_{ECM}$). The testing regression is given by (3.2.8). Following Zivot (2000), we express the conditional error correction model alternatively as

$$a(L)\Delta\alpha' y_t = \delta_1 \alpha' y_{t-1} + b(L)' \Delta y_{2t} + \hat{w}_t' \gamma + \nu_t.$$

$$e_t = \phi' \Delta y_{2t} + \hat{w}_t' \gamma + \nu_t.$$

where $y_t = y_{1t} - \beta y_{2t}$, $a(L) = 1 - C_{11}(L)L$, and $b(L) = (\phi' - \beta') + [C_{12}(L) + C_{11}(L)\beta]L$. The OLS estimator of $\delta_1$ is characterized by

$$T(\hat{\delta}_1 - \delta_1) = \frac{\frac{1}{T}\sum_{t=2}^T y_{t-1}\nu_t - \sum_{t=2}^T y_{t-1} Q_t' (\frac{1}{T}\sum_{t=2}^T Q_t Q_t')^{-1} \frac{1}{T}\sum_{t=2}^T Q_t \nu_t}{\frac{1}{T^2}\sum_{t=2}^T y_{t-1}^2 - \frac{1}{T^2}\sum_{t=2}^T y_{t-1} Q_t' (\frac{1}{T}\sum_{t=2}^T Q_t Q_t')^{-1} \frac{1}{T}\sum_{t=2}^T Q_t y_{t-1}}.$$

We let $Q_t = (\tilde{C}_t', \hat{w}_t{}')'$ with $C_t = (\triangle y_{1,t-1}, \cdots, \triangle y_{1,t-m}, \triangle y_{2,t-1}, \cdots, \triangle y_{2,t-p})$, and $\tilde{C}_t = C_t - T^{-1}\sum_{t=2}^T C_t$. Since $\frac{1}{T}\sum_{t=2}^T \hat{w}_t C_t' = o_p(1)$, and $\frac{1}{T}\sum_{t=2}^T \tilde{C}_t \nu_t = o_p(1)$, we have

$$T(\hat{\delta}_1 - \delta_1) = \frac{\frac{1}{T}\sum_{t=2}^T y_{t-1}\nu_t - \sum_{t=2}^T y_{t-1}\hat{w}_t(\frac{1}{T}\sum_{t=2}^T \hat{w}_t\hat{w}_t')^{-1}\frac{1}{T}\sum_{t=2}^T \hat{w}_t\nu_t}{\frac{1}{T^2}\sum_{t=2}^T y_{t-1}^2}.$$

Moreover, we have

$$T(\hat{\delta}_1 - \delta_1) = \frac{\frac{1}{T}\sum_{t=2}^T y_{t-1}\nu_t}{\frac{1}{T^2}\sum_{t=2}^T y_{t-1}^2} + o_p(1)$$

because $\frac{1}{T}\sum_{t=2}^T \hat{w}_t\nu_t = o_p(1)$ and $\frac{1}{T}\sum_{t=2}^T y_{t-1}\hat{w}_t = O_p(1)$.

Therefore by applying the lemma in Hansen (1995), we have

$$T(\hat{\delta}_1 - \delta_1) = a(1)R\left(\rho\frac{\int_0^1 w_1^c dw_1}{\int_0^1 (w_1^c)^2} + (1-\rho^2)^{(1/2)}\frac{\int_0^1 w_1^c dw_2}{\int_0^1 (w_1^c)^2}\right),$$

where $R = \frac{\sigma_\nu}{\sigma_e}$. Additionally, we have

$$t(\hat{\delta}_1) = \hat{\delta}_1 \cdot \sigma_\nu^{-1} \cdot T\left(\frac{1}{T^2}\sum_{t=2}^T y_{t-1}^2 - \frac{1}{T^2}\sum_{t=2}^T y_{t-1}Q_t'(\sum_{t=2}^T Q_tQ_t')^{-1}\sum_{t=2}^T Q_t y_{t-1}\right)^{\frac{1}{2}}$$

$$= T\hat{\delta}_1 \cdot \sigma_\nu^{-1}\left(\frac{1}{T^2}\sum_{t=2}^T y_{t-1}^2 - \frac{1}{T^2}\sum_{t=2}^T y_{t-1}\hat{w}_t'(\sum_{t=2}^T \hat{w}_t\hat{w}_t')^{-1}\sum_{t=2}^T \hat{w}_t y_{t-1}\right)^{\frac{1}{2}}$$

$$= \sigma_\nu^{-1}\left(\frac{1}{T^2}\sum_{t=2}^T y_{t-1}^2\right)^{1/2} \cdot T\hat{\delta}_1 + o_p(1),$$

and $T\delta = -ca(1)$. Then, the test statistic under the null of $\delta_1 = 0$ can be obtained as

$$t(\hat{\delta}_1) = \sigma_\nu^{-1}\left(a(1)^{-2}\sigma_e^2\int_0^1 (w_1^c)^2\right)^{\frac{1}{2}} \times \left[-ca(1) + a(1)R\left(\rho\frac{\int_0^1 w_1^c dw_1}{\int_0^1 (w_1^c)^2} + (1-\rho^2)^{\frac{1}{2}}\frac{\int_0^1 w_1^c dw_2}{\int_0^1 (w_1^c)^2}\right)\right]$$

$$= -\frac{c}{R}\left(\int_0^1 (w_1^c)^2\right)^{1/2} + \rho\frac{\int_0^1 w_1^c dw_1}{\left(\int_0^1 (w_1^c)^2\right)^{1/2}} + (1-\rho^2)\frac{\int_0^1 w_1^c dw_2}{\left(\int_0^1 (w_1^c)^2\right)^{1/2}}$$

The null of cointegration holds when $c = 0$. Therefore, we have

$$t(\hat{\delta}_1) = \rho t_{\hat{\delta}_1, ECM} + \sqrt{(1-\rho^2)}N(0,1).$$

□

The above results show that the asymptotic distributions of the RALS cointegration tests depend on the nuisance parameter $\rho^2$. However, it is feasible to compute this parameter, as in Hansen (1995) who suggested using a nonparametric estimation. For this, we let

$$\hat{\Omega} = \begin{pmatrix} \hat{\sigma}_e^2 & \hat{\sigma}_{e\nu} \\ \hat{\sigma}_{\nu e} & \hat{\sigma}_\nu^2 \end{pmatrix} = \sum_{k=-\xi}^{\xi} W\left(\tfrac{k}{\xi}\right)\frac{1}{T}\sum_{t=1}^{T}\hat{V}_t\hat{V}_{t-k}' \tag{3.3.9}$$

where $W(\cdot)$ is a Bartlett or Parzen kernel weight function, $\xi$ is a bandwidth parameter and $\hat{V}_t = (\hat{\nu}_t, \hat{e}_t)'$ is constructed from the residuals of the usual and RALS cointegration models, respectively. Then, we can estimate $\rho^2$ as

$$\hat{\rho}^2 = \frac{\hat{\sigma}_{\nu e}^2}{\hat{\sigma}_\nu^2\hat{\sigma}_e^2} \tag{3.3.10}$$

Im et al. (2011) consider RALS unit root tests, as an extension of the usual Dickey-Fuller tests. Their case is a special case of our models. Since the RALS based tests utilize the moment conditions, it will be easy to show that they have the same asymptotic distributions as the GMM tests. In light of the GMM estimator, it is easily expected that estimators (tests) using more moment conditions (3.2.3) will be more efficient (powerful) than those using fewer moment conditions.

In Tables 3.1 and 3.2, we provide the critical values of the RALS cointegration tests $t_{ADL}^*$ and $t_{EG2}^*$ for $T$ = 100 for different values of $\rho^2$ = 0.1, 0.2, $\cdots$, 0.9, and 1.0, when the number of variables ($n$) in the cointegration regression is given as $n$ = 2, 3, 4, 5 and 6.[4] Thus, $n$-1 is the number of integrated regressors, $y_{2t}$. Table 3.1 provides the critical values for the model with a constant and Table 3.2 provides the critical values for the model with a trend in a similar manner. In the Appendix Tables, we report the corresponding asymptotic critical values with $T$ = 500. We have obtained these critical values from the empirical distribution of each of the tests using 50,000 replications from Gauss 9.1 random number generators. It can also be shown that the ECM test of Zivot (2000) based on (3.1.4) has the same limiting distribution in (3.3.1), except that the parameter $\rho^2$ is estimated differently since we have the additional term $\hat{w}_t$ in the RALS testing regressions. However, as noted previously, neither Zivot (2000) nor other popular cointegration tests utilize information on non-normal error distributions.

Next, we provide guidance for how to apply the RALS cointegrating tests in practice. For example, one may wish to use the ADL version of the RALS cointegration test ($t^*_{ADL}$) as follows.

*Step* 1: Obtain the residuals $\hat{e}_t$ from the usual ADL testing equation given by (3.1.7) in order to create the terms, $\hat{w}_{2t}$ and $\hat{w}_{3t}$

$$\hat{w}_t = [\hat{w}_{2t}, \hat{w}_{3t}]' = \left[\hat{e}_t^2 - m_2, \hat{e}_t^3 - m_3 - 3m_2\hat{e}_t\right]'$$

where $m_j = T^{-1}\sum_{t=1}^{T}\hat{e}_t^j$. Then, use these terms to run the RALS-ADL testing regression in (3.2.9).

[4] The critical values for $t^*_{ECM}$ and $t^*_{EG}$ are omitted here. It will be shown that $t^*_{ADL}$ performs better than $t^*_{ECM}$, which shows size distortions under the null. $t^*_{EG}$ is not recommended since it loses power under the alternative as the signal-to-noise ratio increases.

*Step* 2: Calculate $\hat{\rho}^2$ as defined in (3.3.10). More specifically, in order to estimate this nuisance parameter, we use the nonparametric estimation of the long-run variance as in (3.3.9). This estimator utilizes two residuals $(\hat{\nu}_t, \hat{e}_t)'$ from (3.2.9) and (3.1.7), respectively.

*Step* 3: We now use the critical values of $t^*_{ADL}$. The exact critical values will be computed via interpolation using the nearest values of $\hat{\rho}^2$.

Here, the usual residuals $\hat{e}_t$ need to be obtained differently for $t^*_{EG2}$ and $t^*_{ECM}$. The residuals $\hat{e}_t$ for these tests can be obtained from the testing regression that excludes the stationary covariates $\Delta y_{2t}$. Thus, $\hat{e}_t$ can be obtained from (3.1.8) for $t^*_{EG2}$. Similarly, $\hat{e}_t$ can be obtained for $t^*_{ECM}$ from the regression, $\Delta y_{1t} = d_{1t} + \delta_1 z_{1,t-1} + e_t$.

### 3.4. Monte Carlo Experiment

In this section, we examine the finite sample properties of four different RALS cointegration tests ($t^*_{ECM}$, $t^*_{ADL}$, $t^*_{EG}$ and $t^*_{EG2}$) via Monte Carlo simulations. We also compare their performance with the corresponding existing cointegration tests ($t_{ECM}$, $t_{ADL}$, $t_{EG}$ and $t_{EG2}$) which do not utilize information on non-normality of the error term. We consider the following data generating process (DGP) to conduct experiments from various perspectives.

$$\Delta y_{1t} = \delta_1(y_{1,t-1} - \beta y_{2,t-1}) + \phi_1 \Delta y_{2t} + \epsilon_{1t} \tag{3.4.1}$$

$$\Delta y_{2t} = \delta_2(y_{1,t-1} - \beta y_{2,t-1}) + \phi_2 \Delta y_{1t} + \psi \Delta y_{2,t-1} + \epsilon_{2t} \tag{3.4.2}$$

$$\Omega = E(\epsilon_t' \epsilon_t) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

where $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t})$, $\sigma_1^2 = \sigma^2_{\Delta y_{1t}}$, and $\sigma_2^2 = \sigma^2_{\Delta y_{2t}}$. In the DGP we assume a pre-specified cointegrating vector $\beta = 1$. In addition, the correlation structure between two errors is denoted by $\sigma_{12} = \theta$. The first equation (3.4.1) can be considered the conditional expectation of $\Delta y_{1t}$ on

$\Delta y_{2t}$ and information set $I_{t-1}$, which is obtained from the partitioned vector error correction model with respect to $y_{1t}$ and $y_{2t}$. Similarly, the second equation (3.4.2) is the conditional expectation of $\Delta y_{2t}$ on $\Delta y_{1t}$ and $I_{t-1}$. In order to examine the performance of single equation based cointegration tests, we assume weak exogeneity of $y_{2t}$, and set $\delta_2 = 0$ and $\phi_2 = 0$. Here are several features that we will examine throughout the above DGP.

(i) We consider the case where the error processes $\epsilon_{1t}$ and $\epsilon_{2t}$ are possibly correlated with the correlation coefficient, $\theta = \{0.0, 0.3, 0.7\}$.

(ii) We also consider the case where $\Delta y_{2t}$ is persistent and examine how the persistence of $\Delta y_{2t}$ affects the performance of the proposed test. As such, we allow $\Delta y_{2t}$ to have an AR(1) process with the persistence parameter $\psi = \{0.0, 0.6, 0.9\}$.

(iii) We let $\sigma_1^2 = 1$, but we let the ratio of the two variances of $\Delta y_{1t}$ and $\Delta y_{2t}$ vary along with $\sigma_2^2 = \{1, 6, 16\}$. As we will discuss in greater detail later, size distortion in the ECM model depends upon with the relative variance ratio between two errors.

(iv) We examine the effects of another source of the signal-to-noise ratio. As noted previously, we relax the assumption implied in the usual EG test that the long-run ($\beta$) coefficient in the relationship between $y_{1t}$ and $y_{2t}$ is the same as the short-run ($\phi_1$) coefficient in the relationship between $\Delta y_{1t}$ and $\Delta y_{2t}$. As we let $\beta = 1$, this implied assumption does not hold when $\phi_1 \neq 1$. We consider two cases of $\phi_1 = \{1.0, 0.5\}$: (i) the long-run parameter is the same as the short-run parameter such that $\phi_1 = 1.0 = \beta$, and (ii) the long-run parameter is different from the short-run parameter such that $\phi_1 = 0.5 \neq \beta$. The cases for $\phi_1 = 0$ or 0.3 have been also considered, but their results (not reported) show similar patterns of the case where $\phi_1 = 0.5$.

(v) In order to demonstrate change in the size and power properties under non normal errors, we allow the error term in (3.4.1) to follow various types of distribution:

- standard normal distribution
- $\chi^2$ distributions with df = 1,2,3,4
- *t*-distributions with df = 2,3,4

We focus on examining the performance of the proposed tests with a constant using the small sample size of $T$ = 100, and report the results in Tables 3.3 to 3.8. To save space, we report only the results using three distributions: normal, $\chi^2$ distribution with df = 2, and *t*-distribution with df = 2, but similar results are obtained for the cases with other non-normal distributions.[5] The results with larger samples ($T$ = 500) are more pronounced and confirm our results more convincingly. We also obtained similar results for the model with a trend.[6] The rejection frequencies of a one-side test at the 5% significance level are calculated. All of the results are produced with 20,000 replications.

*Size Properties*

We start our discussion of the size properties by assuming the null hypothesis of $\delta_1 = 0$ from the DGP (3.4.1). The 5% rejection rates (sizes) for the baseline case are reported in Table 3.3 when $T$ = 100. Here, the contemporaneous term $\Delta y_{2t}$ is a white noise with $\psi = 0$, and the errors are not correlated so that $\theta = 0$ is assumed. Then, we focus on the pure effect of different values of $\phi_1$ and $\sigma_2^2$. We first notice that three types of the existing tests ($t_{ADL}$, $t_{EG}$, and $t_{EG2}$) show correct sizes close to 5% regardless of the distribution of the error term, either normal or

[5] In the Appendix Tables 3 - 5, we report the additional results using other non-normal distributions in the above DGP. To save space, we provide these results only in the baseline case.
[6] These results are omitted to save space.

various cases of non-normal distributions. We observe minor size distortions from the corresponding RALS cointegration tests ($t^*_{ADL}$, $t^*_{EG}$ and $t^*_{EG2}$), but additional simulation results (not shown here) show that the minor size distortion for $t^*_{ADL}$, $t^*_{EG}$ and $t^*_{EG2}$ decreases significantly in larger samples with *T* = 300. Overall, the size properties of these tests are good. This outcome confirms that the linear cointegration tests including the RALS cointegration tests will remain robust under non-normal errors. Moreover, they are not affected by different values of $\phi_1$ and $\sigma_2^2$ in the DGP. However, the ECM based tests show different behaviors. We observe non-negligible size distortions for the ECM tests for both $t_{ECM}$ and $t^*_{ECM}$. The size distortion problem of $t_{ECM}$ is more pronounced as the signal-to-noise ratio increases when either $\phi_1$ = 0.5 or $\sigma_2^2$ = 6 or 16. Moreover, we observe size distortions for the RALS based ECM test ($t^*_{ECM}$) even when $\phi_1$ = 1 and $\sigma_2^2$ = 1, and the error has the normal distribution. The dependency of the usual ECM test on the signal-to-noise ratio is already known in the literature (see Kremer et al., 1992). The same problem occurs again in the RALS version of the ECM test ($t^*_{ECM}$). As such, we do not recommend $t^*_{ECM}$ for empirical research. For this reason, we examine the size and power results for only the ADL, EG, and EG2 type tests in further discussion.

Now, we examine the case where $\Delta y_{2t}$ is not a white noise, but is modeled as a persistent series with $\Delta y_{2t} = \psi \Delta y_{2,t-1} + \epsilon_{2t}$ while assuming no correlation between errors ($\theta = 0$). The results in Table 3.4 show that the sizes of both tests are not affected much regardless of different error distributions, although there is a tendency of slight size distortions when the persistence parameter $\psi$ gets large. However, all these values are generally acceptable. The next issue in examining the size property is the case where the error terms of $\Delta y_{1t}$ and $\Delta y_{2t}$ are correlated such that $\theta \neq 0$. From Table 3.5, we first note that the correlated errors do not affect the size properties when the contemporaneous term is a white noise $(\psi = 0)$; the results with $\theta = \{0.3, 0.5\}$ do not change much and the results are similar to those in Table 3.4 when $\theta = 0$. Thus, the net effect of correlated errors is not significant when $\psi = 0$. The size properties are still as good as in the baseline case, regardless of different values of $\sigma_2^2$. However, we observe minor size distortions for $t_{ADL}$, $t_{EG2}$, $t^*_{ADL}$, and $t^*_{EG2}$ in the presence of combined effects where $\Delta y_{2t}$ is persistent $(\psi = 0.5)$ and the errors are strongly correlated (for example, $\theta = 0.5$). Despite this, it is encouraging that the size distortions tend to be reduced significantly for these tests, when the ratio of the error variance is large (for example, $\sigma_2^2$ = 16).

*Power Properties*

Our main interest lies in examining the power property. We start by examining the baseline case where $\Delta y_{2t}$ is a white noise ($\psi = 0$), and the errors are not correlated ($\theta = 0$). The size adjusted powers of various tests are presented for the case with *T* = 100 for all results. As shown in Table 3.6, it is clear that the power of the RALS based tests increases drastically in all cases when the error term follows any type of non-normal distribution. This result is one of the key findings of this essay. We observe clearly that RALS cointegration tests ($t^*_{ADL}$, $t^*_{EG}$, and $t^*_{EG2}$) are more powerful than the corresponding existing tests ($t_{ADL}$, $t_{EG}$, and $t_{EG2}$). The increase of the power is noticeable when the error term follows two types of non-normal distributions ($\chi^2$ distribution with df = 2, and *t*-distribution with df = 2). The results using other non-normal distributions (reported in the Appendix) are similar. Clearly, there is an operating advantage of using the RALS cointegration tests in the presence of non-normal errors. When the long-run coefficient is set differently from the short-run coefficient by using the value $\phi_1$ = 0.5 in the DGP, the power of all of the ADL and EG2 type tests ($t_{ADL}$, $t_{EG2}$, $t^*_{ADL}$, and $t^*_{EG2}$) also increases, as the signal-to-noise ratio increases. This pattern is not shown when the long-run coefficient is the same as the short-run coefficient ($\phi_1$ = 1.0). This outcome is as expected. However, we observe that the usual EG test ($t_{EG}$) loses power when $\phi_1 = 0.5$. Under normal errors, the power of the RALS cointegration tests is not necessarily better than that of the existing tests. This result is also as expected.

Table 3.7 reports the size adjusted power of the tests when $\Delta y_{2t}$ is persistent ($\psi \neq 0$), but the errors are not correlated ($\theta = 0$). In essence, the major findings are similar to those in Table 3.6, regarding the effects of various types of non-normal errors and different values of $\sigma_2^2$ and $\phi_1$. In particular, the increase of the power of the RALS based tests is again clear, and the effect is larger when $\phi_1$ = 0.5. The point of interest in Table 3.7 is the effect on the power when the persistent measure $\psi$ changes. We observe that the power of ADL and EG2 type tests is slightly lower in the case where $\psi$ = 0.9 than in the case where $\psi$ = 0.6. The decrease in the power is not highly significant. However, it seems difficult to find any certain pattern in the power property. In contrast to this, without the restriction on the two coefficients, increased persistency helps to improve the power of $t_{ADL}$ and $t^*_{ADL}$. However, there seems to be no clear pattern in $t_{EG}$ and $t^*_{EG2}$, except that these tests tend to lose power as $\sigma_2^2$ increases. The same patterns hold for other results using different magnitudes of the persistency measure and different variance ratios.

In Table 3.8, we examine the power of the tests when the errors are correlated. We are interested in investigating the power in the presence of combined effects of $\psi$ and $\theta$. To begin with, we examine the net effect of correlated errors when $\psi$ = 0. When $\phi_1 = 1$, the results in Table 3.8 show that the size adjusted powers of $t_{ADL}$, $t^*_{ADL}$, $t_{EG2}$, and $t^*_{EG2}$ tend to increase a little when the correlation coefficient ($\theta$) gets large, but the power gain is not significant. When

$\phi_1 = 0.5$, the size adjusted powers of $t_{ADL}$, $t^*_{ADL}$, $t_{EG2}$, and $t^*_{EG2}$ act reversely to the case of $\phi_1 = 1$; they go down. However, when $\psi$ and $\theta$ increase at the same time, power for all tests increases regardless of the restriction on the long run and short run coefficients. At any rate, the power gain of the RALS based tests over the existing tests is clear even in the presence of the combined effects of the persistence measure and correlated errors.

In summary, increasing persistency and covariance changes size and power property, but the changes are not very significant. Under all situations we considered, the major results hold. Sizes are generally good and the size adjusted power is higher in the RALS based tests than in the usual tests when the errors follow non-normal errors. In these cases, we do not require any other outside stationary covariates, which are normally difficult to find. These desirable results are not affected significantly by either correlated errors or the persistence of $\Delta y_{2t}$, although their combined effects can be harmful in some cases. This answers the main question of interest: we can improve the power of the existing tests without specific information on the type of non-normal errors.

## 3.5 Concluding Remarks

In this essay, we have proposed new cointegration tests that utilize information about non-normal errors that has been ignored in the literature. The RALS based cointegration tests that we propose make use of nonlinear moment conditions driven by non-normal errors, but we can still use the usual least squares estimation methods. Clearly, there are operating advantages for our newly suggested tests. It is encouraging that our RALS cointegration tests are more powerful than the usual cointegration tests that do not utilize information on non-normal errors.

The gain in power depends on the nature of the non-normal error distributions, but the power gain is substantial in many cases.

Table 3.1 Critical values of RALS cointegration tests (T = 100)
(model with a constant)

| $\rho^2$ | | 0.1 | | | 0.2 | | | 0.3 | | | 0.4 | | | 0.5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sig. level | | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| *n* = 2 | ADL | -2.956 | -2.263 | -1.902 | -3.212 | -2.539 | -2.176 | -3.397 | -2.718 | -2.369 | -3.550 | -2.881 | -2.535 | -3.692 | -3.026 | -2.669 |
| | EG2 | -2.931 | -2.267 | -1.910 | -3.181 | -2.517 | -2.163 | -3.341 | -2.695 | -2.354 | -3.489 | -2.847 | -2.499 | -3.596 | -2.976 | -2.638 |
| *n* = 3 | ADL | -3.058 | -2.366 | -1.992 | -3.269 | -2.628 | -2.272 | -3.498 | -2.839 | -2.484 | -3.677 | -3.013 | -2.667 | -3.841 | -3.177 | -2.817 |
| | EG2 | -3.099 | -2.419 | -2.060 | -3.364 | -2.702 | -2.349 | -3.573 | -2.910 | -2.564 | -3.732 | -3.099 | -2.754 | -3.880 | -3.257 | -2.919 |
| *n* = 4 | ADL | -3.097 | -2.417 | -2.044 | -3.415 | -2.728 | -2.362 | -3.658 | -2.970 | -2.601 | -3.835 | -3.177 | -2.809 | -4.038 | -3.338 | -2.979 |
| | EG2 | -3.188 | -2.527 | -2.164 | -3.534 | -2.865 | -2.507 | -3.780 | -3.133 | -2.785 | -3.968 | -3.344 | -3.010 | -4.132 | -3.510 | -3.177 |
| *n* = 5 | ADL | -3.113 | -2.452 | -2.100 | -3.479 | -2.802 | -2.445 | -3.755 | -3.064 | -2.702 | -3.964 | -3.282 | -2.904 | -4.143 | -3.463 | -3.100 |
| | EG2 | -3.280 | -2.641 | -2.276 | -3.673 | -3.010 | -2.660 | -3.979 | -3.313 | -2.951 | -4.198 | -3.545 | -3.214 | -4.401 | -3.747 | -3.413 |
| *n* = 6 | ADL | -2.900 | -2.198 | -1.829 | -3.102 | -2.417 | -2.062 | -3.247 | -2.599 | -2.233 | -3.357 | -2.708 | -2.361 | -3.515 | -2.834 | -2.486 |
| | EG2 | -2.837 | -2.160 | -1.790 | -2.998 | -2.327 | -1.976 | -3.124 | -2.464 | -2.122 | -3.235 | -2.588 | -2.242 | -3.297 | -2.666 | -2.335 |
| $\rho^2$ | | 0.6 | | | 0.7 | | | 0.8 | | | 0.9 | | | 1 | | |
| Sig. level | | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| *n* = 2 | ADL | -3.803 | -3.153 | -2.800 | -3.911 | -3.258 | -2.918 | -4.021 | -3.359 | -3.009 | -4.107 | -3.441 | -3.112 | -4.198 | -3.553 | -3.211 |
| | EG2 | -3.698 | -3.082 | -2.760 | -3.760 | -3.170 | -2.843 | -3.873 | -3.245 | -2.939 | -3.935 | -3.332 | -3.015 | -4.003 | -3.420 | -3.107 |
| *n* = 3 | ADL | -3.985 | -3.330 | -2.977 | -4.111 | -3.456 | -3.110 | -4.256 | -3.577 | -3.230 | -4.379 | -3.676 | -3.332 | -4.483 | -3.810 | -3.451 |
| | EG2 | -4.019 | -3.391 | -3.067 | -4.143 | -3.517 | -3.198 | -4.212 | -3.620 | -3.300 | -4.332 | -3.730 | -3.417 | -4.440 | -3.823 | -3.515 |
| *n* = 4 | ADL | -4.158 | -3.496 | -3.142 | -4.326 | -3.640 | -3.274 | -4.445 | -3.774 | -3.415 | -4.567 | -3.897 | -3.539 | -4.666 | -4.029 | -3.677 |
| | EG2 | -4.290 | -3.674 | -3.348 | -4.439 | -3.822 | -3.497 | -4.562 | -3.957 | -3.639 | -4.682 | -4.082 | -3.762 | -4.833 | -4.206 | -3.894 |
| *n* = 5 | ADL | -4.351 | -3.647 | -3.280 | -4.492 | -3.814 | -3.440 | -4.634 | -3.974 | -3.600 | -4.780 | -4.089 | -3.723 | -4.927 | -4.209 | -3.853 |
| | EG2 | -4.597 | -3.956 | -3.618 | -4.742 | -4.104 | -3.781 | -4.881 | -4.261 | -3.942 | -5.005 | -4.417 | -4.101 | -5.151 | -4.546 | -4.216 |
| *n* = 6 | ADL | -3.551 | -2.926 | -2.588 | -3.627 | -3.010 | -2.677 | -3.739 | -3.101 | -2.763 | -3.818 | -3.177 | -2.848 | -3.864 | -3.236 | -2.921 |
| | EG2 | -3.382 | -2.745 | -2.425 | -3.414 | -2.808 | -2.493 | -3.463 | -2.879 | -2.571 | -3.556 | -2.935 | -2.629 | -3.601 | -2.990 | -2.680 |

Table 3.2 Critical values of RALS cointegration tests (T = 100)
(model with a linear trend)

| $\rho^2$ | | 0.1 | | | 0.2 | | | 0.3 | | | 0.4 | | | 0.5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sig. level | | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $n$ = 2 | ADL | -3.059 | -2.367 | -2.009 | -3.290 | -2.644 | -2.295 | -3.489 | -2.852 | -2.492 | -3.656 | -3.035 | -2.689 | -3.800 | -3.172 | -2.829 |
| | EG2 | -2.963 | -2.313 | -1.970 | -3.238 | -2.576 | -2.230 | -3.398 | -2.759 | -2.424 | -3.558 | -2.923 | -2.576 | -3.684 | -3.040 | -2.715 |
| $n$ = 3 | ADL | -3.101 | -2.420 | -2.062 | -3.393 | -2.737 | -2.376 | -3.637 | -2.974 | -2.628 | -3.820 | -3.155 | -2.816 | -4.012 | -3.353 | -3.006 |
| | EG2 | -2.997 | -2.368 | -2.007 | -3.309 | -2.646 | -2.306 | -3.535 | -2.867 | -2.516 | -3.670 | -3.037 | -2.699 | -3.800 | -3.185 | -2.844 |
| $n$ = 4 | ADL | -3.163 | -2.475 | -2.119 | -3.496 | -2.814 | -2.447 | -3.762 | -3.087 | -2.726 | -3.968 | -3.292 | -2.933 | -4.106 | -3.481 | -3.121 |
| | EG2 | -3.139 | -2.454 | -2.089 | -3.434 | -2.769 | -2.415 | -3.646 | -3.013 | -2.667 | -3.847 | -3.184 | -2.852 | -3.990 | -3.364 | -3.038 |
| $n$ = 5 | ADL | -3.171 | -2.501 | -2.149 | -3.552 | -2.885 | -2.527 | -3.814 | -3.164 | -2.809 | -4.061 | -3.396 | -3.037 | -4.270 | -3.613 | -3.252 |
| | EG2 | -3.225 | -2.547 | -2.188 | -3.533 | -2.884 | -2.545 | -3.822 | -3.170 | -2.816 | -4.031 | -3.386 | -3.051 | -4.223 | -3.583 | -3.249 |
| $n$ = 6 | ADL | -3.261 | -2.563 | -2.205 | -3.647 | -2.959 | -2.594 | -3.909 | -3.245 | -2.883 | -4.179 | -3.522 | -3.148 | -4.422 | -3.713 | -3.365 |
| | EG2 | -3.310 | -2.659 | -2.304 | -3.723 | -3.045 | -2.702 | -3.990 | -3.347 | -2.992 | -4.225 | -3.588 | -3.246 | -4.421 | -3.781 | -3.451 |
| $\rho^2$ | | 0.6 | | | 0.7 | | | 0.8 | | | 0.9 | | | 1 | | |
| Sig. level | | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $n$ = 2 | ADL | -3.927 | -3.308 | -2.978 | -4.069 | -3.436 | -3.103 | -4.158 | -3.551 | -3.225 | -4.255 | -3.633 | -3.321 | -4.376 | -3.738 | -3.415 |
| | EG2 | -3.751 | -3.149 | -2.844 | -3.862 | -3.240 | -2.935 | -3.954 | -3.352 | -3.050 | -4.056 | -3.431 | -3.121 | -4.090 | -3.506 | -3.202 |
| $n$ = 3 | ADL | -4.145 | -3.502 | -3.156 | -4.242 | -3.621 | -3.280 | -4.382 | -3.751 | -3.417 | -4.524 | -3.861 | -3.530 | -4.614 | -3.977 | -3.653 |
| | EG2 | -3.901 | -3.294 | -2.979 | -4.022 | -3.405 | -3.079 | -4.089 | -3.498 | -3.192 | -4.195 | -3.587 | -3.286 | -4.274 | -3.678 | -3.372 |
| $n$ = 4 | ADL | -4.290 | -3.633 | -3.294 | -4.462 | -3.794 | -3.451 | -4.563 | -3.921 | -3.581 | -4.744 | -4.064 | -3.721 | -4.842 | -4.169 | -3.835 |
| | EG2 | -4.114 | -3.505 | -3.182 | -4.242 | -3.628 | -3.318 | -4.338 | -3.744 | -3.436 | -4.474 | -3.870 | -3.561 | -4.555 | -3.951 | -3.653 |
| $n$ = 5 | ADL | -4.440 | -3.786 | -3.430 | -4.633 | -3.956 | -3.598 | -4.786 | -4.115 | -3.759 | -4.923 | -4.235 | -3.888 | -5.059 | -4.371 | -4.027 |
| | EG2 | -4.353 | -3.749 | -3.421 | -4.502 | -3.886 | -3.566 | -4.657 | -4.035 | -3.716 | -4.762 | -4.155 | -3.840 | -4.884 | -4.275 | -3.957 |
| $n$ = 6 | ADL | -4.603 | -3.915 | -3.555 | -4.788 | -4.098 | -3.740 | -4.940 | -4.263 | -3.907 | -5.047 | -4.396 | -4.040 | -5.233 | -4.555 | -4.201 |
| | EG2 | -4.564 | -3.963 | -3.635 | -4.793 | -4.155 | -3.826 | -4.921 | -4.308 | -3.996 | -5.072 | -4.450 | -4.127 | -5.201 | -4.583 | -4.259 |

Table 3.3 Size Property: Baseline case $(\psi = \theta = 0)$

| $\phi_1$ | $\sigma_2^2$ | distribution | Existing tests | | | | RALS Cointegration tests | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $t_{ECM}$ | $t_{ADL}$ | $t_{EG}$ | $t_{EG2}$ | $t^*_{ECM}$ | $t^*_{ADL}$ | $t^*_{EG}$ | $t^*_{EG2}$ |
| 0.5 | 1 | Normal | 0.043 | 0.052 | 0.051 | 0.052 | 0.130 | 0.086 | 0.128 | 0.112 |
| | | $\chi^2(2)$ | 0.036 | 0.047 | 0.046 | 0.047 | 0.073 | 0.054 | 0.134 | 0.096 |
| | | $t(2)$ | 0.038 | 0.049 | 0.050 | 0.049 | 0.084 | 0.062 | 0.133 | 0.104 |
| | 6 | Normal | 0.009 | 0.051 | 0.050 | 0.051 | 0.077 | 0.083 | 0.128 | 0.109 |
| | | $\chi^2(2)$ | 0.007 | 0.049 | 0.047 | 0.048 | 0.041 | 0.053 | 0.132 | 0.095 |
| | | $t(2)$ | 0.008 | 0.050 | 0.048 | 0.049 | 0.050 | 0.062 | 0.132 | 0.104 |
| | 16 | Normal | 0.003 | 0.053 | 0.052 | 0.053 | 0.051 | 0.083 | 0.132 | 0.112 |
| | | $\chi^2(2)$ | 0.004 | 0.045 | 0.047 | 0.045 | 0.032 | 0.055 | 0.134 | 0.095 |
| | | $t(2)$ | 0.005 | 0.048 | 0.048 | 0.047 | 0.036 | 0.060 | 0.129 | 0.107 |
| 1.0 | 1 | Normal | 0.047 | 0.046 | 0.047 | 0.046 | 0.200 | 0.082 | 0.126 | 0.103 |
| | | $\chi^2(2)$ | 0.044 | 0.045 | 0.046 | 0.045 | 0.087 | 0.054 | 0.134 | 0.096 |
| | | $t(2)$ | 0.046 | 0.050 | 0.049 | 0.049 | 0.120 | 0.066 | 0.134 | 0.110 |
| | 6 | Normal | 0.051 | 0.052 | 0.049 | 0.051 | 0.325 | 0.086 | 0.127 | 0.110 |
| | | $\chi^2(2)$ | 0.044 | 0.043 | 0.045 | 0.043 | 0.091 | 0.053 | 0.130 | 0.093 |
| | | $t(2)$ | 0.046 | 0.048 | 0.046 | 0.047 | 0.127 | 0.059 | 0.130 | 0.106 |
| | 16 | Normal | 0.048 | 0.049 | 0.048 | 0.049 | 0.322 | 0.081 | 0.128 | 0.108 |
| | | $\chi^2(2)$ | 0.047 | 0.048 | 0.048 | 0.048 | 0.092 | 0.055 | 0.132 | 0.096 |
| | | $t(2)$ | 0.044 | 0.045 | 0.046 | 0.045 | 0.129 | 0.061 | 0.130 | 0.107 |

Table 3.4 Size Property: Effects of Persistent Contemporaneous term ($\psi \neq 0$)

| $\psi$ | $\sigma_2^2$ | distribution | Existing tests | | | RALS Cointegration tests | | |
|---|---|---|---|---|---|---|---|---|
| | | | $t_{ADL}$ | $t_{EG}$ | $t_{EG2}$ | $t^*_{ADL}$ | $t^*_{EG}$ | $t^*_{EG2}$ |
| 0.6 | 1 | Normal | 0.059 | 0.038 | 0.058 | 0.094 | 0.085 | 0.109 |
| | | $\chi^2(2)$ | 0.056 | 0.033 | 0.055 | 0.056 | 0.117 | 0.079 |
| | | $t(2)$ | 0.056 | 0.034 | 0.055 | 0.064 | 0.104 | 0.087 |
| | 16 | Normal | 0.058 | 0.034 | 0.057 | 0.093 | 0.083 | 0.109 |
| | | $\chi^2(2)$ | 0.054 | 0.033 | 0.053 | 0.056 | 0.117 | 0.078 |
| | | $t(2)$ | 0.053 | 0.033 | 0.052 | 0.060 | 0.102 | 0.086 |
| 0.9 | 1 | Normal | 0.087 | 0.042 | 0.076 | 0.129 | 0.085 | 0.126 |
| | | $\chi^2(2)$ | 0.080 | 0.038 | 0.069 | 0.061 | 0.129 | 0.076 |
| | | $t(2)$ | 0.085 | 0.041 | 0.073 | 0.075 | 0.114 | 0.084 |
| | 16 | Normal | 0.085 | 0.040 | 0.075 | 0.127 | 0.085 | 0.125 |
| | | $\chi^2(2)$ | 0.079 | 0.037 | 0.070 | 0.064 | 0.126 | 0.074 |
| | | $t(2)$ | 0.086 | 0.043 | 0.076 | 0.077 | 0.116 | 0.089 |

Note: These results are obtained for the case where $\phi_1 = 0.5$. The results for $\phi_1 = 1$ and $\sigma_2^2 = 6$ show similar patterns.

Table 3.5 Size Property: Combined Effects of Persistent Contemporaneous term and Correlated Errors

| $\psi$ | $\theta$ | $\sigma_2^2$ | distribution | Existing tests | | | RALS Cointegration tests | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | $t_{ADL}$ | $t_{EG}$ | $t_{EG2}$ | $t^*_{ADL}$ | $t^*_{EG}$ | $t^*_{EG2}$ |
| 0.0 | 0.3 | 1 | Normal | 0.050 | 0.050 | 0.051 | 0.079 | 0.127 | 0.106 |
| | | | $\chi^2(2)$ | 0.045 | 0.046 | 0.045 | 0.055 | 0.134 | 0.097 |
| | | | $t(2)$ | 0.047 | 0.049 | 0.047 | 0.060 | 0.129 | 0.101 |
| | | 16 | Normal | 0.047 | 0.049 | 0.047 | 0.082 | 0.125 | 0.104 |
| | | | $\chi^2(2)$ | 0.044 | 0.046 | 0.044 | 0.054 | 0.133 | 0.096 |
| | | | $t(2)$ | 0.050 | 0.046 | 0.050 | 0.062 | 0.132 | 0.103 |
| | 0.5 | 1 | Normal | 0.048 | 0.048 | 0.048 | 0.080 | 0.126 | 0.106 |
| | | | $\chi^2(2)$ | 0.047 | 0.047 | 0.047 | 0.056 | 0.134 | 0.096 |
| | | | $t(2)$ | 0.048 | 0.048 | 0.048 | 0.059 | 0.132 | 0.105 |
| | | 16 | Normal | 0.050 | 0.049 | 0.050 | 0.083 | 0.130 | 0.110 |
| | | | $\chi^2(2)$ | 0.046 | 0.046 | 0.046 | 0.054 | 0.130 | 0.094 |
| | | | $t(2)$ | 0.049 | 0.049 | 0.049 | 0.060 | 0.131 | 0.107 |
| 0.5 | 0.3 | 1 | Normal | 0.072 | 0.039 | 0.071 | 0.109 | 0.094 | 0.132 |
| | | | $\chi^2(2)$ | 0.070 | 0.035 | 0.068 | 0.082 | 0.114 | 0.116 |
| | | | $t(2)$ | 0.073 | 0.037 | 0.072 | 0.087 | 0.110 | 0.120 |
| | | 16 | Normal | 0.057 | 0.035 | 0.057 | 0.092 | 0.088 | 0.110 |
| | | | $\chi^2(2)$ | 0.051 | 0.032 | 0.052 | 0.056 | 0.114 | 0.079 |
| | | | $t(2)$ | 0.054 | 0.036 | 0.053 | 0.066 | 0.109 | 0.096 |
| | 0.5 | 1 | Normal | 0.119 | 0.048 | 0.116 | 0.159 | 0.114 | 0.186 |
| | | | $\chi^2(2)$ | 0.118 | 0.044 | 0.115 | 0.136 | 0.114 | 0.178 |
| | | | $t(2)$ | 0.118 | 0.048 | 0.116 | 0.134 | 0.115 | 0.179 |
| | | 16 | Normal | 0.061 | 0.040 | 0.060 | 0.091 | 0.091 | 0.113 |
| | | | $\chi^2(2)$ | 0.055 | 0.034 | 0.053 | 0.063 | 0.115 | 0.089 |
| | | | $t(2)$ | 0.055 | 0.033 | 0.054 | 0.066 | 0.106 | 0.094 |

Note: See the footnote of Table 3.4

Table 3.6 Power Property: Baseline case $(\psi = \theta = 0)$

| $\phi_1$ | $\sigma_2^2$ | distribution | Existing tests | | | RALS Cointegration tests | | |
|---|---|---|---|---|---|---|---|---|
| | | | $t_{ADL}$ | $t_{EG}$ | $t_{EG2}$ | $t^*_{ADL}$ | $t^*_{EG}$ | $t^*_{EG2}$ |
| 0.5 | 1 | Normal | 0.291 | 0.186 | 0.292 | 0.253 | 0.161 | 0.250 |
| | | $\chi^2(2)$ | 0.312 | 0.200 | 0.314 | 0.990 | 0.762 | 0.951 |
| | | $t(2)$ | 0.306 | 0.188 | 0.307 | 0.948 | 0.614 | 0.853 |
| | 6 | Normal | 0.959 | 0.067 | 0.945 | 0.945 | 0.066 | 0.917 |
| | | $\chi^2(2)$ | 0.957 | 0.073 | 0.944 | 1.000 | 0.190 | 0.995 |
| | | $t(2)$ | 0.959 | 0.071 | 0.946 | 1.000 | 0.144 | 0.986 |
| | 16 | Normal | 1.000 | 0.052 | 1.000 | 1.000 | 0.051 | 0.999 |
| | | $\chi^2(2)$ | 1.000 | 0.061 | 1.000 | 1.000 | 0.146 | 1.000 |
| | | $t(2)$ | 1.000 | 0.057 | 1.000 | 1.000 | 0.120 | 1.000 |
| 1.0 | 1 | Normal | 0.266 | 0.232 | 0.271 | 0.234 | 0.198 | 0.235 |
| | | $\chi^2(2)$ | 0.261 | 0.224 | 0.265 | 0.988 | 0.933 | 0.972 |
| | | $t(2)$ | 0.247 | 0.217 | 0.251 | 0.923 | 0.802 | 0.869 |
| | 6 | Normal | 0.249 | 0.228 | 0.254 | 0.226 | 0.200 | 0.229 |
| | | $\chi^2(2)$ | 0.268 | 0.222 | 0.274 | 0.986 | 0.936 | 0.971 |
| | | $t(2)$ | 0.259 | 0.234 | 0.264 | 0.930 | 0.818 | 0.885 |
| | 16 | Normal | 0.254 | 0.219 | 0.259 | 0.234 | 0.187 | 0.225 |
| | | $\chi^2(2)$ | 0.250 | 0.217 | 0.254 | 0.986 | 0.930 | 0.970 |
| | | $t(2)$ | 0.262 | 0.225 | 0.269 | 0.926 | 0.811 | 0.877 |

Table 3.7 Power Property: Effects of Persistent Contemporaneous term ($\psi \neq 0$, $\theta = 0$)

| $\phi_1$ | $\psi$ | $\sigma_2^2$ | distribution | Existing tests | | | RALS Cointegration tests | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | $t_{ADL}$ | $t_{EG}$ | $t_{EG2}$ | $t^*_{ADL}$ | $t^*_{EG}$ | $t^*_{EG2}$ |
| 0.5 | 0.6 | 1 | Normal | 0.470 | 0.062 | 0.447 | 0.419 | 0.060 | 0.364 |
| | | | $\chi^2(2)$ | 0.503 | 0.068 | 0.477 | 0.997 | 0.450 | 0.979 |
| | | | $t(2)$ | 0.499 | 0.064 | 0.475 | 0.979 | 0.348 | 0.930 |
| | | 6 | Normal | 0.999 | 0.012 | 0.995 | 0.999 | 0.017 | 0.990 |
| | | | $\chi^2(2)$ | 0.999 | 0.013 | 0.995 | 1.000 | 0.037 | 0.999 |
| | | | $t(2)$ | 0.999 | 0.012 | 0.995 | 1.000 | 0.033 | 0.999 |
| | | 16 | Normal | 1.000 | 0.014 | 0.999 | 1.000 | 0.020 | 0.998 |
| | | | $\chi^2(2)$ | 1.000 | 0.014 | 0.999 | 1.000 | 0.041 | 1.000 |
| | | | $t(2)$ | 1.000 | 0.015 | 0.999 | 1.000 | 0.036 | 1.000 |
| | 0.9 | 1 | Normal | 0.867 | 0.011 | 0.590 | 0.831 | 0.016 | 0.513 |
| | | | $\chi^2(2)$ | 0.870 | 0.013 | 0.603 | 1.000 | 0.091 | 0.903 |
| | | | $t(2)$ | 0.868 | 0.012 | 0.584 | 0.999 | 0.072 | 0.847 |
| | | 6 | Normal | 1.000 | 0.036 | 0.847 | 1.000 | 0.056 | 0.822 |
| | | | $\chi^2(2)$ | 1.000 | 0.038 | 0.853 | 1.000 | 0.071 | 0.909 |
| | | | $t(2)$ | 1.000 | 0.037 | 0.857 | 1.000 | 0.073 | 0.897 |
| | | 16 | Normal | 1.000 | 0.038 | 0.861 | 1.000 | 0.061 | 0.838 |
| | | | $\chi^2(2)$ | 1.000 | 0.041 | 0.862 | 1.000 | 0.082 | 0.920 |
| | | | $t(2)$ | 1.000 | 0.036 | 0.861 | 1.000 | 0.081 | 0.900 |
| 1.0 | 0.6 | 1 | Normal | 0.252 | 0.270 | 0.258 | 0.229 | 0.226 | 0.231 |
| | | | $\chi^2(2)$ | 0.251 | 0.280 | 0.262 | 0.984 | 0.955 | 0.98 |
| | | | $t(2)$ | 0.249 | 0.277 | 0.261 | 0.917 | 0.868 | 0.906 |
| | | 6 | Normal | 0.243 | 0.261 | 0.249 | 0.226 | 0.228 | 0.228 |
| | | | $\chi^2(2)$ | 0.266 | 0.283 | 0.272 | 0.983 | 0.952 | 0.979 |
| | | | $t(2)$ | 0.243 | 0.266 | 0.251 | 0.914 | 0.866 | 0.900 |
| | | 16 | Normal | 0.251 | 0.256 | 0.255 | 0.219 | 0.223 | 0.223 |
| | | | $\chi^2(2)$ | 0.252 | 0.283 | 0.259 | 0.985 | 0.953 | 0.980 |
| | | | $t(2)$ | 0.244 | 0.263 | 0.252 | 0.922 | 0.868 | 0.906 |
| | 0.9 | 1 | Normal | 0.224 | 0.253 | 0.242 | 0.209 | 0.218 | 0.224 |
| | | | $\chi^2(2)$ | 0.224 | 0.248 | 0.244 | 0.972 | 0.939 | 0.977 |
| | | | $t(2)$ | 0.234 | 0.248 | 0.245 | 0.895 | 0.850 | 0.907 |
| | | 6 | Normal | 0.229 | 0.244 | 0.248 | 0.211 | 0.210 | 0.215 |
| | | | $\chi^2(2)$ | 0.224 | 0.256 | 0.243 | 0.976 | 0.942 | 0.980 |
| | | | $t(2)$ | 0.220 | 0.240 | 0.233 | 0.894 | 0.845 | 0.903 |
| | | 16 | Normal | 0.231 | 0.258 | 0.246 | 0.204 | 0.219 | 0.223 |
| | | | $\chi^2(2)$ | 0.212 | 0.255 | 0.234 | 0.976 | 0.940 | 0.979 |
| | | | $t(2)$ | 0.218 | 0.247 | 0.238 | 0.897 | 0.851 | 0.909 |

Table 3.8 Power Property: Combined Effects of Persistent Contemporaneous term and Correlated Errors

| $\phi_1$ | $\psi$ | $\theta$ | distribution | Existing tests | | | RALS Cointegration tests | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | $t_{ADL}$ | $t_{EG}$ | $t_{EG2}$ | $t^*_{ADL}$ | $t^*_{EG}$ | $t^*_{EG2}$ |
| 0.5 | 0.0 | 0.3 | Normal | 0.816 | 0.091 | 0.798 | 0.775 | 0.089 | 0.739 |
| | | | $\chi^2(2)$ | 0.827 | 0.096 | 0.807 | 0.999 | 0.248 | 0.977 |
| | | | $t(2)$ | 0.810 | 0.094 | 0.794 | 0.998 | 0.198 | 0.943 |
| | | 0.5 | Normal | 0.800 | 0.092 | 0.781 | 0.757 | 0.085 | 0.712 |
| | | | $\chi^2(2)$ | 0.816 | 0.097 | 0.800 | 0.999 | 0.270 | 0.978 |
| | | | $t(2)$ | 0.803 | 0.088 | 0.788 | 0.998 | 0.201 | 0.941 |
| | 0.5 | 0.3 | Normal | 0.973 | 0.010 | 0.955 | 0.964 | 0.014 | 0.929 |
| | | | $\chi^2(2)$ | 0.972 | 0.012 | 0.957 | 1.000 | 0.057 | 0.997 |
| | | | $t(2)$ | 0.973 | 0.011 | 0.958 | 1.000 | 0.043 | 0.990 |
| | | 0.5 | Normal | 0.967 | 0.009 | 0.946 | 0.955 | 0.013 | 0.914 |
| | | | $\chi^2(2)$ | 0.965 | 0.011 | 0.947 | 1.000 | 0.056 | 0.995 |
| | | | $t(2)$ | 0.964 | 0.011 | 0.946 | 1.000 | 0.044 | 0.990 |
| 1.0 | 0.0 | 0.3 | Normal | 0.274 | 0.216 | 0.279 | 0.247 | 0.182 | 0.243 |
| | | | $\chi^2(2)$ | 0.281 | 0.212 | 0.285 | 0.988 | 0.863 | 0.961 |
| | | | $t(2)$ | 0.282 | 0.207 | 0.284 | 0.939 | 0.713 | 0.865 |
| | | 0.5 | Normal | 0.324 | 0.185 | 0.324 | 0.293 | 0.168 | 0.288 |
| | | | $\chi^2(2)$ | 0.347 | 0.191 | 0.349 | 0.992 | 0.721 | 0.944 |
| | | | $t(2)$ | 0.319 | 0.179 | 0.320 | 0.958 | 0.573 | 0.854 |
| | 0.5 | 0.3 | Normal | 0.343 | 0.252 | 0.348 | 0.312 | 0.219 | 0.305 |
| | | | $\chi^2(2)$ | 0.357 | 0.263 | 0.360 | 0.983 | 0.871 | 0.967 |
| | | | $t(2)$ | 0.340 | 0.247 | 0.343 | 0.920 | 0.746 | 0.878 |
| | | 0.5 | Normal | 0.475 | 0.223 | 0.475 | 0.436 | 0.190 | 0.426 |
| | | | $\chi^2(2)$ | 0.499 | 0.215 | 0.494 | 0.982 | 0.674 | 0.946 |
| | | | $t(2)$ | 0.488 | 0.224 | 0.489 | 0.933 | 0.549 | 0.862 |

Note: These results are obtained for the case where $\sigma_2^2 = 16$. The results for $\sigma_2^2 = 1, 6$ show similar patterns.

Appendix Table 1. Critical values of RALS cointegration tests (T=500)
(model with a constant)

| $\rho^2$ | | 0.1 | | | 0.2 | | | 0.3 | | | 0.4 | | | 0.5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sig. level | | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $n$ = 2 | ADL | -2.847 | -2.202 | -1.839 | -3.059 | -2.412 | -2.063 | -3.267 | -2.576 | -2.225 | -3.364 | -2.713 | -2.355 | -3.436 | -2.813 | -2.479 |
| | EG2 | -2.818 | -2.167 | -1.799 | -2.977 | -2.352 | -1.997 | -3.137 | -2.494 | -2.146 | -3.212 | -2.584 | -2.261 | -3.302 | -2.694 | -2.367 |
| $n$ = 3 | ADL | -2.946 | -2.273 | -1.904 | -3.210 | -2.528 | -2.175 | -3.359 | -2.707 | -2.359 | -3.497 | -2.882 | -2.531 | -3.678 | -3.025 | -2.673 |
| | EG2 | -2.940 | -2.274 | -1.920 | -3.205 | -2.513 | -2.154 | -3.330 | -2.697 | -2.353 | -3.485 | -2.857 | -2.519 | -3.577 | -2.957 | -2.641 |
| $n$ = 4 | ADL | -3.012 | -2.351 | -1.984 | -3.295 | -2.638 | -2.269 | -3.548 | -2.865 | -2.503 | -3.684 | -3.019 | -2.668 | -3.789 | -3.176 | -2.833 |
| | EG2 | -3.061 | -2.392 | -2.039 | -3.349 | -2.693 | -2.337 | -3.544 | -2.913 | -2.571 | -3.732 | -3.108 | -2.781 | -3.834 | -3.229 | -2.906 |
| $n$ = 5 | ADL | -3.106 | -2.418 | -2.047 | -3.416 | -2.731 | -2.378 | -3.635 | -2.974 | -2.609 | -3.807 | -3.173 | -2.828 | -3.966 | -3.331 | -2.980 |
| | EG2 | -3.167 | -2.502 | -2.153 | -3.505 | -2.848 | -2.507 | -3.751 | -3.099 | -2.763 | -3.946 | -3.322 | -2.986 | -4.101 | -3.486 | -3.164 |
| $n$ = 6 | ADL | -3.176 | -2.473 | -2.112 | -3.484 | -2.812 | -2.463 | -3.720 | -3.055 | -2.712 | -3.947 | -3.303 | -2.937 | -4.128 | -3.466 | -3.125 |
| | EG2 | -3.304 | -2.633 | -2.274 | -3.683 | -3.001 | -2.646 | -3.918 | -3.285 | -2.944 | -4.152 | -3.525 | -3.190 | -4.322 | -3.705 | -3.379 |
| $\rho^2$ | | 0.6 | | | 0.7 | | | 0.8 | | | 0.9 | | | 1 | | |
| Sig. level | | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $n$ = 2 | ADL | -3.577 | -2.931 | -2.590 | -3.629 | -3.009 | -2.686 | -3.688 | -3.088 | -2.758 | -3.739 | -3.137 | -2.833 | -3.792 | -3.209 | -2.896 |
| | EG2 | -3.369 | -2.774 | -2.455 | -3.441 | -2.838 | -2.527 | -3.476 | -2.892 | -2.587 | -3.535 | -2.948 | -2.652 | -3.549 | -2.990 | -2.702 |
| $n$ = 3 | ADL | -3.762 | -3.129 | -2.785 | -3.855 | -3.226 | -2.897 | -3.939 | -3.336 | -3.013 | -4.022 | -3.424 | -3.105 | -4.124 | -3.522 | -3.201 |
| | EG2 | -3.669 | -3.062 | -2.747 | -3.755 | -3.174 | -2.854 | -3.830 | -3.250 | -2.942 | -3.900 | -3.326 | -3.025 | -3.938 | -3.393 | -3.103 |
| $n$ = 4 | ADL | -3.957 | -3.322 | -2.976 | -4.089 | -3.433 | -3.103 | -4.173 | -3.554 | -3.228 | -4.285 | -3.673 | -3.340 | -4.370 | -3.761 | -3.434 |
| | EG2 | -4.001 | -3.366 | -3.043 | -4.084 | -3.489 | -3.174 | -4.168 | -3.588 | -3.291 | -4.255 | -3.695 | -3.395 | -4.348 | -3.778 | -3.486 |
| $n$ = 5 | ADL | -4.132 | -3.497 | -3.155 | -4.266 | -3.625 | -3.283 | -4.415 | -3.762 | -3.420 | -4.525 | -3.903 | -3.566 | -4.598 | -3.990 | -3.665 |
| | EG2 | -4.242 | -3.646 | -3.319 | -4.388 | -3.785 | -3.479 | -4.481 | -3.921 | -3.611 | -4.588 | -4.028 | -3.725 | -4.692 | -4.124 | -3.839 |
| $n$ = 6 | ADL | -4.299 | -3.658 | -3.296 | -4.445 | -3.793 | -3.453 | -4.580 | -3.944 | -3.602 | -4.693 | -4.067 | -3.736 | -4.816 | -4.199 | -3.864 |
| | EG2 | -4.489 | -3.889 | -3.573 | -4.623 | -4.044 | -3.732 | -4.757 | -4.182 | -3.883 | -4.888 | -4.327 | -4.036 | -4.983 | -4.447 | -4.162 |

Appendix Table 2. Critical values of RALS cointegration tests ($T$ = 500)
(model with a linear trend)

| $\rho^2$ | | 0.1 | | | 0.2 | | | 0.3 | | | 0.4 | | | 0.5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sig. level | | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $n$ = 2 | ADL | -3.079 | -2.382 | -2.010 | -3.293 | -2.654 | -2.307 | -3.512 | -2.847 | -2.502 | -3.633 | -3.023 | -2.696 | -3.791 | -3.159 | -2.836 |
| | EG2 | -3.010 | -2.337 | -1.978 | -3.225 | -2.581 | -2.243 | -3.388 | -2.770 | -2.435 | -3.512 | -2.909 | -2.586 | -3.662 | -3.054 | -2.727 |
| $n$ = 3 | ADL | -3.119 | -2.420 | -2.062 | -3.410 | -2.737 | -2.385 | -3.600 | -2.969 | -2.626 | -3.811 | -3.162 | -2.811 | -3.971 | -3.337 | -2.988 |
| | EG2 | -3.075 | -2.388 | -2.034 | -3.311 | -2.669 | -2.332 | -3.499 | -2.874 | -2.531 | -3.674 | -3.047 | -2.719 | -3.807 | -3.185 | -2.866 |
| $n$ = 4 | ADL | -3.174 | -2.490 | -2.126 | -3.484 | -2.810 | -2.468 | -3.730 | -3.077 | -2.729 | -3.942 | -3.285 | -2.946 | -4.093 | -3.471 | -3.132 |
| | EG2 | -3.157 | -2.462 | -2.101 | -3.420 | -2.794 | -2.439 | -3.654 | -3.007 | -2.664 | -3.854 | -3.208 | -2.874 | -3.976 | -3.370 | -3.053 |
| $n$ = 5 | ADL | -3.186 | -2.538 | -2.172 | -3.589 | -2.916 | -2.561 | -3.856 | -3.180 | -2.826 | -4.035 | -3.395 | -3.057 | -4.231 | -3.591 | -3.245 |
| | EG2 | -3.226 | -2.559 | -2.202 | -3.559 | -2.915 | -2.558 | -3.812 | -3.177 | -2.835 | -4.006 | -3.390 | -3.060 | -4.170 | -3.564 | -3.253 |
| $n$ = 6 | ADL | -3.289 | -2.586 | -2.228 | -3.624 | -2.971 | -2.619 | -3.932 | -3.267 | -2.917 | -4.169 | -3.485 | -3.153 | -4.370 | -3.725 | -3.390 |
| | EG2 | -3.282 | -2.628 | -2.282 | -3.711 | -3.028 | -2.679 | -3.974 | -3.322 | -2.990 | -4.183 | -3.560 | -3.236 | -4.376 | -3.753 | -3.434 |
| $\rho^2$ | | 0.6 | | | 0.7 | | | 0.8 | | | 0.9 | | | 1 | | |
| Sig. level | | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $n$ = 2 | ADL | -3.949 | -3.310 | -2.981 | -4.002 | -3.397 | -3.089 | -4.109 | -3.514 | -3.203 | -4.177 | -3.602 | -3.296 | -4.247 | -3.684 | -3.385 |
| | EG2 | -3.752 | -3.175 | -2.848 | -3.833 | -3.240 | -2.935 | -3.920 | -3.342 | -3.049 | -3.967 | -3.407 | -3.116 | -4.047 | -3.485 | -3.200 |
| $n$ = 3 | ADL | -4.092 | -3.481 | -3.144 | -4.187 | -3.596 | -3.273 | -4.335 | -3.726 | -3.407 | -4.421 | -3.846 | -3.531 | -4.539 | -3.950 | -3.637 |
| | EG2 | -3.897 | -3.310 | -2.990 | -3.981 | -3.408 | -3.095 | -4.063 | -3.497 | -3.209 | -4.180 | -3.605 | -3.312 | -4.200 | -3.662 | -3.386 |
| $n$ = 4 | ADL | -4.260 | -3.623 | -3.291 | -4.393 | -3.777 | -3.450 | -4.522 | -3.906 | -3.579 | -4.635 | -4.041 | -3.712 | -4.740 | -4.139 | -3.827 |
| | EG2 | -4.081 | -3.512 | -3.193 | -4.218 | -3.631 | -3.327 | -4.316 | -3.756 | -3.454 | -4.408 | -3.845 | -3.554 | -4.509 | -3.942 | -3.661 |
| $n$ = 5 | ADL | -4.408 | -3.769 | -3.431 | -4.572 | -3.943 | -3.599 | -4.689 | -4.075 | -3.749 | -4.808 | -4.215 | -3.887 | -4.963 | -4.351 | -4.030 |
| | EG2 | -4.335 | -3.718 | -3.412 | -4.481 | -3.874 | -3.564 | -4.584 | -4.013 | -3.713 | -4.673 | -4.112 | -3.820 | -4.759 | -4.214 | -3.931 |
| $n$ = 6 | ADL | -4.530 | -3.909 | -3.567 | -4.720 | -4.093 | -3.747 | -4.895 | -4.246 | -3.911 | -5.011 | -4.385 | -4.056 | -5.133 | -4.521 | -4.195 |
| | EG2 | -4.534 | -3.934 | -3.623 | -4.683 | -4.108 | -3.806 | -4.835 | -4.260 | -3.970 | -4.955 | -4.383 | -4.093 | -5.068 | -4.516 | -4.225 |

Appendix Table 3. Size Properties of the RALS cointegration tests compared with existing tests ($\delta_1 = 0, T = 100$)

| $\phi_1$ | $\sigma_2^2$ | distribution | Existing tests | | | | RALS Cointegration tests | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $t_{ECM}$ | $t_{ADL}$ | $t_{EG}$ | $t_{EG2}$ | $t^*_{ECM}$ | $t^*_{ADL}$ | $t^*_{EG}$ | $t^*_{EG2}$ |
| 1 | 1 | $\chi^2(1)$ | 0.046 | 0.045 | 0.047 | 0.047 | 0.061 | 0.062 | 0.071 | 0.078 |
| | | $\chi^2(3)$ | 0.048 | 0.046 | 0.049 | 0.049 | 0.077 | 0.083 | 0.084 | 0.086 |
| | | $t(3)$ | 0.055 | 0.053 | 0.055 | 0.055 | 0.042 | 0.020 | 0.058 | 0.058 |
| | | $t(4)$ | 0.051 | 0.050 | 0.050 | 0.052 | 0.077 | 0.097 | 0.060 | 0.061 |
| 0.5 | 6 | $\chi^2(1)$ | 0.007 | 0.046 | 0.049 | 0.048 | 0.016 | 0.063 | 0.075 | 0.079 |
| | | $\chi^2(3)$ | 0.008 | 0.049 | 0.050 | 0.051 | 0.018 | 0.080 | 0.074 | 0.087 |
| | | $t(3)$ | 0.025 | 0.052 | 0.053 | 0.054 | 0.019 | 0.021 | 0.055 | 0.058 |
| | | $t(4)$ | 0.008 | 0.050 | 0.051 | 0.052 | 0.013 | 0.098 | 0.058 | 0.059 |
| 0.5 | 16 | $\chi^2(1)$ | 0.004 | 0.045 | 0.047 | 0.046 | 0.009 | 0.062 | 0.076 | 0.083 |
| | | $\chi^2(3)$ | 0.004 | 0.046 | 0.047 | 0.049 | 0.010 | 0.082 | 0.079 | 0.084 |
| | | $t(3)$ | 0.009 | 0.053 | 0.055 | 0.055 | 0.010 | 0.021 | 0.055 | 0.058 |
| | | $t(4)$ | 0.004 | 0.052 | 0.052 | 0.054 | 0.007 | 0.100 | 0.050 | 0.055 |

Note: The model with a constant is considered in this table. There is one integrated regressor in the model ($n$ = 2).

Appendix Table 4. Power Properties of the RALS cointegration tests compared with existing tests ($\delta_1 = -0.1, T = 100$)

| $\phi_1$ | $\sigma_2^2$ | distribution | Existing tests | | | | RALS Cointegration tests | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $t_{ECM}$ | $t_{ADL}$ | $t_{EG}$ | $t_{EG2}$ | $t^*_{ECM}$ | $t^*_{ADL}$ | $t^*_{EG}$ | $t^*_{EG2}$ |
| 1 | 1 | $\chi^2(1)$ | 0.314 | 0.238 | 0.207 | 0.241 | 0.998 | 0.987 | 0.966 | 0.985 |
| | | $\chi^2(3)$ | 0.321 | 0.247 | 0.219 | 0.250 | 0.971 | 0.861 | 0.823 | 0.853 |
| | | $t(3)$ | 0.317 | 0.241 | 0.211 | 0.245 | 0.822 | 0.466 | 0.430 | 0.431 |
| | | $t(4)$ | 0.319 | 0.246 | 0.216 | 0.250 | 0.765 | 0.381 | 0.348 | 0.343 |
| 0.5 | 6 | $\chi^2(1)$ | 0.995 | 0.955 | 0.066 | 0.940 | 1.000 | 1.000 | 0.128 | 0.998 |
| | | $\chi^2(3)$ | 0.998 | 0.961 | 0.068 | 0.947 | 1.000 | 0.999 | 0.122 | 0.993 |
| | | $t(3)$ | 0.995 | 0.959 | 0.068 | 0.947 | 1.000 | 0.992 | 0.117 | 0.991 |
| | | $t(4)$ | 0.996 | 0.955 | 0.067 | 0.943 | 1.000 | 0.981 | 0.122 | 0.984 |
| 0.5 | 16 | $\chi^2(1)$ | 1.000 | 1.000 | 0.058 | 1.000 | 1.000 | 1.000 | 0.101 | 1.000 |
| | | $\chi^2(3)$ | 1.000 | 1.000 | 0.052 | 1.000 | 1.000 | 1.000 | 0.094 | 1.000 |
| | | $t(3)$ | 1.000 | 0.999 | 0.055 | 0.999 | 1.000 | 1.000 | 0.101 | 1.000 |
| | | $t(4)$ | 1.000 | 1.000 | 0.058 | 1.000 | 1.000 | 1.000 | 0.102 | 1.000 |

Note: The model with a constant is considered in this table. There is one integrated regressor in the model (*n* = 2).

Appendix Table 5. Size adjusted Power Properties of the RALS cointegration tests compared with existing tests ($\delta_1 = -0.1, T = 100$)

| $\phi_1$ | $\sigma_2^2$ | distribution | Existing tests | | | RALS Cointegration tests | | |
|---|---|---|---|---|---|---|---|---|
| | | | $t_{ADL}$ | $t_{EG}$ | $t_{EG2}$ | $t^*_{ADL}$ | $t^*_{EG}$ | $t^*_{EG2}$ |
| 1 | 1 | $\chi^2(1)$ | 0.257 | 0.225 | 0.264 | 0.986 | 0.962 | 0.982 |
| | | $\chi^2(3)$ | 0.251 | 0.222 | 0.256 | 0.839 | 0.777 | 0.805 |
| | | $t(3)$ | 0.244 | 0.220 | 0.249 | 0.482 | 0.418 | 0.459 |
| | | $t(4)$ | 0.248 | 0.216 | 0.254 | 0.366 | 0.317 | 0.340 |
| 0.5 | 6 | $\chi^2(1)$ | 0.959 | 0.070 | 0.947 | 1.000 | 0.254 | 0.999 |
| | | $\chi^2(3)$ | 0.959 | 0.068 | 0.946 | 1.000 | 0.162 | 0.991 |
| | | $t(3)$ | 0.956 | 0.065 | 0.944 | 0.992 | 0.106 | 0.986 |
| | | $t(4)$ | 0.957 | 0.066 | 0.944 | 0.982 | 0.088 | 0.971 |
| 0.5 | 16 | $\chi^2(1)$ | 1.000 | 0.060 | 0.999 | 1.000 | 0.163 | 1.000 |
| | | $\chi^2(3)$ | 1.000 | 0.060 | 1.000 | 1.000 | 0.106 | 1.000 |
| | | $t(3)$ | 1.000 | 0.052 | 0.999 | 1.000 | 0.077 | 1.000 |
| | | $t(4)$ | 1.000 | 0.055 | 1.000 | 1.000 | 0.068 | 1.000 |

Note: The model with a constant has been considered in this table. There is one integrated regressor in the model ($n$ = 2).

# CONCLUSION

In this dissertation, we have tried to find ways to improve the power of cointegration tests. Indeed, the task of seeking more powerful tests is not a trivial concern. This is so, since the power of the usual unit root and cointegration tests is low and the usual tests often fail to reject the null hypothesis. This dissertation is focused on finding more powerful cointegration tests. First, the proposed modified Engle and Granger (1987, EG) tests appear promising. When the first differenced regressors are added to the testing regression, the power of the modified EG tests will improve significantly. Although the modified EG test has not been considered in the literature as it will induce a nuisance parameter, this does not pose a problem when we use the relevant critical values that are based on the estimated value of the nuisance parameter. In fact, the modified EG test show the good size property under the null. More importantly, it is shown that the power in the modified EG test increases significantly as the signal-to-noise ratio increases, which was an important source of power loss in the usual EG tests. In contrast, it is very encouraging that the signal-to-noise ratio in the modified test becomes a useful factor to improve power without affecting performance under the null. As such, the modified EG test provides a solution to the problem of the common factor restriction (CFR) that was initially examined in Kremers et al. (1992). One caveat of the modified EG test is that the procedure works in a single equation framework where weak exogeneity of the regressors is assumed. However, it would be interesting to see if the same advantage of improving the power of the cointegration test will occur in a system of equations when we add similar terms in the vector error correction models (VECM). This remains as a topic for future research.

In this dissertation, we have also examined the performance of some popular nonlinear unit root tests in the presence of non-normal errors. Surprisingly, this topic has not been examined in the literature. Intuitively, we conjecture that many popular nonlinear unit root tests will suffer from a loss of power in the presence of non-normal errors, which is not the case in linear unit root tests. The intuition is simple. The nonlinear models will be mis-specified in the presence of non-normal errors, but the critical values are often derived by the model assuming normality of the error term. This problem will not occur in the linear models due to the central limit theorem, and the linear estimator is the most efficient under certain assumptions that do not include the normality assumption. Although we have not examined nonlinear cointegration tests, we expect that the same phenomenon will be observed. This finding provides a warning to users of the popular nonlinear unit root and cointegration tests that the underlying assumption of the error term should be correctly specified; otherwise, those tests can lose power. In the second essay, we did not examine the threshold unit root tests of Enders and Granger (1998) and the threshold cointegration tests of Enders and Siklos (2001). As such, it remains to be seen if they are robust to the presence of non-normal errors. If they are robust, this finding will add another important feature to threshold unit root and cointegration tests. This topic remains for future research.

In the third essay, we have suggested new cointegration tests that are more powerful in the presence of non-normal errors. The information of non-normal errors is utilized in a linearized testing regression. The suggested tests have the same property as the tests based on the generalized methods of moments where nonlinear moment conditions of the residuals are employed to capture unknown forms of non-normal distributions. The underlying idea is to adopt a two-step procedure following the "residual augmented least squares" (RALS) method of Im and Schmidt (2008), which can make use of nonlinear moment conditions driven by non-normal errors. If the errors are non-normal, the higher moments of the error term (or residuals) will contain information on the nature of the non-normal errors. If we can utilize this information in the higher moments of the residuals, then we can potentially obtain more powerful cointegration tests. The suggested RALS testing procedure is easy to implement because it does not require non-linear estimation techniques even though we utilize nonlinear moment conditions associated with the non-normal errors. Then, additional terms are added to existing testing regressions in the linearized testing regression, which works to reduce the variance of the error term under certain regularity conditions. As we know already, a smaller variance leads to a smaller confidence interval and this makes the estimator more efficient. The conditions employed in this essay are quite general and seem to work over various non-normal distributions. That is, this methodology does not require any pre-testing procedures to find a specific type of distribution that satisfies desired features. In addition, the newly added term is easily obtained inside the system. The finite sample based simulation confirms this conjecture of better performance in terms of efficiency: compared to the existing cointegration tests, the ability to reject a false null is significantly improved and does not rely on any type of non-normal distribution. One important feature is that we still utilize the linearized testing regression. In light of the finding in the second essay that the usual nonlinear tests will suffer from loss of power, using a linearized regression will be appealing as we can rely on the nice features of linear based estimators.

It will be interesting to examine whether these encouraging results showing improved power will work in more general models in a system of equations. This question is important since our promising results shown in the third essay are based on a single equation model. However, it seems likely that the procedure will also work in a system error correction model. If so, it seems likely that the power in the Johansen-type cointegration test can be improved. In addition, it seems promising to see if this same outcome will hold in nonlinear unit root and cointegration tests. These issues can be examined in future research.

REFERENCES


Banerjee, A., J.J. Dolado, D.F. Hendry, and G.W. Smith (1986), Exploring Equilibrium Relationship in Econometrics through Static Models: Some Monte Carlo Evidence. *Oxford Bulletin of Economics and Statistics* 48, 253-277.

Banerjee, A., J.J. Dolado, and R. Mestre (1998), "Error-Correction Mechanism Tests for Cointegration in a Single-Equation Framework," *Journal of Time Series Analysis*, 19 (3), 267-283.

Boswijk, H.P. and P.H. Frances (1992), "Dynamic Specification and Cointegration," *Oxford Bulletin of Economics and Statistics* 54, 369-381.

Boswijk, H.P. (1994), "Testing for an Unstable Root in Conditional and Structural Error Correction Models," *Journal of Econometrics* 63, 37-60.

Choi, C.Y. and Y.K. Moh (2007), "How Useful are Tests for Unit-root in Distinguishing Unit-root Processes from Stationary but Non-linear Processes?" *Econometrics Journal*, 10, 82-112.

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

Enders, W. and C.W.J. Granger (1998), "Unit-Root Tests and Asymmetric Adjustment with an Example Using the Term Structure of Interest Rates," *Journal of Business & Economic Statistics*, 16 (3), 304-311.

Enders, W., K. Im, J. Lee, and M.C. Strazicich (2010), "IV Threshold Cointegration Tests and the Taylor Rule," *Economic Modeling* 27, 1463-1472.

Enders, W. and P.L. Siklos (2001), "Cointegration and Threshold Adjustment," *Journal of Business & Economic Statistics*, 19 (2), 166-176.

Engle, R.F. and C.W.J. Granger (1987), "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251-276.

Ericsson, N.R. and J.G. MacKinnon (2002), "Distributions of Error Correction Tests for Cointegration," *Econometrics Journal*, 5, 285-318.

Hall, P. and C.C. Heyde (1980), "Martingale Limit Theory and Its Applications," Academic, New York.

Hansen, B. (1995), "Rethinking the Univariate Approach to Unit Root Tests: How to Use Covariates to Increase Power," *Econometric Theory*, 11, 1148-1171.

Harbo, I., S. Johansen, B. Nielsen, and A. Behbek (1998), "Asymptotic Inference on Cointegrating Rank in Partial Systems," *Journal of Business and Economic Statistics*, 16 (4), 388-399.

Haug, A.A. (1993), "Residual Based Tests for Cointegration: A Monte Carlo Study of Size Distortion," *Economics Letters*, 41, 89-115.

Haug, A.A. (1996), "Tests for Cointegration: A Monte Carlo Comparison," *Journal of Econometrics*, 71, 89-115.

Im, K., J. Lee and M.A. Tieslau (2011), "More Powerful Unit Root Tests with Non-normal Errors," manuscript.

Im, K. and P. Schmidt (2008), "More Efficient Estimation under Non-normality When Higher Moments Do Not Depend on the Regressors, Using Residual-augmented Least Squares," *Journal of Econometrics*, 144, 219-233.

Kapetanios, G., Y. Shin, and A. Snell, (2003), "Testing for a Unit Root in the Nonlinear STAR Framework," *Journal of Econometrics*, 112, 359-379.

Kim, K. and P. Schmidt (1993), "Unit Root Tests with Conditional Heteroskedasticity, " *Journal of Econometrics*, Elsevier, 59, 287-300.

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.

Lee, H. (2011), "Performance of the ECM Based Cointegration Tests under the Null in the Presence of a Nuisance Parameter," unpublished manuscript.

Li, J. and J. Lee (2010), "ADL tests for Threshold Cointegration," *Journal of Time Series Analysis*, 31 (4), 241-254.

Park, J.Y. and M. Shintani (2005), "Testing for a Unit Root against Transitional Autoregressive Models," Working Paper, Department of Economics, Vanderbilt University.

Park, J.Y. and P.C.B. Phillips (1999), "Asymptotics for Nonlinear Transformations of Integrated Time Series," *Econometric Theory*, Cambridge University Press, 15, 269-298.

Pesavento, E. (2004), “Analytical Evaluation of the Power of Tests for the Absence of Cointegration”, *Journal of Econometrics*, 122 (2), 349-384.

Phillips, P.C.B. and J.Y. Park (1988), “Asymptotic Equivalence of Ordinary Least Squares and Generalized Least Squares in Regressions with Integrated Regressors,” *Journal of the American Statistical Association*, 83, 111-115.

Phillips, P.C.B. and P. Perron (1988), “Testing for a Unit Root in Time Series Regression,” *Biometrika*, 75, 335-346.

So, B. and D.W. Shin (2001), “An Invariant Sign Test for Random Walks Based on Recursive Median Adjustment,” *Journal of Econometrics*, 102, 197-229.

Zivot, E. (2000), “The Power of Single Equation Tests for Cointegration When the Cointegrating Vector is Prespecified,” *Econometric Theory*, 16, 407-439.