

ABSTRACT

Sen, Kapildeb. Unit Root Tests in Time Series and Stochastic Volatility Models (Under the direction of Dr Sastry G. Pantula).

Providing appropriate forecasts of time series data into the future depends crucially on whether the time series under consideration is non-stationary (i.e. has a unit root) or stationary. In the context of a Stochastic Volatility Model (SVM), the presence of a unit root in financial data has important implications for the pricing of various financial instruments. We propose a unit root test for the volatility process based on the Simulation-Extrapolation (SIMEX) approach. We express the SVM as a measurement error model and propose a Simulation-Extrapolation (SIMEX)-based approach to test for the unit root hypothesis. The asymptotic theory of the Ordinary Least Squares (OLS) and Weighted Symmetric (WS) estimators are exploited to obtain SIMEX-based tests and simulation studies are provided to demonstrate that the SIMEX-based test compares favorably with some of the well known unit root tests already available in the literature.

We also propose a unit root test based on the maximum order statistic in a simple autoregressive (AR) model of order 1. The asymptotic distribution of the test statistic under the null hypothesis is derived and the approximate percentiles are also provided. Through simulation studies, the proposed test is compared with the Dickey-Fuller (DF) test under various specifications for the error distributions.

In the final chapter of this dissertation, we propose a procedure to test the null hypothesis of stationarity in AR (1) models. The procedure is based on the Intersection-Union tests used in Bio-Equivalence studies. The performance of the test based on finite sample percentiles as well as asymptotic percentiles is assessed using simulation studies.

**UNIT ROOT TESTS IN TIME SERIES AND STOCHASTIC VOLATILITY
MODELS**

by

KAPILDEB SEN

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

STATISTICS

Raleigh,
2002

APPROVED BY:

Sastry G. Pantula
(Chair of Advisory Committee)

Leonard A. Stefanski

David A. Dickey

Atsushi Inoue

TO PRAMITA SEN & BASUDEB SEN

BIOGRAPHY

Kapildeb Sen was born to Basudeb Sen and Pramita Sen on November 21, 1976 in Calcutta, India. He moved along with his family to Bombay, the financial capital of India in 1983.

Kapil completed the B.Sc. degree in Statistics from St. Xavier's College, Bombay in 1997. Before joining the Doctoral program in the department of Statistics at North Carolina State University, he completed a Master's degree in Applied Statistics and Informatics from the Indian Institute of Technology (I.I.T.), Bombay in 1999.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my deepest gratitude to my professors at NCSU. In particular, I am indebted to Professor Sastry Pantula for providing me with every opportunity to broaden, develop and enhance my knowledge and skills throughout my academic pursuit at NCSU. I consider myself extremely fortunate to have been a beneficiary of his innovative research ideas. Working under the guidance of Prof. Pantula has been one of my most memorable experiences.

I am thankful to Prof. Leonard Stefanski for his invaluable suggestions and constructive critiques on my dissertation proposal and seminar presentations. His insights were of immense help in charting the course of this dissertation. Prof. David Dickey's repository of amazing SAS programming techniques was instrumental in the successful completion of the simulation studies in this document. Learning about Dickey-Fuller tests from Prof. Dickey himself was a treat that I will always cherish. His teaching and encouraging words have been a constant source of inspiration for me. I would also like to thank Prof. Atsushi Inoue for expressing interest in my research work and providing me with valuable feedback on my draft that resulted in a better presentation of my work. I am grateful to Janice Gaddy for assisting me with an uncountable number of important tasks.

Special thanks are due to my elder brother Sowmdeb Sen and my granny Maya Chakravarty for encouragement and support in various forms. Most importantly, I am thankful to God for my wonderful parents, especially my mother Pramita Sen, who uncomplainingly undertook 24-hour long exhausting journeys to rejuvenate me with her presence whenever I needed her.

CONTENTS

LIST OF TABLES.....	VI
LIST OF FIGURES.....	X
CHAPTER 1 SIMEX-BASED UNIT ROOT TEST FOR STOCHASTIC VOLATILITY MODELS .	1
SECTION 1. INTRODUCTION	1
SECTION 2. STANDARD UNIT ROOT TESTS FOR VOLATILITY	5
SECTION 3. APPLYING THE SIMEX METHOD TO TEST THE UNIT ROOT HYPOTHESIS	9
SECTION 4. PERFORMANCE OF SIMEX-BASED TEST USING A DIFFERENT MECHANISM TO GENERATE PSEUDO-ERRORS.....	23
SECTION 5. COMPARING SIMEX, ADF AND INSTRUMENTAL VARIABLES APPROACH	29
SECTION 6. SIMEX APPLIED TO THE IV ESTIMATOR	32
SECTION 7. A SIMEX TEST BASED ON P-VALUES.....	35
SECTION 8. EXTRAPOLATING FUNCTION BASED ON ASYMPTOTIC THEORY	39
SECTION 9. PERFORMANCE OF SIMEX-BASED TEST WHEN σ_{η}^2 IS ESTIMATED.....	50
SECTION 10. COMBINING THE TWO PROMISING EXTRAPOLATING FUNCTIONS.....	55
SECTION 11. CONCLUSION	61
CHAPTER 2 A UNIT ROOT TEST BASED ON THE MAXIMUM ORDER STATISTIC	62
SECTION 1. INTRODUCTION	62
SECTION 2. THE TEST STATISTIC.....	63
SECTION 3. FINITE SAMPLE AND ASYMPTOTIC PERCENTILES	64
SECTION 4. COMPARATIVE POWER STUDY	72
SECTION 5. CONCLUSIONS.....	84
CHAPTER 3 TESTING THE NULL HYPOTHESIS OF STATIONARITY USING INTERSECTION-UNION PRINCIPLE	85
SECTION 1. INTRODUCTION	85
SECTION 2. INTERSECTION UNION TEST FOR NEAR UNIT ROOT HYPOTHESIS.....	87
SECTION 3. ASYMPTOTIC THEORY	91
SECTION 4. MONTE CARLO STUDY	93
SECTION 5. A DIFFERENT SEQUENCE OF HYPOTHESIS.....	107
SECTION 6. CONCLUSIONS.....	112
BIBLIOGRAPHY	113

List of Tables

Table 1-1: $P(\hat{\tau}_{OLS} < -2.90 \phi = \phi_0)$ and $P(\hat{\tau}_{WS} < -2.55 \phi = \phi_0)$, $n = 100$	9
Table 1-2: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 \phi = \phi_0)$ when $\sigma_\eta^2 = 1$, $n = 100$	18
Table 1-3: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 \phi = \phi_0)$ when $\sigma_\eta^2 = 1.5$, $n = 100$	18
Table 1-4: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 \phi = \phi_0)$ when $\sigma_\eta^2 = 2$, $n = 100$	18
Table 1-5: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 \phi = \phi_0)$ when $\sigma_\eta^2 = 10$, $n = 100$	19
Table 1-6: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 \phi = \phi_0)$ when $\sigma_\eta^2 = 0.5$, $n = 100$	20
Table 1-7: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 \phi = \phi_0)$ when $\sigma_\eta^2 = 0.1$, $n = 100$	20
Table 1-8: $P(\hat{\tau}_{SIMEX,WS} < -2.55 \phi = \phi_0)$ when $\sigma_\eta^2 = 1$, $n = 100$	21
Table 1-9: $P(\hat{\tau}_{SIMEX,WS} < -2.55 \phi = \phi_0)$ when $\sigma_\eta^2 = 1.5$, $n = 100$	21
Table 1-10: $P(\hat{\tau}_{SIMEX,WS} < -2.55 \phi = \phi_0)$ when $\sigma_\eta^2 = 2$, $n = 100$	21
Table 1-11: $P(\hat{\tau}_{SIMEX,WS} < -2.55 \phi = \phi_0)$ when $\sigma_\eta^2 = 10$, $n = 100$	22
Table 1-12: $P(\hat{\tau}_{SIMEX,WS} < -2.55 \phi = \phi_0)$ when $\sigma_\eta^2 = 0.5$, $n = 100$	22
Table 1-13: $P(\hat{\tau}_{SIMEX,WS} < -2.55 \phi = \phi_0)$ when $\sigma_\eta^2 = 0.1$, $n = 100$	23
Table 1-14: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 \phi = \phi_0)$ when $\sigma_\eta^2 = 1$, $n = 100$	24
Table 1-15: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 \phi = \phi_0)$ when $\sigma_\eta^2 = 1.5$, $n = 100$	25
Table 1-16: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 \phi = \phi_0)$ when $\sigma_\eta^2 = 2$, $n = 100$	25
Table 1-17: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 \phi = \phi_0)$ when $\sigma_\eta^2 = 10$, $n = 100$	25
Table 1-18: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 \phi = \phi_0)$ when $\sigma_\eta^2 = 0.5$, $n = 100$	26
Table 1-19: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 \phi = \phi_0)$ when $\sigma_\eta^2 = 0.1$, $n = 100$	26
Table 1-20: $P(\hat{\tau}_{SIMEX,WS} < -2.55 \phi = \phi_0)$ when $\sigma_\eta^2 = 1$, $n = 100$	27
Table 1-21: $P(\hat{\tau}_{SIMEX,WS} < -2.55 \phi = \phi_0)$ when $\sigma_\eta^2 = 1.5$, $n = 100$	27
Table 1-22: $P(\hat{\tau}_{SIMEX,WS} < -2.55 \phi = \phi_0)$ when $\sigma_\eta^2 = 2$, $n = 100$	28
Table 1-23: $P(\hat{\tau}_{SIMEX,WS} < -2.55 \phi = \phi_0)$ when $\sigma_\eta^2 = 10$, $n = 100$	28
Table 1-24: $P(\hat{\tau}_{SIMEX,WS} < -2.55 \phi = \phi_0)$ when $\sigma_\eta^2 = 0.5$, $n = 100$	28
Table 1-25: $P(\hat{\tau}_{SIMEX,WS} < -2.55 \phi = \phi_0)$ when $\sigma_\eta^2 = 0.1$, $n = 100$	29

Table 1-26: Comparing ADF, IV and SIMEX-based unit root test when $\sigma_{\eta}^2 = 1$, $n = 100$	32
Table 1-27: SIMEX applied to the IV estimator, $n = 100$	35
Table 1-28: Rejection frequencies when SIMEX is applied to p-values (λ - approach), $n = 100$	37
Table 1-29: Rejection frequencies when SIMEX is applied to p-values ($Non - \lambda$ - approach), $n = 100$	37
Table 1-30: Comparison of ADF, IV and SIMEX-based tests for different values of σ_{η}^2 , $n = 100$	42
Table 1-31: Relationship between σ_{η}^2 and δ	43
Table 1-32: Power of SIMEX-based test using $a + b\delta^c$ as the extrapolating function ($\sigma_{\eta}^2 = 0.1, 0.2, 0.3$), $n = 100$	47
Table 1-33: Power of SIMEX-based test using $a + b\delta^c$ as the extrapolating function ($\sigma_{\eta}^2 = 0.4, 0.5, 0.6$), $n = 100$	47
Table 1-34: Power of SIMEX-based test using $a + b\delta^c$ as the extrapolating function ($\sigma_{\eta}^2 = 0.7, 0.8, 0.9$), $n = 100$	47
Table 1-35: Power of SIMEX-based test using $a + b\delta^c$ as the extrapolating function ($\sigma_{\eta}^2 = 1, 2, 3$), $n = 100$	48
Table 1-36: Power of SIMEX-based test using $a + b\delta^c$ as the extrapolating function ($\sigma_{\eta}^2 = 4, 5, 10$), $n = 100$	48
Table 1-37: Power of SIMEX-based test using $a + b\delta^c$ as the extrapolating function ($\sigma_{\eta}^2 = 50, 100$), $n = 100$	48
Table 1-38: Performance of $a + b\delta^{1.2}$ for different values of $\sigma_{\eta}^2 \geq 1$, $n = 100$	49
Table 1-39: Performance of $a + b\delta^{1.2}$ for different values of $\sigma_{\eta}^2 < 1$, $n = 100$	50
Table 1-40: Power of SIMEX-based test using $a + b\hat{\delta}_{\lambda}^c$ as extrapolating function ($\sigma_{\eta}^2 = 0.1, 0.2, 0.3$), $n = 100$	51
Table 1-41: Power of SIMEX-based test using $a + b\hat{\delta}_{\lambda}^c$ as extrapolating function ($\sigma_{\eta}^2 = 0.4, 0.5, 0.6$), $n = 100$	52
Table 1-42: Power of SIMEX-based test using $a + b\hat{\delta}_{\lambda}^c$ as extrapolating function ($\sigma_{\eta}^2 = 0.7, 0.8, 0.9, 1$), $n = 100$	52
Table 1-43: Power of SIMEX-based test using $a + b\hat{\delta}_{\lambda}^c$ as extrapolating function ($\sigma_{\eta}^2 = 2, 3$), $n = 100$	53

Table 1-44: Power of SIMEX-based test using $a + b\hat{\delta}_\lambda^c$ as extrapolating function $(\sigma_\eta^2 = 4, 5, 8), n = 100$	53
Table 1-45: Power of SIMEX-based test using $a + b\hat{\delta}_\lambda^c$ as extrapolating function $(\sigma_\eta^2 = 10, 100), n = 100$	53
Table 1-46: Comparing $a + b\hat{\delta}_\lambda$ and $a + b\hat{\delta}_\lambda^{1.2}$ as extrapolating functions, $n = 100$	54
Table 1-47: Different rules for combining the extrapolating functions $a + b\hat{\delta}_\lambda$ and $a + b\hat{\delta}_\lambda^{1.2}$	56
Table 1-48: Estimated size of the SIMEX test using different rules, $n = 100$	58
Table 1-49: Estimated power of the SIMEX test using different rules at $H_1: \phi = 0.7$, $n = 100$	59
Table 1-50: Power and Size of various rules when Sample Size $n = 1000$ and $0.01 \leq \sigma_\eta^2 \leq 0.10$	60
Table 2-1: Percentiles from the Simulated distribution of τ	64
Table 2-2: Power Comparison for $n = 25$	65
Table 2-3: Power Comparison for $n = 100$	65
Table 2-4: Power Comparison for $n = 1000$	65
Table 2-5: Asymptotic Percentiles of τ	72
Table 2-6: Size and Power based on Finite Sample Percentiles under various specifications for the Error distribution	78
Table 2-7: Size and Power based on Asymptotic Percentiles under various specifications for the Error distribution	79
Table 2-8: Size and Power based on Finite Sample Percentiles under centered Skew- Normal, Skew-t, Uniform and Exponential distributions.	80
Table 2-9: Size and Power based on Asymptotic Percentiles under centered Skew- Normal, Skew-t, Uniform and Exponential distributions.	81
Table 2-10: Rejection Probabilities when $e_t \sim \{Exponential(1)-1\}$ or $Uniform(-0.5, 0.5)$, $n = 25$ based on asymptotic percentiles.....	82
Table 2-11: Rejection Probabilities when $e_t \sim \{Exponential(1)-1\}$ or $Uniform(-0.5, 0.5)$, based on finite sample ($n = 25$) percentiles.....	82
Table 2-12: Rejection Probabilities when $e_t \sim \{Exponential(1)-1\}$ or $Uniform(-0.5, 0.5)$, $n = 50$ based on asymptotic percentiles.....	82
Table 2-13: Rejection Probabilities when $e_t \sim \{Exponential(1)-1\}$ or $Uniform(-0.5, 0.5)$, based on finite sample ($n = 50$) percentiles.....	83
Table 2-14: Rejection probabilities when the error distribution is given by the cone- shaped distribution (2.5.1) and the sample size $n = 50$	84
Table 2-15: Rejection probabilities when the error distribution is given by the cone- shaped distribution (2.5.1) and the sample size $n = 25$	84
Table 3-1: 95 th percentile based on OLS statistic $\tau_{OLS,n}^{(2)}$	93
Table 3-2: 5 th percentile based on OLS statistic $\tau_{OLS,n}^{(2)}$	94

Table 3-3: 95 th percentile based on OLS statistic $\tau_{WS,n}^{(1)}$	94
Table 3-4: 5 th percentile based on OLS statistic $\tau_{WS,n}^{(2)}$	95
Table 3-5: Size and Power of the IUT based on OLS, $n = 100$	97
Table 3-6: Size and Power of IUT based on OLS, $n = 500$	97
Table 3-7: Size and Power based of IUT based on OLS, $n = 1000$	97
Table 3-8: Size and Power based on WS statistic, $n = 100$	98
Table 3-9: Size and Power based on WS statistic, $n = 500$	99
Table 3-10: Size and Power based on WS statistic, $n = 1000$	99
Table 3-11: Asymptotic Percentiles corresponding to various values of γ	101
Table 3-12: Rejection probabilities under the null and alternative hypothesis based on asymptotic percentiles.....	102
Table 3-13: Size and Power of TS1 for seven different DC and three different sample sizes.....	105
Table 3-14: Size and Power of TS2 for seven different DC and three different sample sizes.....	105
Table 3-15: Size and Power of TS3 for seven different DC and three different sample sizes.....	106
Table 3-16: Size and Power of TS4 for seven different DC and three different sample sizes.....	106
Table 3-17: Size and Power of TS5 for seven different DC and three different sample sizes.....	106
Table 3-18: 95 th percentile based on the distribution of $\tau_{OLS,\sqrt{n}}^{(1)}$	108
Table 3-19: 5 th percentile based on the distribution of $\tau_{OLS,\sqrt{n}}^{(2)}$	109
Table 3-20: 95 th percentile based on the distribution of $\tau_{ws,\sqrt{n}}^{(1)}$	109
Table 3-21: 5 th percentile based on the distribution of $\tau_{ws,\sqrt{n}}^{(2)}$	109
Table 3-22: Size and Power of the IUT based on $\tau_{OLS,\sqrt{n}}^{(1)}$ and $\tau_{OLS,n}^{(2)}$, $n = 100$	111
Table 3-23: Size and Power of the IUT based on $\tau_{OLS,\sqrt{n}}^{(1)}$ and $\tau_{OLS,n}^{(2)}$, $n = 500$	111
Table 3-24: Size and Power of the IUT based on $\tau_{OLS,\sqrt{n}}^{(1)}$ and $\tau_{OLS,n}^{(2)}$, $n = 1000$	111

List of Figures

Figure 1-1: Linear extrapolating function $a + b\lambda$	15
Figure 1-2: A quadratic extrapolating function $a + b\lambda + c\lambda^2$	16
Figure 1-3: The extrapolating function $a + b/(\lambda + 2)$	16
Figure 1-4: (a) Sample Plot of $Z_{b,IV}(\lambda)$ against λ	33
Figure 1-5: (b) Sample Plot of $Z_{b,IV}(\lambda)$ against λ	33
Figure 1-6: (a) Plot of p-values against k : # of measurement errors	36
Figure 1-7: (b) Plot of p-values against k : # of measurement errors	38
Figure 1-8: Plot of $\hat{\tau}_{OLS}$ against λ	45
Figure 2-1:Skew-Normal Density Curve with $\lambda = 0.5$	73
Figure 2-2: Skew-Normal Density with $\lambda = 1.5$	74
Figure 2-3: Skew t-distribution with 10 d.f. and $\lambda = 1$	74
Figure 2-4: Skew-t distribution with 10 d.f. and $\lambda = 0.5$	75
Figure 2-5: Skew-t distribution with 10 d.f. and $\lambda = 1.5$	75
Figure 2-6: Density Curve corresponding to (2.5.1).....	83

Chapter 1 SIMEX-BASED UNIT ROOT TEST FOR STOCHASTIC VOLATILITY MODELS

Section 1. Introduction

Modeling the volatility of the price of an asset is central to the determination of the price of various financial instruments. Engle (1982) proposed the Autoregressive Conditional Heteroskedasticity (ARCH) models to analyze economic times series. In the ARCH models, the conditional variance of the white noise process is allowed to change over time. This aspect of time-dependent conditional variance is the consequence of modeling the square of the white noise error process as an Autoregressive (AR) process. The lack of parsimonious ARCH models led to the development of generalized autoregressive conditionally heteroscedastic (GARCH) models. Bollerslev (1986) proposed the GARCH models that model the square of the white noise process as an Autoregressive Moving Average (ARMA) process.

Stochastic Volatility Model (SVM) serves as an alternative to ARCH and GARCH models. Taylor (1982) introduced the first discrete-time SVM in which the observed time series is expressed in terms of an unobserved variance component and an independent and identically distributed (i.i.d.) error.

Let r_t be an observed time series (e.g. stock returns). Then we consider the following SVM for r_t :

$$\begin{aligned} r_t &= u_t \exp(h_t/2), \\ (h_t - \mu) &= \phi(h_{t-1} - \mu) + \eta_t, \quad t = 1, \dots, n, \end{aligned} \tag{1.1.1}$$

where u_t and η_t are zero-mean white noise with variance 1 and σ_η^2 , respectively. Also, u_t and η_t are assumed to be stochastically independent.

Equation (1.1.1) shows that the unobserved variance component h_t has been modeled as an autoregressive process with mean μ and AR parameter ϕ .

In equation (1.1.1), if $\phi=1$ then the SVM is non-stationary (unit root process) and volatility is said to be persistent. On the other hand, if the magnitude of ϕ is less than one, the volatility is said to be transient in nature implying that the effect of a shock on the future volatility will eventually die out.

The persistence of volatility is a very important topic, which has been investigated by Poterba and Summers (1986), Chou (1988) and Poon and Taylor (1992). The SVM specified by (1.1.1) implies that the logarithm of the squared time series r_t is an ARMA process. This ARMA process has the same autoregressive root as that of the volatility process, h_t . It is thus possible to test for a unit root in the unobserved variance component by testing for a unit root in the logarithm of the squared time series. Poterba and Summers (1986) examined the persistence of volatility over the 1928-84 period as well as the shorter postwar period (post-1950) studied by Pindyck (1984). They computed the volatility estimates from daily returns on the Standard and Poor's (S&P) Composite Index. Following Dickey and Fuller (1981) and Fuller (1976), they tested the hypothesis that measured volatility is a nonstationary series. For both sample periods, the null hypothesis of nonstationarity was rejected in favor of the alternative hypothesis of stationarity. Chou (1988) investigated the issue of volatility persistence in the stock market using the GARCH models. They analyzed weekly returns of the New York Stock Exchange (NYSE) value-weighted index and could not reject the unit-root hypothesis. Poon and Taylor (1992), again using GARCH models could not reject the hypothesis of the presence of a unit root in the volatility in U.K. stock markets. Harvey, Ruiz and Shepherd (1994), using the augmented Dickey-Fuller tests, rejected the null hypothesis of unit-root in each of the four series of exchange rates they considered.

So, Lam and Li (1997) addressed the issue of persistence of volatility in seven Southeast Asian markets, by the SVM defined in (1.1.1). They concluded that the shocks to volatility are transient in the whole period (1980-91) in all the seven Southeast Asian indices. A Bayesian approach to testing for a unit-root in SVM was proposed by So and Li (1999). They applied their method to the seven indices of So et al. (1997) and the S & P 500 data. Their method leads to slightly different results as compared with the ADF tests. The null hypothesis of nonstationarity is rejected in all cases according to the classical test. Using the posterior odds ratio, they found strong evidence favoring stationary volatility for most of the market indices they considered.

In this paper, we focus on unit root tests for the volatility process based on a frequentist approach. The stochastic volatility model implies that the log of the squared time series, i.e., $\log(r_t^2)$ is an ARMA (1,1) process with the autoregressive root being the same as the autoregressive root of the volatility process h_t . One may test for a unit root in the unobserved volatility process h_t by testing for a unit root in the log of the squared time series $\log(r_t^2)$. Unfortunately, for small values of σ_η^2 , the MA parameter in $\log(r_t^2)$ process has a large negative moving average root and the standard unit root tests are known to suffer from extreme size distortions in the presence of negative MA roots as shown in Pantula (1991). Because the standard unit root tests reject the null hypothesis of a unit root too often when the moving average parameter is very close to one, they are unreliable. In this paper, our goal is to investigate whether the procedure of testing for a unit root in the $\log(r_t^2)$ process based on ordinary least squares (OLS), weighted symmetric estimators (WS) and instrumental variables (IV) can be modified to obtain test procedures which maintain the correct level of significance and still deliver enough power to differentiate between the stationary and non-stationary series. Towards that end, we view model (1.1.1) as a model with measurement error. On taking logarithms on both sides of the representation of r_t^2 in (1.1.1), we get $\log(r_t^2) = h_t + \log(u_t^2)$. Note that the observed data $\log(r_t^2)$ has been expressed as a sum of the unobserved true value h_t and

the error $\log(u_t^2)$ (in the measurement of $\log(r_t^2)$). The regression model is given by the second equation in (1.1.1), $(h_t - \mu) = \phi(h_{t-1} - \mu) + \eta_t$.

Cook and Stefanski (1994) developed the Simulation–Extrapolation (SIMEX) procedure in response to the need for fitting nonstandard, generalized linear measurement error models. It is a simulation-based method of estimating and reducing bias due to measurement error. In this method, estimates are obtained by adding additional measurement error to the data, identifying a trend in the measurement error-induced statistic and extrapolating this trend back to the case of no measurement error. In this article, we investigate SIMEX-based procedures for testing $H_0: \phi = 1$ against $H_1: \phi < 1$.

The remainder of the article is organized into seven sections. The next section discusses the application of the standard unit root tests to the volatility process and presents two estimation methods used by several of these tests. Section 3 describes the SIMEX procedure and how it can be used to derive a unit root test for the SVM. Section 4 presents an alternative way to generate the pseudo-errors in the SIMEX procedure. Section 5 presents the IV-based unit root test and the ADF test and compares their performance with the SIMEX-based test. Section 6 explores the applicability of the SIMEX method to the IV estimator. In Section 7, an attempt is made to apply the SIMEX procedure to the p-values instead of the test statistic values. Section 8 draws insight from asymptotic theory of OLS and WS estimators in ARMA models to specify an extrapolating function that would be valid irrespective of the value of σ_η^2 . In practice, the value of σ_η^2 will seldom be known and so Section 9 attempts to validate the SIMEX-based test in situations that require the estimation of σ_η^2 . To optimize the size and power performance of the SIMEX-based test, the simulation results of Section 9 suggest the use of different extrapolating functions depending on whether σ_η^2 is less than 1 or not. Section 10 seeks to address this issue by evaluating different rules of selecting an extrapolating function based on an estimate of σ_η^2 .

Section 2. Standard unit root tests for volatility

Note that on taking logarithm on both sides of the first equation in (1.1.1), we get, $\log(r_t^2) = h_t + \log(u_t^2)$. It follows that $\log(r_t^2)$ has the same covariance structure as that of a stationary ARMA(1, 1) process,

$$(\log(r_t^2) - \mu) = \phi(\log(r_{t-1}^2) - \mu) + \varepsilon_t - \theta\varepsilon_{t-1}, |\phi| < 1. \quad (1.2.1)$$

But if $\phi = 1$, $\log(r_t^2)$ is an ARIMA(0,1,1) process,

$$(\log(r_t^2) - \mu) = (\log(r_{t-1}^2) - \mu) + \varepsilon_t - \theta\varepsilon_{t-1}, \quad (1.2.2)$$

where the ε_t 's are white noise with variance σ^2 . If we let $e_t = \log(u_t^2)$, then the mean and the variance of e_t are known to be approximately -1.2704 and $\pi^2/2$ respectively; see Abramovitz & Stegun (1970, Pg 943). We can compute the values of the parameters θ and σ^2 in the model (1.2.2) by solving the equations

$$\frac{\theta}{1+\theta^2} = \frac{-\sigma_e^2}{2\sigma_e^2 + \sigma_\eta^2} \quad \text{and} \quad \theta\sigma^2 = \sigma_e^2. \quad (1.2.3)$$

Based on the ARMA representation in (1.2.1), we may use any one of the unit root tests available in the literature to test for a unit root in $\log(r_t^2)$ and hence to test for a unit root in the volatility process, h_t . Said and Dickey (1984) suggested a test that extended the results of Dickey and Fuller (1979) to ARMA models of unknown order. They approximated an ARMA process by a long AR process and used the test criteria suggested by Dickey and Fuller (1979). This approach is known as the Augmented Dickey-Fuller approach. Said and Dickey (1985) proposed a test based on the one-step Gauss-Newton procedure. Phillips (1987) and Phillips and Perron (1988) proposed tests based on non-parametric correction of the Dickey-Fuller type test statistics. Hall (1989) proposed a new approach to testing for unit roots in time series with moving average innovations based on an instrumental variable estimator.

We now present two pivotal test statistics associated with two different kinds of estimators that exist in the literature for testing the null hypothesis $H_0 : \phi = 1$, against the alternative hypothesis $H_a : \phi < 1$ in the AR model

$$Y_t - \mu = \phi (Y_{t-1} - \mu) + \varepsilon_t, Y_0 = 0, \varepsilon_t \sim NID(0, \sigma^2). \quad (1.2.4)$$

The following two statistics are not appropriate for testing the hypothesis of nonstationarity in the $\log(r_t^2)$ process because $\log(r_t^2)$ follows an ARMA structure. But we present them to serve two purposes. The first purpose is to illustrate the size distortion of such tests when the ARMA(1,1) model is incorrectly specified to be a first-order AR process. The second purpose is to show, in a later section, how the SIMEX method applied to these statistics, controls the size distortion.

1. *The ordinary least squares (OLS) statistic $\hat{\tau}_{OLS}$*

It follows from equation (1.2.4) that $Y_t = \mu(1-\phi) + \phi Y_{t-1} + \varepsilon_t = \theta_0 + \phi Y_{t-1} + \varepsilon_t$ where $\theta_0 = \mu(1-\phi)$. The regression t statistic for testing that the coefficient of Y_{t-1} is 1 in the regression of Y_t on '1' and Y_{t-1} is,

$$\hat{\tau}_{OLS} = s_1^{-1} \left(\sum_{t=2}^n y_{t-1}^2 \right)^{1/2} (\hat{\phi} - 1), \quad (1.2.5)$$

where

$$s_1^2 = (n-2)^{-1} \sum_{t=2}^n (y_t - \hat{\phi} y_{t-1})^2$$

and

$$\hat{\phi} = \left(\sum_{t=2}^n y_{t-1}^2 \right)^{-1} \sum_{t=2}^n y_{t-1} y_t$$

and

$$y_t = Y_t - \bar{y}, \quad \bar{y} = \frac{1}{n} \sum_{t=1}^n Y_t$$

2. The weighted symmetric (WS) statistic $\hat{\tau}_{ws}$

The WS estimator is a member of the class of symmetric estimators—the symmetry referring to the fact that if a normal stationary AR process satisfies (1.2.4), then it also satisfies the equation $Y_t - \mu = \phi(Y_{t+1} - \mu) + \varepsilon_t^*$, $\varepsilon_t^* \sim NID(0, \sigma^2)$. The WS estimator of ϕ , given by

$$\hat{\phi}_{ws} = \frac{\sum_{t=2}^n y_{t-1} y_t}{\sum_{t=2}^{n-1} y_t^2 + n^{-1} \sum_{t=1}^n y_t^2}, \quad (1.2.6)$$

was studied by Park and Fuller (1993). It is obtained by minimizing

$$Q_w(\phi) = \sum_{t=2}^n w_t (y_t - \phi y_{t-1})^2 + \sum_{t=1}^{n-1} (1 - w_{t+1}) (y_t - \phi y_{t+1})^2, \text{ where } w_t = n^{-1}(t-1), \quad t = 2, 3, \dots, n,$$

$$y_t = Y_t - \bar{y} \text{ and } \bar{y} = n^{-1} \sum_{t=1}^n Y_t.$$

The pivotal statistic corresponding to $\hat{\phi}_{ws}$ is given by

$$\hat{\tau}_{ws} = \left[\hat{\phi}_{ws} - 1 \right] \left[\sum_{t=2}^{n-1} y_t^2 + n^{-1} \sum_{t=1}^n y_t^2 \right]^{1/2} \hat{\sigma}_{ws}^{-1}, \quad (1.2.7)$$

where

$$\hat{\sigma}_{ws}^2 = (n-2)^{-1} Q_w(\hat{\phi}_{ws}) \quad (1.2.8)$$

Monte Carlo study

In this section we present the results of a simulation study of the test statistics presented in the previous section. Our model is

$$\begin{aligned}\log(r_t^2) &= h_t + \log(u_t^2) \\ (h_t - \mu) &= \phi(h_{t-1} - \mu) + \eta_t, \quad t = 1, \dots, n,\end{aligned}\tag{1.2.9}$$

where u_t and η_t are zero-mean white noise with variance 1 and σ_η^2 , respectively. Also, u_t and η_t are assumed to be stochastically independent.

We let $h_0 = 0$ and $\mu = 0$. The RANNOR function in SAS (V8) was used to generate the η_t 's as independent standard normal variables ($\sigma_\eta^2 = 1$). The same function was used again to generate the u_t 's as independent standard normal variables independent of the η_t 's. We generated samples of size $n=100$ with $\phi = 0.8, 0.85, 0.9, 0.95, 0.98$. The two statistics $\hat{\tau}_{OLS}$ and $\hat{\tau}_{ws}$ were then computed by plugging in $Y_t = \log(r_t^2)$ in each of the expressions (1.2.5) and (1.2.7).

The finite sample ($n=100$) critical values (for a 5%-level test of $H_0 : \phi = 1$ against $H_a : \phi < 1$ when the true process follows the AR model (1.2.4), are computed to be -2.90 and -2.55 (Tables 10.A.2 and 10.A.4 in Fuller (1996)) for $\hat{\tau}_{OLS}$ and $\hat{\tau}_{ws}$ respectively. These are the critical values that we used to test $H_0 : \phi = 1$ against $H_a : \phi < 1$ in the ARMA model for $\log(r_t^2)$. So, for the OLS statistic, the null hypothesis $H_0 : \phi = 1$ was rejected when $\hat{\tau}_{OLS} \leq -2.90$ and for the WS statistic, the null hypothesis was rejected when $\hat{\tau}_{ws} \leq -2.55$. The powers of the tests were computed on the basis of 5000 Monte Carlo replications and are reported in the following table.

Table 1-1: $P(\hat{\tau}_{OLS} < -2.90 | \phi = \phi_0)$ and $P(\hat{\tau}_{WS} < -2.55 | \phi = \phi_0)$, $n = 100$

Test Statistic	ϕ_0					
	1	0.98	0.95	0.90	0.85	0.80
$\hat{\tau}_{OLS}$	0.76	0.91	0.97	0.99	1.00	1.00
$\hat{\tau}_{WS}$	0.80	0.95	0.99	0.99	1.00	1.00

As is evident from Table 1, both the OLS and the WS statistics have severe size distortions. More specifically, the WS statistic wrongly rejects the null hypothesis of nonstationarity more often than the OLS statistic does. Hence these tests are not reliable even though they produce high power at values of ϕ very close to 1.

For the above simulations, $\sigma_\eta^2 = 1$ and $\sigma_e^2 = \pi^2/2$. Using equations (1.2.3), we get $\theta = -0.6399$ (we use the value of θ that is less than 1 in absolute value so that we have an invertible process) and $\sigma^2 = 7.71183$. As shown in Pantula (1991), such size distortions become worse as the moving average parameter θ in the ARMA model (1.2.2) becomes closer to 1.

Section 3. Applying the SIMEX method to test the unit root hypothesis

The SIMEX method was introduced by Cook and Stefanski (1994) as a simulation-based method of inference for measurement error models. If the measurement error variance is unknown, one can still apply the method using an estimate of the measurement error variance computed from replicate measurements. The SIMEX estimate is obtained by adding additional measurement error to the data, establishing a trend between these estimates and the variance of the added errors and extrapolating this trend back to the case of no measurement error. Cook and Stefanski (1994) showed that the method is equivalent or asymptotically equivalent to method-of-moments estimation in linear

measurement error modeling. We will first present the SIMEX procedure in a formal setting and then show how our model (1.2.9) can be fitted into the SIMEX framework. We will borrow heavily from Cook and Stefanski (1994) to formally describe the SIMEX method.

Let Y, V, U and X denote the response variable, a covariate measured without error, the true predictor, and the measured predictor. Assume that the true predictor and the measured predictor are related by the following equation:

$$X = U + \sigma Z$$

where Z is a standard normal random variable independent of U, V and X and σ^2 is the known measurement error variance. Suppose that the observed data is given by $\{Y_i, V_i, X_i\}_1^n$.

Let

$$\hat{\tau}_{true} = T(\{Y_i, V_i, U_i\}_1^n)$$

be an estimator of the parameter τ . $\hat{\tau}_{true}$ is not an estimator in the strict sense since it depends on the unknown $\{U_i\}_1^n$. But it is useful to have a notation for the random quantity.

For $\lambda \geq 0$, define

$$X_{b,i}(\lambda) = X_i + \lambda^{1/2} \sigma Z_{b,i},$$

where $\{Z_{b,i}\}_1^n$ are mutually independent, independent of $\{Y_i, V_i, U_i, X_i\}_1^n$ and identically distributed standard normal random variables. Define

$$\hat{\tau}_b(\lambda) = T(\{Y_i, V_i, X_{b,i}(\lambda)\}_1^n)$$

and

$$\hat{\tau}(\lambda) = E(\hat{\tau}_b(\lambda) | \{Y_i, V_i, X_i\}_1^n);$$

The expectation above is with respect to the distribution of $\{Z_{b,i}\}_1^n$ only.

Exact determination of $\hat{\tau}(\lambda)$ for $\lambda > 0$ is generally not feasible, but it can always be estimated arbitrarily well by generating a large number of independent measurement error vectors, $\{\{Z_{b,i}\}_{i=1}^n\}_{b=1}^B$, computing $\hat{\tau}_b(\lambda)$ for $b=1, \dots, B$, and approximating $\hat{\tau}(\lambda)$ by the sample mean of $\{\hat{\tau}_b(\lambda)\}_{b=1}^B$. $\{Z_{b,i}\}_{i=1}^n$ are called pseudo-errors. This is the simulation component of the SIMEX method.

The extrapolation part of the method comprises modeling $\hat{\tau}(\lambda)$ as a function of λ for $\lambda \geq 0$ and using the model to extrapolate back to $\lambda = -1$. This yields the simulation extrapolation estimator, denoted by $\hat{\tau}_{SIMEX}$. Here, $\lambda = -1$ corresponds to no measurement error in the model.

Assuming that $\hat{\tau}_{true}$, $\hat{\tau}_b(\lambda)$ and $\hat{\tau}(\lambda)$ have finite expectations and that a weak law of large number applies, each statistic converges in probability to its expectation. One can regard $E(\hat{\tau}_b(\lambda))$ to be a function of the true estimand τ_0 and the variance of the total measurement error in $X_i(\lambda)$; that is $Var(Z_i + \sigma \sqrt{\lambda} Z_{b,i}) = \sigma^2(1 + \lambda)$. If we denote this function by $T(\tau_0, \sigma^2(1 + \lambda))$, then $\hat{\tau}_b(\lambda)$ converges in probability to $T(\tau_0, \sigma^2(1 + \lambda))$. Also, if $\sigma^2 = 0$ (i.e., if there is no measurement error), then $\hat{\tau}_{true} = \hat{\tau}_b(\lambda) = \hat{\tau}(\lambda)$ (almost surely). So if $\hat{\tau}_{true}$ is a consistent estimator of τ_0 , then $\tau_0 = T(\tau_0, 0)$. Now, provided $T(.,.)$ is a continuous function, we are also led to the conclusion that $\lim_{\lambda \rightarrow -1} E(\hat{\tau}(\lambda)) = \lim_{\lambda \rightarrow -1} E(\hat{\tau}_b(\lambda)) = \lim_{\lambda \rightarrow -1} T(\tau_0, \sigma^2(1 + \lambda)) = T(\tau_0, 0) = \tau_0$.

Thus the extrapolation of the fitted model to $\lambda = -1$, which is $\hat{\tau}_{SIMEX}$, is an *approximately consistent estimator* of τ_0 . The SIMEX estimator is only approximately consistent in general, because the estimated extrapolation function is generally an approximation.

To see how the SIMEX method can be applied to the model (1.1.1), consider

$$\begin{aligned}
\ln r_t^2 &= h_t + \log(u_t^2) \\
&= (h_t + (-1.27)) + (\log(u_t^2) - (-1.27)) \\
&= h_t^* + e_t
\end{aligned}$$

where $E(\ln u_t^2) = -1.27$ and $e_t = \ln u_t^2 + 1.27$ are independent $\log(\chi^2)$ variables with mean zero and variance $\pi^2/2$.

So h_t^* can be viewed as the true variate subject to measurement error, e_t , and $\log(r_t^2)$ can be looked upon as the measurement of h_t^* . Let $Y_t = \log(r_t^2)$. Then Y_t is an ARMA(1,1) process satisfying $Y_t = Y_{t-1} + \varepsilon_t^{(1)} - \theta^{(1)} \varepsilon_{t-1}^{(1)}$, $\varepsilon_t^{(1)}$ being white noise with variance $\sigma_{\varepsilon^{(1)}}^2$.

Define

$$\begin{aligned}
h'_t &= h_t - \bar{h}, \quad \bar{h} = n^{-1} \sum_{t=1}^n h_t \\
\hat{\tau}_{OLS} &= \frac{\hat{\phi}_{OLS} - 1}{\hat{\sigma}_{OLS}} \left(\sum_{t=2}^n h_{t-1}'^2 \right)^{1/2} \\
\hat{\tau}_{ws} &= \left[\hat{\phi}_{ws} - 1 \right] \left[\sum_{t=2}^{n-1} h_t'^2 + n^{-1} \sum_{t=1}^n h_t'^2 \right]^{1/2} \hat{\sigma}_{ws}^{-1}
\end{aligned}$$

where

$$\begin{aligned}
\hat{\phi}_{OLS} &= \left(\sum_{t=2}^n h_{t-1}'^2 \right)^{-1} \sum_{t=2}^n h_{t-1}' h_t' \\
\hat{\sigma}_{OLS}^2 &= (n-2)^{-1} \sum_{t=2}^n (h_t' - \hat{\phi}_{OLS} h_{t-1}')^2 \\
\hat{\phi}_{ws} &= \frac{\sum_{t=2}^n h_{t-1}' h_t'}{\sum_{t=2}^{n-1} h_t'^2 + n^{-1} \sum_{t=1}^n h_t'^2}
\end{aligned}$$

$$\hat{\sigma}_{ws}^2 = (n-2)^{-1} Q_w(\hat{\phi}_{ws})$$

$$Q_w(\phi) = \sum_{t=2}^n w_t (h'_t - \phi h'_{t-1})^2 + \sum_{t=1}^{n-1} (1-w_{t+1})(h'_t - \phi h'_{t+1})^2.$$

$\hat{\tau}_{OLS}$ and $\hat{\tau}_{WS}$ are the OLS and WS statistics respectively based on the unknown $\{h_t\}_1^n$. They are akin to the $\hat{\tau}_{true}$ estimator defined above in the description of the SIMEX procedure. Next, we define

$$Y_{b,t}(\lambda) = Y_t + \sqrt{\lambda} e_{b,t}, \quad \lambda > 0$$

$$b = 1, 2, \dots, B \quad (1.3.1)$$

where $e_{b,t}$ are independent $\log(\chi^2)$ variables with mean zero and variance $\pi^2/2$.

Just as we had defined $\hat{\tau}_b(\lambda)$ above, similarly we define

$$\hat{\tau}_{b,OLS}(\lambda) = \frac{\hat{\phi}_{b,OLS}(\lambda) - 1}{\hat{\sigma}_{b,OLS}} \left(\sum_{t=2}^n y_{b,t-1}^2(\lambda) \right)^{1/2} \quad (1.3.2)$$

and

$$\hat{\tau}_{b,WS}(\lambda) = \left[\hat{\phi}_{b,WS}(\lambda) - 1 \right] \left[\sum_{t=2}^{n-1} y_{b,t}^2(\lambda) + n^{-1} \sum_{t=1}^n y_{b,t}^2(\lambda) \right]^{1/2} \hat{\sigma}_{b,WS}^{-1} \quad (1.3.3)$$

where $\hat{\phi}_{b,OLS}(\lambda)$, $\hat{\phi}_{b,WS}(\lambda)$, $\hat{\sigma}_{b,OLS}$ and $\hat{\sigma}_{b,WS}^{-1}$ are now based on $y_{b,t}(\lambda) = Y_{b,t}(\lambda) - \bar{Y}_{b,t}(\lambda)$

and $\bar{Y}_{b,t}(\lambda) = n^{-1} \sum_{t=1}^n Y_{b,t}(\lambda)$.

For $\lambda > 0$, we can generate a large number of independent pseudo-errors $\{\{e_{b,t}\}_{t=1}^n\}_{b=1}^B$ and compute $\hat{\tau}_{b,OLS}(\lambda)$ and $\hat{\tau}_{b,ws}(\lambda)$ for $b=1,2,\dots,B$. The final step of the SIMEX method calls for modeling $\hat{\tau}_{b,OLS}(\lambda)$ and $\hat{\tau}_{b,ws}(\lambda)$ as functions of λ for $\lambda \geq 0$ and using the model to extrapolate back to $\lambda = -1$. This yields $\hat{\tau}_{SIMEX,OLS}$ and $\hat{\tau}_{SIMEX,ws}$ as estimates of $\hat{\tau}_{OLS}$ and $\hat{\tau}_{ws}$ respectively. We can now compare $\hat{\tau}_{SIMEX,OLS}$ and $\hat{\tau}_{SIMEX,ws}$ with the percentiles given in Fuller (1996), for $\hat{\tau}_\mu$ and $\hat{\tau}_w$ respectively to either reject or fail to reject $H_0 : \phi = 1$.

Monte Carlo Study:

In this simulation study, we again set $h_0 = 0$ and $\mu = 0$ in the model (1.2.9). The η_t 's are generated as independent standard normal variables using the RANNOR function in SAS (V8). To generate e_t 's as independent $\log(\chi^2)$ variables with mean zero and variance $\pi^2/2$, the u_t 's were first generated as independent standard normal variables independent of the η_t 's and then the value 1.2704 was subtracted from the log of the square of each u_t . We generated samples of size $n=100$ with $\phi = 0.8, 0.85, 0.9, 0.95, 0.98, 1$. The two statistics $\hat{\tau}_\mu$ and $\hat{\tau}_{ws}$ were then computed from the expressions (1.2.5) and (1.2.7) respectively. Note that so far, we have just computed what we would usually compute when using the standard unit root test (based on OLS and WS) on the $\ln r_t^2$ process.

The next step is the simulation-step of the SIMEX method. The $e_{b,t}$'s for each b were generated as independent $\log(\chi^2)$ variables with mean zero and variance $\pi^2/2$ in the same way as the e_t 's were generated. For $\lambda = 0.5, 1.5, 2.5$ and 3.5 , we then generated samples of size 100 on the variables $\{Y_{b,t}(\lambda)\}_{b=1}^{10}$ using equation (1.3.1) and computed $\{\hat{\tau}_{b,OLS}(\lambda)\}_{b=1}^{10}$ and $\{\hat{\tau}_{b,ws}(\lambda)\}_{b=1}^{10}$ using expressions (1.3.2) and (1.3.3).

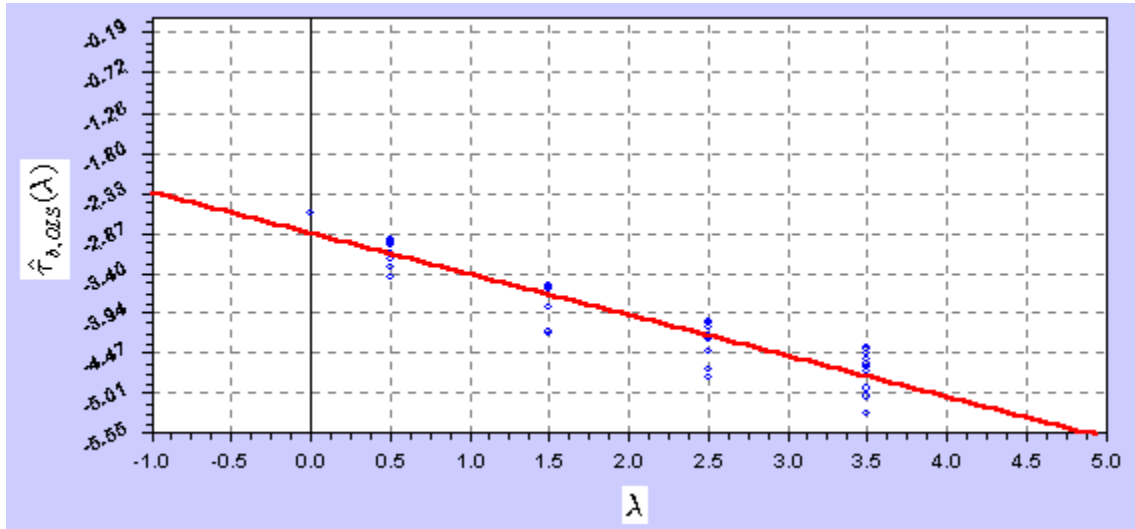


Figure 1-1: Linear extrapolating function $a + b\lambda$

The extrapolation step of the SIMEX method required some preliminary investigation. We set $\phi = 1$ and performed a few replications of the entire procedure described in the previous two paragraphs and plotted $\{\hat{\tau}_{b,OLS}(\lambda)\}_{b=1}^{10}$ and $\{\hat{\tau}_{b,ws}(\lambda)\}_{b=1}^{10}$ against λ . Figure 1-1 and Figure 1-2 show the fit of the linear $(a + b\lambda)$ and quadratic fit $(a + b\lambda + c\lambda^2)$ to a representative plot of $\{\hat{\tau}_{b,OLS}(\lambda)\}_{b=1}^{10}$. Figure 1-3 shows the fit of the curve, $a + b/(\lambda + 2)$ to the same plot. Note that the lone value plotted at $\lambda = 0$ corresponds to the statistic for $Y_i = \log(r_i^2)$.

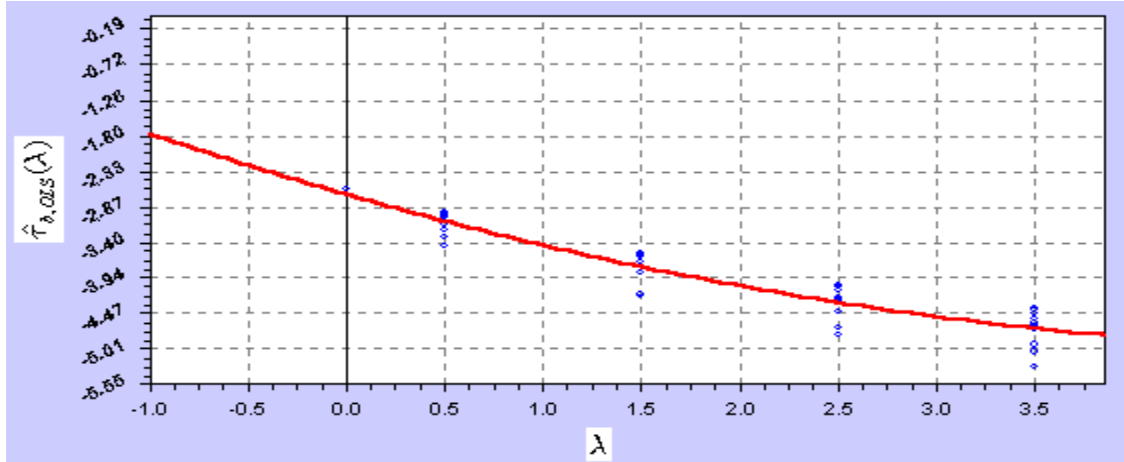


Figure 1-2: A quadratic extrapolating function $a + b\lambda + c\lambda^2$

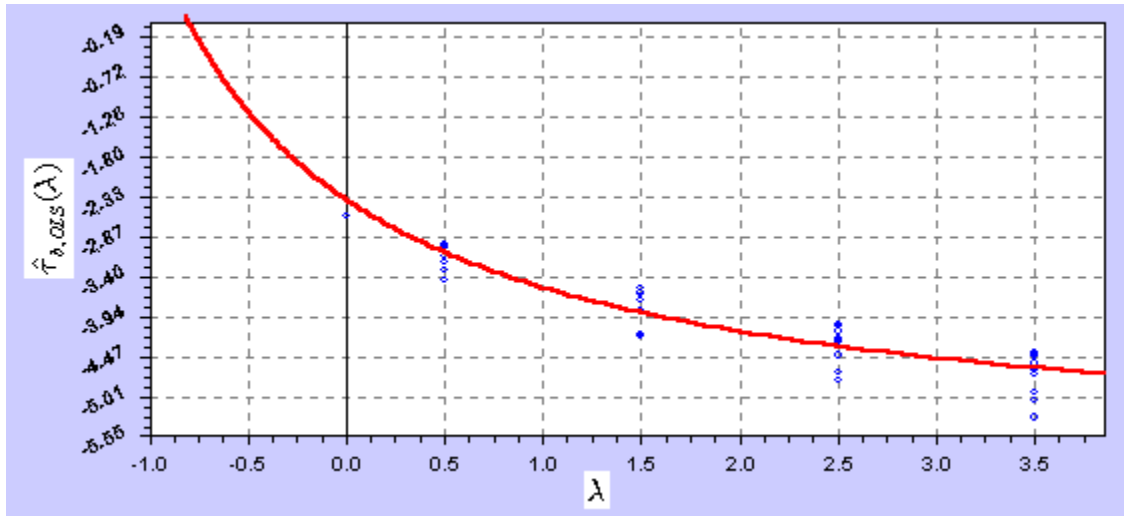


Figure 1-3: The extrapolating function $a + b/(\lambda + 2)$

Using each of the above three models, the $\hat{\tau}_{b,OLS}(\lambda)$ value was extrapolated at $\lambda = -1$ and was compared with the 5% critical value of -2.90 . The null hypothesis $H_0 : \phi = 1$ was rejected if the extrapolated value was less than -2.90 . To compare the three regression models, the p-value for the lack-of-fit test was computed for each of the three regression fits. The null hypothesis of no lack-of-fit was rejected if the corresponding p-value was less than 0.05 . On the basis of 5000 replications, the following table reports (a) the

proportion of times the null hypothesis $H_0 : \phi = 1$ was rejected and (b) the proportion of times the null hypothesis of no lack-of-fit was rejected for each of the three regression models, when the data was generated using $\phi = 0.8, 0.85, 0.9, 0.95, 0.98, 1$.

Regression curves similar to the ones obtained in figures 1, 2 and 3, were also obtained for $\{\hat{\tau}_{b,ws}(\lambda)\}_{b=1}^{10}$. Again, all the three regression models were applied and in each case, $\hat{\tau}_{b,ws}(\lambda)$ was extrapolated at $\lambda = -1$. The null hypothesis $H_0 : \phi = 1$ was rejected if the extrapolated value was less than -2.55 .

From the following tables of simulated power computations and Lack-of-fit tests, we observe that the results for the OLS estimator are very much similar to those for the WS estimator. When $\sigma_\eta^2 = 1$, the regression function $a + b/(\lambda + 2)$ provides the best fit to the OLS and the WS statistic for any value of ϕ and also maintains the size close to the nominal significance level of 0.05. While in the OLS case, the simulated size falls within 3σ limits, $(0.05 - 3\sqrt{0.05 * 0.95/5000}, 0.05 + 3\sqrt{0.05 * 0.95/5000}) = (0.041, 0.059)$ of 0.05, the simulated size in the WS case is 0.063, slightly above the upper 3σ limit, 0.059. As we let σ_η^2 assume larger values, the fit of the regression function $a + b/(\lambda + 2)$ deteriorates for values of ϕ , closer to 1. The simulated size of the test using this regression function also falls well below the lower 3σ limit, 0.041. In contrast to the performance of the inverse linear relationship, the performance of the linear and the quadratic polynomials improve as σ_η^2 increases from 1 to 10. When $\sigma_\eta^2 = 1.5$ or $\sigma_\eta^2 = 2$, none of the regression functions considered here yield a size close to the nominal size, 0.05. On one hand, the linear polynomial function delivers a rejection rate that is too high under the unit root null hypothesis to produce a reliable test and on the other hand, the inverse linear function, $a + b/(\lambda + 2)$ delivers a rejection rate that is too low to produce a test with sufficient power. The quadratic polynomial provides the best fit to the OLS and the WS statistics when $\sigma_\eta^2 = 2$.

Table 1-2: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 | \phi = \phi_0)$ when $\sigma_\eta^2 = 1, n = 100$

Regression	ϕ_0							
	0.70	0.80	0.85	0.90	0.93	0.95	0.98	1.00
$a + b\lambda$	1.000	1.000	1.000	0.997	0.982	0.957	0.874	0.690
Lack-of-fit	0.027	0.035	0.042	0.050	0.036	0.057	0.053	0.048
$a + b\lambda + c\lambda^2$	0.996	0.983	0.955	0.881	0.798	0.716	0.567	0.400
Lack-of-fit	0.061	0.078	0.090	0.105	0.113	0.118	0.114	0.101
$a + b/(\lambda + 2)$	0.640	0.467	0.354	0.232	0.168	0.129	0.081	0.051
Lack-of-fit	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002

Table 1-3: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 | \phi = \phi_0)$ when $\sigma_\eta^2 = 1.5, n = 100$

Regression	ϕ_0							
	0.70	0.80	0.85	0.90	0.93	0.95	0.98	1.00
$a + b\lambda$	1.000	1.000	0.997	0.976	0.928	0.869	0.728	0.522
Lack-of-fit	0.039	0.047	0.053	0.058	0.060	0.060	0.054	0.0428
$a + b\lambda + c\lambda^2$	0.986	0.935	0.854	0.714	0.600	0.503	0.358	0.235
Lack-of-fit	0.087	0.109	0.152	0.126	0.126	0.125	0.113	0.096
$a + b/(\lambda + 2)$	0.396	0.226	0.154	0.083	0.054	0.037	0.023	0.014
Lack-of-fit	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.005

Table 1-4: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 | \phi = \phi_0)$ when $\sigma_\eta^2 = 2, n = 100$

Regression	ϕ_0							
	0.70	0.80	0.85	0.90	0.93	0.95	0.98	1.00
$a + b\lambda$	1.000	0.998	0.986	0.935	0.846	0.759	0.585	0.393
Lack-of-fit	0.045	0.053	0.054	0.060	0.060	0.055	0.049	0.038
$a + b\lambda + c\lambda^2$	0.963	0.852	0.735	0.562	0.441	0.350	0.231	0.149
Lack-of-fit	0.103	0.121	0.127	0.127	0.125	0.123	0.107	0.086
$a + b/(\lambda + 2)$	0.239	0.113	0.065	0.033	0.020	0.014	0.008	0.005
Lack-of-fit	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.008

If we set $\sigma_\eta^2 = 10$, the empirical size obtained by employing the first-degree polynomial in λ is 0.035—still falls outside the 3σ -interval but the test yields higher power than that obtained by the other two regression functions, at lower values of ϕ .

While the fit provided by the function $a + b\lambda$ to the OLS or the WS statistic when $\sigma_\eta^2 = 10$ is not as good a fit as that provided by $a + b/(\lambda + 2)$ when $\sigma_\eta^2 = 1$, the power obtained by $a + b\lambda$ when $\sigma_\eta^2 = 10$ is much higher than that obtained by $a + b/(\lambda + 2)$ when $\sigma_\eta^2 = 1$.

We also performed simulations by setting $\sigma_\eta^2 = 0.5, 0.1$. For both these cases, although the fit provided by some of the regression functions considered are very good, the proportion of false rejections obtained at $\phi = 1$ is much larger than the nominal significance level.

Table 1-5: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 | \phi = \phi_0)$ when $\sigma_\eta^2 = 10, n = 100$

Regression	ϕ_0							
	0.70	0.80	0.85	0.90	0.93	0.95	0.98	1.00
$a + b\lambda$	0.947	0.667	0.425	0.210	0.116	0.079	0.045	0.035
Lack-of-fit	0.009	0.008	0.007	0.006	0.004	0.005	0.006	0.007
$a + b\lambda + c\lambda^2$	0.628	0.247	0.119	0.044	0.024	0.014	0.009	0.017
Lack-of-fit	0.023	0.021	0.021	0.018	0.017	0.017	0.014	0.017
$a + b/(\lambda + 2)$	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Lack-of-fit	0.001	0.003	0.007	0.013	0.019	0.024	0.038	0.041

Table 1-6: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 | \phi = \phi_0)$ when $\sigma_\eta^2 = 0.5$, $n = 100$

Regression	ϕ_0	
	0.7	1
$a + b\lambda$	1.000	0.860
Lack-of-fit	0.009	0.060
$a + b\lambda + c\lambda^2$	1.000	0.720
Lack-of-fit	0.021	0.082
$a + b/(\lambda + 2)$	0.896	0.200
Lack-of-fit	0.000	0.000
$a + b/(\lambda + 2)^2$	0.350	0.060
Lack-of-fit	0.000	0.000

Table 1-7: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 | \phi = \phi_0)$ when $\sigma_\eta^2 = 0.1$, $n = 100$

Regression	ϕ_0	
	0.70	1.00
$a + b\lambda$	1.000	1.000
Lack-of-fit	0.000	0.000
$a + b\lambda + c\lambda^2$	1.000	1.000
Lack-of-fit	0.000	0.010
$a + b/(\lambda + 2)$	1.000	0.849
Lack-of-fit	0.000	0.000
$a + b/(\lambda + 2)^2$	0.912	0.456
Lack-of-fit	0.000	0.000

The following tables report the power of the SIMEX test based on the WS estimator and the results of the lack-of-fit test on the basis of 5000 replications, for various values of σ_η^2 .

Table 1-8: $P(\hat{\tau}_{SIMEX,WS} < -2.55 | \phi = \phi_0)$ when $\sigma_\eta^2 = 1, n = 100$

Regression	ϕ_0							
	0.70	0.80	0.85	0.90	0.93	0.95	0.98	1.00
$a + b\lambda$	1.000	1.000	1.000	0.986	0.981	0.972	0.845	0.673
Lack-of-fit	0.022	0.023	0.040	0.051	0.052	0.059	0.057	0.046
$a + b\lambda + c\lambda^2$	0.992	0.989	0.945	0.865	0.801	0.742	0.543	0.367
Lack-of-fit	0.071	0.071	0.104	0.104	0.111	0.118	0.122	0.101
$a + b/(\lambda + 2)$	0.766	0.579	0.451	0.310	0.223	0.170	0.102	0.063
Lack-of-fit	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001

Table 1-9: $P(\hat{\tau}_{SIMEX,WS} < -2.55 | \phi = \phi_0)$ when $\sigma_\eta^2 = 1.5, n = 100$

Regression	ϕ_0							
	0.70	0.80	0.85	0.90	0.93	0.95	0.98	1.00
$a + b\lambda$	1.000	1.000	1.000	0.986	0.900	0.896	0.706	0.542
Lack-of-fit	0.033	0.059	0.066	0.055	0.089	0.061	0.044	0.015
$a + b\lambda + c\lambda^2$	0.956	0.922	0.841	0.722	0.615	0.509	0.351	0.237
Lack-of-fit	0.086	0.108	0.154	0.136	0.116	0.111	0.112	0.083
$a + b/(\lambda + 2)$	0.398	0.2522	0.156	0.088	0.054	0.040	0.012	0.015
Lack-of-fit	0.000	0.000	0.000	0.000	0.000	0.0010	0.001	0.005

Table 1-10: $P(\hat{\tau}_{SIMEX,WS} < -2.55 | \phi = \phi_0)$ when $\sigma_\eta^2 = 2, n = 100$

Regression	ϕ_0							
	0.70	0.80	0.85	0.90	0.93	0.95	0.98	1.00
$a + b\lambda$	1.000	0.998	0.956	0.932	0.880	0.746	0.555	0.336
Lack-of-fit	0.043	0.052	0.054	0.063	0.065	0.055	0.048	0.039
$a + b\lambda + c\lambda^2$	0.950	0.851	0.715	0.587	0.492	0.356	0.200	0.160
Lack-of-fit	0.101	0.103	0.127	0.127	0.110	0.138	0.160	0.084
$a + b/(\lambda + 2)$	0.225	0.135	0.066	0.032	0.022	0.014	0.005	0.005
Lack-of-fit	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.010

Table 1-11: $P(\hat{\tau}_{SIMEX,WS} < -2.55 | \phi = \phi_0)$ when $\sigma_\eta^2 = 10$, $n = 100$

Regression	ϕ_0							
	0.70	0.80	0.85	0.90	0.93	0.95	0.98	1.00
$a + b\lambda$	0.948	0.662	0.455	0.228	0.120	0.088	0.046	0.035
Lack-of-fit	0.009	0.008	0.007	0.006	0.004	0.004	0.005	0.007
$a + b\lambda + c\lambda^2$	0.628	0.247	0.119	0.044	0.024	0.014	0.009	0.017
Lack-of-fit	0.021	0.022	0.025	0.020	0.016	0.020	0.015	0.017
$a + b/(\lambda + 2)$	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Lack-of-fit	0.001	0.001	0.005	0.013	0.020	0.022	0.036	0.048

Table 1-12: $P(\hat{\tau}_{SIMEX,WS} < -2.55 | \phi = \phi_0)$ when $\sigma_\eta^2 = 0.5$, $n = 100$

Regression	ϕ_0	
	0.70	1.00
$a + b\lambda$	1.00	0.89
Lack-of-fit	0.00	0.03
$a + b\lambda + c\lambda^2$	1.00	0.77
Lack-of-fit	0.00	0.08
$a + b/(\lambda + 2)$	0.96	0.25
Lack-of-fit	0.00	0.00
$a + b/(\lambda + 2)^2$	0.40	0.09
Lack-of-fit	0.00	0.00

Table 1-13: $P(\hat{\tau}_{SIMEX,WS} < -2.55 | \phi = \phi_0)$ when $\sigma_\eta^2 = 0.1$, $n = 100$

	ϕ_0	
	0.70	1.00
Regression		
$a + b\lambda$	1.00	1.00
Lack-of-fit	0.00	0.00
$a + b\lambda + c\lambda^2$	1.00	1.00
Lack-of-fit	0.00	0.00
$a + b/(\lambda + 2)$	1.00	0.86
Lack-of-fit	0.00	0.00
$a + b/(\lambda + 2)^2$	0.89	0.51
Lack-of-fit	0.00	0.00

Section 4. Performance of SIMEX-based test using a different mechanism to generate pseudo-errors

The method of adding additional measurement errors to the original data according to (1.3.1) can be slightly modified. We shall refer to the following procedure as the *non-lambda approach* in the later sections. Let us define $Y_{b,t}^{(1)} = \log(r_t^2)$ and subsequently the following:

$$Y_{b,t}^{(k)} = Y_{b,t}^{(k-1)} + e_{b,t}^{(k)}, \quad k = 2, 3, \dots \quad (1.4.1)$$

where $e_{b,t}^{(k)}$ are independent $\log(\chi^2)$ variables with mean zero and variance $\pi^2/2$.

$Y_{b,t}^{(k)}$ is the result of adding ' k ' additional independent measurement errors to h_t^* . Let us denote the OLS and the WS statistics based on $Y_{b,t}^{(k)}$ by $\hat{\tau}_{b,OLS}^{(k)}$ and $\hat{\tau}_{b,ws}^{(k)}$ respectively, for $b = 1, 2, \dots, B$. For $k = 2, 3, \dots$, we can generate a large number of independent pseudo-errors $\{\{e_{b,t}^{(k)}\}_{t=1}^n\}_{b=1}^B$ and compute $\hat{\tau}_{b,OLS}^{(k)}$ and $\hat{\tau}_{b,ws}^{(k)}$ for $b = 1, 2, \dots, B$. Finally, $\hat{\tau}_{b,OLS}^{(k)}$ and $\hat{\tau}_{b,ws}^{(k)}$ can be modeled as functions of k for $k = 1, 2, 3, \dots$. To get the SIMEX estimators $\hat{\tau}_{SIMEX,OLS}$ and $\hat{\tau}_{SIMEX,ws}$, we use the estimated model to extrapolate back to $k = 0$. Note

that the case $k = 0$ corresponds to the unobserved h_t^* that is devoid of any measurement error. At the nominal significance level of 0.05, we compare $\hat{\tau}_{SIMEX,OLS}$ and $\hat{\tau}_{SIMEX,ws}$ with the percentiles -2.90 (for $\hat{\tau}_{OLS}$) and -2.55 (for $\hat{\tau}_{ws}$) respectively to either reject or fail to reject $H_0 : \phi = 1$.

The power computations for the OLS and the WS estimators based on the above approach of generating pseudo-errors, are given below. The results are very similar to the ones obtained in the previous section.

For $\sigma_\eta^2 = 1$, the regression function $a + b/(1+k)$ provides the best fit to the OLS as well as the WS statistics. Again, the simulated size of 0.0558 in the case of OLS falls within the 3σ limits of the nominal size of 0.05. In the case of WS, the simulated size is 0.0652 which falls outside the 3σ limits of the nominal size of 0.05.

Table 1-14: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 | \phi = \phi_0)$ when $\sigma_\eta^2 = 1$, $n = 100$

Regression	ϕ_0							
	0.70	0.80	0.85	0.90	0.93	0.95	0.98	1.00
$a + bk$	1.000	1.000	1.000	0.999	0.994	0.980	0.920	0.759
Lack-of-fit	0.021	0.031	0.037	0.044	0.047	0.049	0.050	0.046
$a + bk + ck^2$	0.997	0.984	0.961	0.896	0.825	0.754	0.615	0.445
Lack-of-fit	0.062	0.075	0.086	0.103	0.108	0.114	0.115	0.105
$a + b/(1+k)$	0.636	0.471	0.364	0.253	0.181	0.141	0.083	0.056
Lack-of-fit	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.010

Table 1-15: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 | \phi = \phi_0)$ when $\sigma_\eta^2 = 1.5$, $n = 100$

Regression	ϕ_0							
	0.70	0.80	0.85	0.90	0.93	0.95	0.98	1.00
$a + bk$	1.000	1.000	0.999	0.991	0.965	0.921	0.803	0.602
Lack-of-fit	0.033	0.044	0.047	0.053	0.053	0.053	0.050	0.043
$a + bk + ck^2$	0.986	0.939	0.879	0.754	0.650	0.562	0.413	0.277
Lack-of-fit	0.079	0.010	0.109	0.121	0.125	0.127	0.119	0.101
$a + b/(1+k)$	0.415	0.251	0.167	0.096	0.065	0.048	0.026	0.015
Lack-of-fit	0.000	0.000	0.000	0.001	0.001	0.002	0.006	0.022

Table 1-16: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 | \phi = \phi_0)$ when $\sigma_\eta^2 = 2$, $n = 100$

Regression	ϕ_0							
	0.70	0.80	0.85	0.90	0.93	0.95	0.98	1.00
$a + bk$	1.000	0.999	0.997	0.969	0.912	0.840	0.681	0.474
Lack-of-fit	0.041	0.0482	0.051	0.052	0.052	0.050	0.048	0.043
$a + bk + ck^2$	0.965	0.876	0.774	0.619	0.492	0.409	0.276	0.178
Lack-of-fit	0.094	0.111	0.120	0.128	0.129	0.126	0.113	0.094
$a + b/(1+k)$	0.257	0.125	0.079	0.039	0.022	0.015	0.010	0.006
Lack-of-fit	0.000	0.000	0.000	0.001	0.003	0.004	0.012	0.033

Table 1-17: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 | \phi = \phi_0)$ when $\sigma_\eta^2 = 10$, $n = 100$

Regression	ϕ_0							
	0.70	0.80	0.85	0.90	0.93	0.95	0.98	1.00
$a + bk$	0.975	0.768	0.548	0.301	0.200	0.123	0.068	0.051
Lack-of-fit	0.011	0.014	0.015	0.014	0.013	0.012	0.011	0.0132
$a + bk + ck^2$	0.665	0.306	0.159	0.066	0.041	0.023	0.015	0.019
Lack-of-fit	0.036	0.038	0.039	0.038	0.034	0.035	0.034	0.034
$a + b/(1+k)$	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Lack-of-fit	0.008	0.026	0.044	0.072	0.098	0.109	0.137	0.141

Table 1-18: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 | \phi = \phi_0)$ when $\sigma_\eta^2 = 0.5$, $n = 100$

Regression	ϕ_0	
	0.70	1.00
$a + bk$	1.000	0.751
Lack-of-fit	0.000	0.050
$a + bk + ck^2$	1.000	0.551
Lack-of-fit	0.000	0.050
$a + b/(1+k)$	0.900	0.151
Lack-of-fit	0.000	0.000
$a + b/(1+k)^2$	0.252	0.100
Lack-of-fit	0.000	0.000

Table 1-19: $P(\hat{\tau}_{SIMEX,OLS} < -2.90 | \phi = \phi_0)$ when $\sigma_\eta^2 = 0.1$, $n = 100$

Regression	ϕ_0	
	0.70	1.00
$a + bk$	1.000	0.946
Lack-of-fit	0.000	0.001
$a + bk + ck^2$	0.976	0.941
Lack-of-fit	0.000	0.001
$a + b/(1+k)$	0.968	0.701
Lack-of-fit	0.000	0.000
$a + b/(1+k)^2$	0.801	0.300
Lack-of-fit	0.000	0.000

For the cases $\sigma_\eta^2 = 1.5$ and $\sigma_\eta^2 = 2$, the regression function, $a + b/(1+k)$ provides the best fit to both, the OLS and the WS statistics. But, using this function to get the SIMEX estimate results in gross under-estimation of the nominal significance level.

Finally, when $\sigma_\eta^2 = 10$, the linear regression function, $a + bk$ delivers an empirical size of 0.0512 (within the 3σ limits) for both, the OLS and the WS statistics. The fit of this

function is not as good as the fit of $a + b/(1+k)$ when $\sigma_{\eta}^2 = 1$. For the cases, $\sigma_{\eta}^2 = 0.1$ and $\sigma_{\eta}^2 = 0.5$, none of the regression functions considered here yielded any favorable results.

Table 1-20: $P(\hat{\tau}_{SIMEX,WS} < -2.55 | \phi = \phi_0)$ when $\sigma_{\eta}^2 = 1$, $n = 100$

Regression	ϕ_0							
	0.70	0.80	0.85	0.90	0.93	0.95	0.98	1.00
$a + bk$	1.000	1.000	1.000	1.000	1.000	0.994	0.960	0.806
Lack-of-fit	0.025	0.029	0.0384	0.047	0.050	0.053	0.057	0.058
$a + bk + ck^2$	0.999	0.996	0.982	0.948	0.900	0.849	0.705	0.496
Lack-of-fit	0.065	0.079	0.095	0.109	0.116	0.120	0.128	0.136
$a + b/(1+k)$	0.715	0.553	0.436	0.311	0.231	0.180	0.107	0.065
Lack-of-fit	0.000	0.003	0.000	0.000	0.001	0.001	0.003	0.010

Table 1-21: $P(\hat{\tau}_{SIMEX,WS} < -2.55 | \phi = \phi_0)$ when $\sigma_{\eta}^2 = 1.5$, $n = 100$

Regression	ϕ_0							
	0.70	0.80	0.85	0.90	0.93	0.95	0.98	1.00
$a + bk$	1.000	1.000	0.998	0.995	0.955	0.945	0.813	0.602
Lack-of-fit	0.032	0.043	0.047	0.052	0.057	0.052	0.052	0.040
$a + bk + ck^2$	0.960	0.947	0.888	0.739	0.658	0.562	0.423	0.260
Lack-of-fit	0.076	0.090	0.129	0.132	0.124	0.127	0.104	0.111
$a + b/(1+k)$	0.398	0.246	0.124	0.097	0.056	0.047	0.027	0.015
Lack-of-fit	0.000	0.000	0.000	0.001	0.001	0.002	0.006	0.021

Table 1-22: $P(\hat{\tau}_{SIMEX,WS} < -2.55 | \phi = \phi_0)$ when $\sigma_\eta^2 = 2, n = 100$

Regression	ϕ_0							
	0.70	0.80	0.85	0.90	0.93	0.95	0.98	1.00
$a + bk$	1.000	1.000	0.967	0.930	0.911	0.800	0.646	0.480
Lack-of-fit	0.041	0.042	0.055	0.055	0.048	0.050	0.045	0.041
$a + bk + ck^2$	0.985	0.896	0.751	0.625	0.491	0.423	0.255	0.187
Lack-of-fit	0.015	0.111	0.116	0.134	0.129	0.130	0.112	0.094
$a + b/(1+k)$	0.260	0.134	0.068	0.040	0.040	0.012	0.010	0.006
Lack-of-fit	0.000	0.000	0.000	0.001	0.002	0.004	0.012	0.032

Table 1-23: $P(\hat{\tau}_{SIMEX,WS} < -2.55 | \phi = \phi_0)$ when $\sigma_\eta^2 = 10, n = 100$

Regression	ϕ_0							
	0.70	0.80	0.85	0.90	0.93	0.95	0.98	1.00
$a + bk$	0.978	0.770	0.512	0.311	0.220	0.130	0.056	0.051
Lack-of-fit	0.010	0.014	0.015	0.011	0.011	0.013	0.011	0.012
$a + bk + ck^2$	0.667	0.359	0.154	0.059	0.043	0.012	0.015	0.017
Lack-of-fit	0.038	0.034	0.039	0.032	0.044	0.035	0.039	0.033
$a + b/(1+k)$	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Lack-of-fit	0.006	0.026	0.045	0.072	0.094	0.112	0.140	0.141

Table 1-24: $P(\hat{\tau}_{SIMEX,WS} < -2.55 | \phi = \phi_0)$ when $\sigma_\eta^2 = 0.5, n = 100$

Regression	ϕ_0	
	0.70	1.00
$a + bk$	1.000	0.765
Lack-of-fit	0.000	0.050
$a + bk + ck^2$	0.980	0.551
Lack-of-fit	0.000	0.052
$a + b/(1+k)$	0.911	0.160
Lack-of-fit	0.000	0.000
$a + b/(1+k)^2$	0.250	0.091
Lack-of-fit	0.000	0.000

Table 1-25: $P(\hat{\tau}_{SIMEX,WS} < -2.55 | \phi = \phi_0)$ when $\sigma_\eta^2 = 0.1$, $n = 100$

Regression	ϕ_0	
	0.70	1.00
$a + bk$	0.998	0.991
Lack-of-fit	0.000	0.000
$a + bk + ck^2$	0.956	0.940
Lack-of-fit	0.000	0.000
$a + b/(1+k)$	0.999	0.721
Lack-of-fit	0.001	0.000
$a + b/(1+k)^2$	0.799	0.312
Lack-of-fit	0.000	0.000

Section 5. Comparing SIMEX, ADF and Instrumental Variables approach

Hall (1989) proposed a new approach to testing for unit roots in time series with moving average innovations based on an instrumental variable estimator. Consider the problem of testing the null hypothesis $H_0 : \phi = 1$, against the alternative hypothesis $H_a : \phi < 1$ in the ARMA model

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \varepsilon_t - \theta\varepsilon_{t-1}, Y_0 = 0, \varepsilon_t \sim NID(0, \sigma^2) \quad (1.5.1)$$

If we let $u_t = \varepsilon_t - \theta\varepsilon_{t-1}$, then Y_{t-1} is correlated with u_t . This correlation gives rise to inconsistent OLS estimators of ϕ . In the Instrumental Variables (IV) approach, we circumvent this problem by choosing an *instrument* that is uncorrelated with the error u_t , but has strong correlation with Y_{t-1} . If Y_{t-2} is chosen as the instrument, the IV estimator given by

$$\hat{\phi}_{IV} = \frac{\sum_{t=2}^n (Y_t - \bar{y})(Y_{t-2} - \bar{y})}{\sum_{t=2}^n (Y_{t-1} - \bar{y})(Y_{t-2} - \bar{y})} \quad (1.5.2)$$

turns out to be a consistent estimator of ϕ in this case. Furthermore, Pantula (1989) showed that among the IV estimators that use $Y_{t-k}, k \geq 2$ as an instrument for ϕ in the model (1.5.1) considered by Hall (1989), the IV estimator with Y_{t-2} as the instrument has the smallest asymptotic variance when $|\phi| < 1$. The test statistic corresponding to $\hat{\phi}_{IV}$ that can be used to test the unit root hypothesis is given by

$$n(\hat{\phi}_{IV} - 1) \quad (1.5.3)$$

The asymptotic null distribution of $n(\hat{\phi}_{IV} - 1)$ is as tabulated by Fuller (Table 10.A.1. in Fuller, 1996).

Monte Carlo Study

In this simulation study, we again consider the model (1.2.9) and set $h_0 = 0$ and $\mu = 0$. The η_t 's are generated as independent standard normal variables using the RANNOR function in SAS (V8). The e_t 's were generated as independent $\log(\chi^2)$ variables with mean zero and variance $\pi^2/2$ as was done in the Monte Carlo study of Section 3. We generated samples of size $n=100$ with $\phi=1$ and computed $n(\hat{\phi}_{IV} - 1)$ based on $Y_t = \log(r_t^2)$ for each of the 10,000 samples. The 5% empirical critical value from the simulated distribution of $n(\hat{\phi}_{IV} - 1)$ was computed to be -22.49 . Next, we generated samples of size $n=100$ for $\phi = 0.7, 0.8, 0.85, 0.9, 0.95, 0.98$. The IV statistic was computed and compared with the critical value of -22.49 . The unit-root null hypothesis was rejected if the computed value of $n(\hat{\phi}_{IV} - 1)$ was less than -22.49 . The procedure was replicated 5,000 times for each value of ϕ and the power was computed as the proportion

of times the null hypothesis was rejected. Here is the table of simulated power computations.

ϕ	0.7	0.8	0.85	0.9	0.95	0.98
Power	0.6778	0.5084	0.3896	0.2612	0.1468	0.0894

If we use the asymptotic distribution of $n(\hat{\phi}_{IV} - 1)$ to perform the test, we have to compare the computed statistic value with the 5% critical value of -14.1 (Appendix 10.A. in Fuller, 1996). In that case we get following power performance:

ϕ	0.8	0.85	0.90	0.95	0.98	1
Power	0.6652	0.572	0.4344	0.2806	0.1898	0.1218

From the above table we see that as concluded by Hall (1989), the use of an instrumental variable estimator does not solve the size problem since the test statistic requires much larger sample size to converge to its asymptotic distribution.

Augmented Dickey Fuller (ADF)

According to Said and Dickey (1984), we can approximate the ARMA representation of the process Y_t in (1.5.1) by a long AR process and use the test criteria suggested by Dickey and Fuller (1979). In other words, we can first regress $\dot{y}_t = (Y_t - \bar{y}) - (Y_{t-1} - \bar{y}) = y_t - y_{t-1}$ on $y_{t-1}, \dot{y}_{t-1}, \dot{y}_{t-2}, \dots, \dot{y}_{t-k}$ for a suitably large k and then compare the studentized statistic of the estimated coefficient on y_{t-1} to the $\hat{\tau}_\mu$ tables in Fuller (1996).

Monte Carlo study

The data is generated in exactly the same way as it was generated in the Monte Carlo study of the IV estimator. To determine the appropriate value of k , we started with $k = 12$, and used PROC ARIMA in SAS (V8) to compute the p-value for the ADF test (single mean estimated). Based on 1000 replications, we estimated the size of the test as the ratio of the number of times the null hypothesis was rejected (p-value < 0.05) to the total number of replications (1000). The estimated size when using 12 lags was about

0.03 and so we kept reducing the number of lags. At $k = 5, 6$ and 7 , the size was approximately 0.04, 0.05 and 0.06 respectively. The following table reports the size and the power of the ADF test with lags 6 and 12 and compares them with the IV-based test and the SIMEX-based test.

Table 1-26: Comparing ADF, IV and SIMEX-based unit root test when $\sigma_{\eta}^2 = 1$, $n = 100$

Test Statistic	ϕ_0					
	1	0.98	0.95	0.90	0.85	0.80
<i>ADF using 12 lags</i>	0.03	0.05	0.06	0.09	0.13	0.17
<i>ADF using 6 lags</i>	0.05	0.08	0.09	0.21	0.34	0.46
<i>$\hat{\tau}_{SIMEX,OLS}$ using $a + b/(\lambda + 2)$ as the extrapolating function.</i>	0.05	0.08	0.130	0.23	0.35	0.47
Lack-of-fit. (# times LOF test rejected/5000)	0.00	0.00	0.00	0.00	0.00	0.00
<i>Instrumental Variable (IV) approach</i>	0.12	0.19	0.28	0.43	0.57	0.66

We observe that the SIMEX-based test performs as well as the ADF test using 6 lags.

Section 6. SIMEX applied to the IV estimator

In this subsection, we present the application of the SIMEX procedure to the IV estimator. Observations were again simulated from the model (1.2.9) as was done in the monte carlo study of Section 3. Let Z_{IV} denote the IV statistic based on $\{h_t\}_1^n$. We generated samples of size $n = 100$ with $\phi = 1$ and computed Z_{IV} for each of the 10,000 samples. The 5% empirical critical value from the simulated distribution of Z_{IV} was computed to be -14.94852 .

For a given value of λ , let $Z_{b,IV}(\lambda)$ denote the IV statistic based on $Y_{b,t}(\lambda)$, $b = 1, 2, \dots, B$. Here are two representative plots of $Z_{b,IV}(\lambda)$ against λ , $b = 1, 2, \dots, 100$.

The plots clearly show that the variance of the $Z_{b,IV}(\lambda)$ values increases with increasing λ .

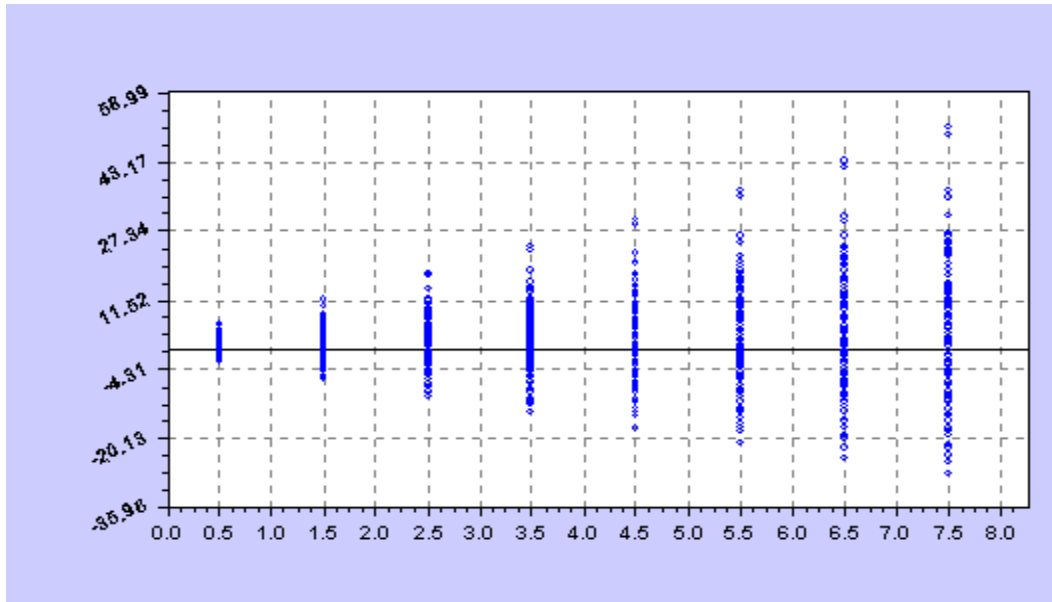


Figure 1-4: (a) Sample Plot of $Z_{b,IV}(\lambda)$ against λ

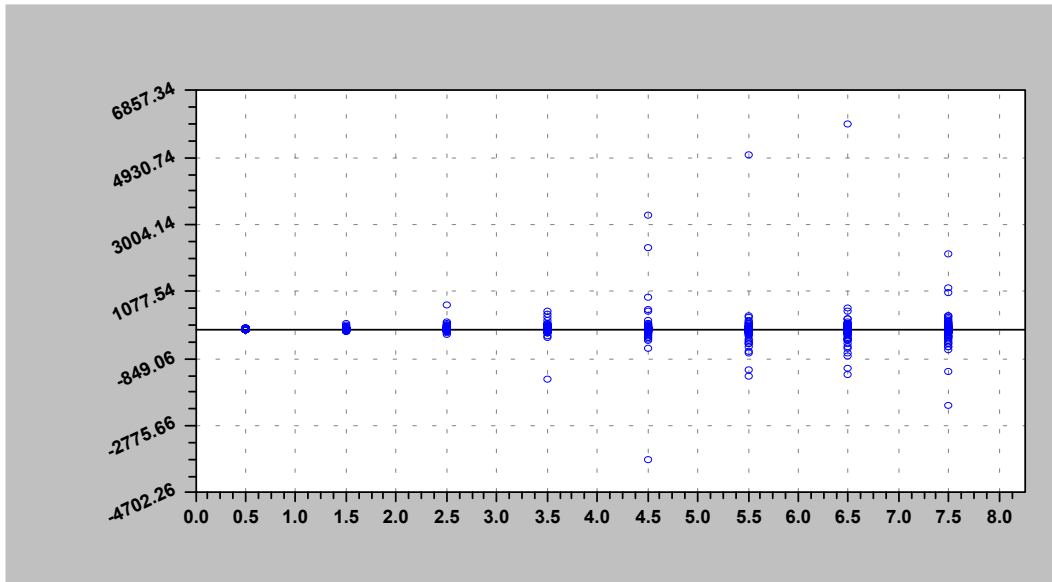


Figure 1-5: (b) Sample Plot of $Z_{b,IV}(\lambda)$ against λ

We consider the Weighted Least Squares (WLS) approach to account for the heterogeneous variance structure. Suppose we are fitting the following model to the Y (response) values:

$$Y = X\beta + u \quad (1.6.1)$$

with

$$E(u) = 0$$

and

$$E(uu') = V$$

where X is a matrix of known constants.

The minimum variance unbiased linear estimator of β under the above setup is given by Weighted Least Squares estimator

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y \quad (1.6.2)$$

where V is assumed to be a known, symmetric, positive-definite matrix.

The estimator given by (1.6.2) accounts for possible heteroscedasticity of the errors u .

In our problem, Y is a 801×1 vector consisting of the value $Z_{IV}(0)$ and the values $\{Z_{b,IV}(\lambda)\}_{b=1}^{100}, \lambda = 0.5, 1.5, 2.5, \dots, 7.5$. We considered three different specifications for X -- the first one being just the column of ones, $X = [1, 1, \dots, 1]'_{1 \times 801}$. The two other X matrices were constructed by augmenting the column of ones by the column of λ values and then, the column of λ^2 values. We assumed a diagonal structure for the variance matrix V and estimated the actual variances by their sample versions. For instance, the estimated variance of $Z_{IV}(\lambda)$ is given by $\widehat{Var}(Z_{IV}(\lambda)) = (1/99) \sum_{b=1}^{100} (Z_{b,IV}(\lambda) - \bar{Z}_{b,IV}(\lambda))^2$ for $\lambda = 0.5, 1.5, \dots, 7.5$. Within the SIMEX procedure, computing $\widehat{Var}(Z_{IV}(0))$ is not possible because $Z_{IV}(0)$ corresponds to the IV statistic for the series $\ln r_t^2$ and there is no replication at $\lambda = 0$. In order to compute $\widehat{Var}(Z_{IV}(0))$, we generated samples of size 100 on h_t and $\log(r_t^2)$ using (1.1.1) under the assumptions, $\phi = 1$ and $\sigma_\eta^2 = 1$. The process

was replicated 100 times and each time, $Z_{IV}(0)$ was computed. The variance of the $Z_{IV}(0)$ values so generated was computed to be 3.71.

We thus estimated the regression equation given by (1.6.1) using (1.6.2). The value of $Z_{IV}(\lambda)$ was then extrapolated at $\lambda = -1$ and compared with the critical value, -14.94. Here are the power computations for the three regression models. None of these maintain the nominal significance level at 0.05.

Table 1-27: SIMEX applied to the IV estimator, $n = 100$

Regression	$\phi = 1$	$\phi = 0.7$
Constant	0.1356	0.7028
$a + b\lambda$	0.1122	0.7164
$a + b\lambda + c\lambda^2$	0.2056	0.7506

The linear model in λ gives the smallest rejection rate under the null hypothesis, $\phi = 1$. It also has a slightly better power than the ‘Constant’ fit at $\phi = 0.7$. The size of the test is also marginally better than that obtained by using the ordinary IV estimator based on $\log(r_i^2)$ (simulations in the previous section) and comparing it with the asymptotic percentiles.

Section 7. A SIMEX test based on p-values

The tests that have been discussed in the previous sections were based on extrapolating the value of the test statistic for the process h_t and then computing the rejection probabilities under the null and the alternative hypothesis.

Another approach would be to extrapolate the p-value for the unobserved h_t , and then reject or accept the null hypothesis depending on whether the p-value is less or greater than the set level of significance. Again here, we have analyzed the performance of the OLS and the WS estimators under the two different ways of generating pseudo-errors.

The Dickey-Fuller test based on OLS rely on the innovation process being white noise and so, these tests have poor power when the innovations are moving averages. This loss of power shows up in the very small and decreasing p-values when the unit root test is performed successively on $\log(r_t^2)$, $Y_{b,t}^{(2)}$, $Y_{b,t}^{(3)}$, $Y_{b,t}^{(4)}$, ... Under $H_0 : \phi = 1$, if the pseudo-errors are generated using the *non-lambda approach*, we get the following representative plots of the p-values against k .

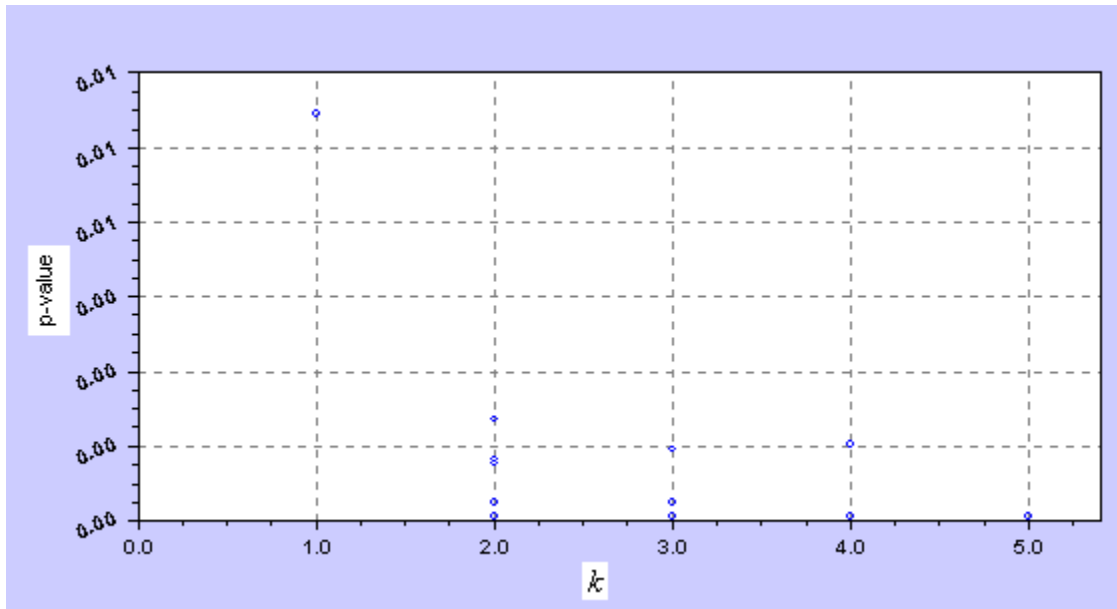


Figure 1-6: (a) Plot of p-values against k : # of measurement errors

Since we are dealing with probabilities, a logit transformation of the p-values is appropriate so that a sensible p-value (between 0 and 1) could be extrapolated at $k = 0$. The graphs of the transformed p-values were also indicative of a negative relationship

between the p-values and k . But several attempts to model the p-value as a function of k , failed to deliver the correct size for the test. Among the regression functions that were used to model the relationship, polynomials in k of up to degree 4 were estimated and each in turn was used to extrapolate the p-value at $k = 0$. The procedure was replicated 5,000 times and proportion of p-values less than 0.05 was computed for each regression function. These proportions are reported in the following table for the three different regression functions that produced the lowest values. All of these are much higher than 0.05.

Table 1-28: Rejection frequencies when SIMEX is applied to p-values (λ – approach), $n = 100$

$\sigma_{\eta}^2 = 1$	ϕ
Regression	1
$a + bk$	0.846
$a + bk + ck^2$	0.930
$a + b/(1 + k)$	0.631
$a + b/(1 + k)^2$	0.457

Table 1-29: Rejection frequencies when SIMEX is applied to p-values ($Non - \lambda$ – approach), $n = 100$

$\sigma_{\eta}^2 = 1$	ϕ
Regression	1
$a + b\lambda$	0.850
$a + b\lambda + c\lambda^2$	0.913
$a + b/(1 + \lambda)$	0.632
$a + b/(1 + \lambda)^2$	0.460

The major reason for getting a higher size than the nominal level is the presence of data sets that produce zero (or a value very close to zero) as the p-value at $k=1$ (see the figure below). A p-value of zero at $k=0$ corresponds to a large negative value of the OLS test statistic for the $\ln r_t^2$ process. As we add pseudo-errors to derive the successive random variables, $Y_{2,t}, Y_{3,t}$, the MA parameters of the corresponding ARMA processes have still larger negative values and hence, still larger negative values of the test statistic. Thus, the p-values computed at $k=2,3,4,5$ continue to be zero.

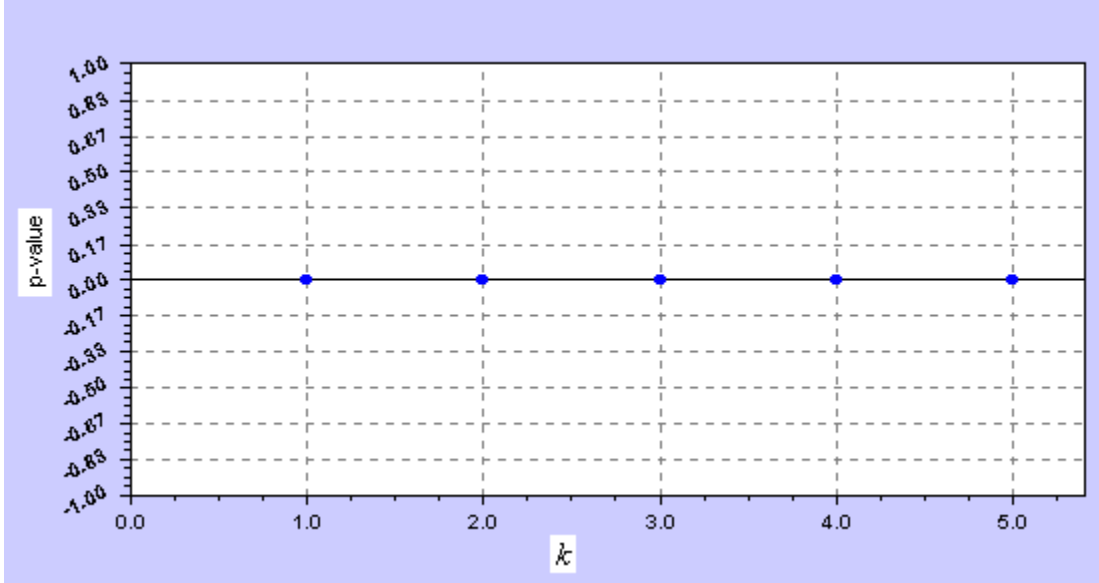


Figure 1-7: (b) Plot of p-values against k : # of measurement errors

Using the *lambda approach* to generate pseudo-errors does not yield favorable results due to the same reason cited above. During the simulation process, once the p-value turns out to be zero at $\lambda = 0$ ($k = 0$), it remains at zero for any larger value of λ (k).

The p-values corresponding to WS estimator exhibit the same pattern as exhibited by those obtained from the OLS estimator and hence our attempts at modeling the p-values as a function of λ or k results in rejection rates, much higher than the nominal significance levels under the null hypothesis.

Section 8. Extrapolating function based on Asymptotic theory

So far, in all our simulation studies, we have always assumed that the value of σ_η^2 is known. We found that the extrapolating function $a + \frac{b}{\lambda + 2}$ serves as a good extrapolating function that not only provides a good fit to the $\hat{\tau}_{OLS}$ estimates as a function of λ , but also maintains the nominal level of the test and delivers as good a power as the ADF test. However this observation was valid only when σ_η^2 took the value 1. As we increased or decreased σ_η^2 the level and power properties based on the same extrapolating function degraded even when the fit of the function was very good in some cases. The simulation studies in the previous sections have not provided any conclusive approach to choosing an extrapolating function as the value of σ_η^2 changes. In this section, we resort to the asymptotic theory of $\hat{\tau}_{OLS}$ under the ARMA representation of the $\log(r_t^2)$ process to gain a better understanding of the effect of measurement error on the $\hat{\tau}_{OLS}$ estimates.

In the previous sections, our method of selecting the appropriate extrapolating function for a particular value of σ_η^2 was entirely based on the visual approach. Also, the changing value of σ_η^2 was not incorporated into the regressor variable. We always plotted the $\hat{\tau}_{OLS}$ values against λ irrespective of the value of σ_η^2 . We thus find a scope for improvement by attempting to incorporate the information of σ_η^2 into the regressor. One way of achieving this goal would be to use the ARMA representation of the $\log(r_t^2)$ process. We have already discussed the ARMA representation in Section 2. We will elaborate on it in the framework of the SIMEX procedure. The pseudo errors will henceforth be generated only using the λ -approach unless otherwise stated. Under the null hypothesis of a unit root, the $\log(r_t^2)$ process and the pseudo-processes $Y_{b,t}(\lambda) = Y_t + \sqrt{\lambda}e_{b,t}$, have an ARMA(1,1) representation with ‘1’ as the common AR parameter but the MA parameters defined as follows:

$$\theta_\lambda = \begin{cases} 0 & \lambda = -1 \\ \frac{-(2\sigma_{e,\lambda}^2 + \sigma_\eta^2) + \sqrt{\sigma_\eta^4 + 4\sigma_{e,\lambda}^2 \sigma_\eta^2}}{2\sigma_{e,\lambda}^2} & \lambda > -1 \end{cases} \quad (1.8.1)$$

where $\sigma_{e,\lambda}^2 = (1+\lambda) \frac{\pi^2}{2}$. Thus, each of the B pseudo-random variables $Y_{b,t}(\lambda)$ created at a particular value of λ has the same MA parameter θ_λ .

Note that the case of no measurement error is now identified by $\theta_\lambda = 0$. Now we will try to exploit the asymptotic distribution of $\hat{\tau}_{OLS}$ under the ARIMA (0,1,1) specification to identify an extrapolating function based on θ_λ . Phillips (Time Series Regression with a Unit Root, Econometrica, 1987) shows that under very general conditions on u_t in the model

$$\begin{aligned} y_t &= \phi y_{t-1} + u_t, \quad t = 1, 2, \dots, n \\ \phi &= 1, y_0 = 0. \end{aligned} \quad (1.8.2)$$

the OLS statistic

$$t_\phi = \left(n^{-1} \sum_1^n y_{t-1}^2 \right)^{1/2} (\hat{\phi} - 1) / s \quad (1.8.3)$$

where

$$\hat{\phi} = \frac{\sum_{t=1}^n y_t y_{t-1}}{\sum_{t=1}^n y_{t-1}^2} \quad (1.8.4)$$

and,

$$s^2 = \frac{\sum_{t=1}^n (y_t - \hat{\phi} y_{t-1})^2}{n} \quad (1.8.5)$$

satisfies

$$t_\phi \Rightarrow \left(\frac{\sigma}{2\sigma_u} \right) \left[W(1)^2 - \frac{\sigma_u^2}{\sigma^2} \right] / \left\{ \int_0^1 W(r)^2 dr \right\}^{1/2} \quad (1.8.6)$$

where

$$\begin{aligned} \sigma_u^2 &= \lim_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n E(u_t^2) \\ \sigma^2 &= \lim_{n \rightarrow \infty} E(n^{-1} S_n^2) \\ S_t &= \sum_{j=1}^t u_j \end{aligned} \quad (1.8.7)$$

Now suppose

$$u_t = \varepsilon_t + \theta \varepsilon_{t-1}, \quad \varepsilon_t \sim i.i.d.(0, \sigma^2) \quad (1.8.8)$$

then,

$$\sigma_u^2 = (1 + \theta^2) \sigma_\varepsilon^2 \quad (1.8.9)$$

$$\sigma^2 = (1 + \theta)^2 \sigma_\varepsilon^2$$

$$\therefore t_\phi \Rightarrow \left(\frac{(1+\theta)}{2\sqrt{1+\theta^2}} \right) \left[W(1)^2 - \frac{(1+\theta^2)}{(1+\theta)^2} \right] / \left\{ \int_0^1 W(r)^2 dr \right\}^{1/2} \quad (1.8.10)$$

If we let $\delta = \frac{(1+\theta)}{\sqrt{1+\theta^2}}$, then

$$t_\phi \Rightarrow \left(\frac{\delta}{2} \right) \left[W(1)^2 - \frac{1}{\delta^2} \right] / \left\{ \int_0^1 W(r)^2 dr \right\}^{1/2} \quad (1.8.11)$$

Phillips and Perron (1988) derived the limiting distribution of t_ϕ after centering the y_t values. In other words, the authors showed that the limiting distribution of $\hat{\tau}_{OLS}$ (1.2.5) is given by the following expression:

$$\hat{\tau}_{OLS} \Rightarrow \frac{\sigma}{\sigma_u} \frac{\left(\int W_* dW + \frac{1}{2} \right)}{\left[\int W_*^2 dr \right]^{1/2}} - \frac{\sigma_u}{2\sigma} \frac{1}{\left[\int W_*^2 dr \right]^{1/2}} \quad (1.8.12)$$

where

$$W_*(r) = W(r) - \int W(r) dr \quad (1.8.13)$$

Again, if we substitute the expressions for σ and σ_u into the above expression in (1.8.12), we obtain,

$$\hat{\tau}_{OLS} \Rightarrow \delta \frac{\left(\int W_* dW + \frac{1}{2} \right)}{\left[\int W_*^2 dr \right]^{1/2}} - \frac{1}{2\delta} \frac{1}{\left[\int W_*^2 dr \right]^{1/2}} \quad (1.8.14)$$

The above expressions suggest that we should attempt to model t_ϕ and $\hat{\tau}_{OLS}$ as linear functions of δ and $1/\delta$ in the Extrapolation-step of the SIMEX procedure and then use the estimated function to extrapolate back to $\delta = 1$. Note that now, $\theta = 0$ ($\lambda = -1$), and hence $\delta = 1$ corresponds to the case of no-measurement error (the simple AR (1) model).

Such an extrapolated value would give the SIMEX estimate of the OLS statistic based on the unobserved values h_t .

Table 1-30 provides the results of a simulation study. In this study we evaluate the performance of the simple linear regressions $\hat{\tau}_{OLS} = a + b\delta + c/\delta$ and $\hat{\tau}_{OLS} = b\delta + c/\delta$ as the extrapolating functions on the basis of 1000 replications. The frequency of rejection under the null hypothesis and the alternative hypothesis $\phi = 0.7$ is reported in the following table for various values of σ_η^2 . The ADF tests using 6 and 12 lags and the IV-based test are also appended to the table to facilitate a quick comparison.

All the tests listed below overestimate the nominal size (0.05) of the test when $\sigma_\eta^2 = 0.1$. When $\sigma_\eta^2 = 0.5$, the SIMEX-based test using $a + b\delta + c/\delta$ as the extrapolating function is the only test that maintains the size below the 0.05 level. As σ_η^2 increases, the ADF test (6 or 12 lags) maintains the nominal size of the test and delivers at least 55% power. The SIMEX-based tests grossly (in some cases) under-estimate the nominal size even when $\sigma_\eta^2 \geq 1$ and hence shows lack of power. However, at large values of σ_η^2 , the power of the SIMEX-based tests is much larger than that of the ADF tests.

Table 1-30: Comparison of ADF, IV and SIMEX-based tests for different values of σ_η^2 , $n = 100$

Test Statistic	ϕ_0									
	$\sigma_\eta^2 = 0.1$		$\sigma_\eta^2 = 0.5$		$\sigma_\eta^2 = 1$		$\sigma_\eta^2 = 10$		$\sigma_\eta^2 = 100$	
	1.00	0.70	1.00	0.70	1.00	0.70	1.00	0.70	1.00	0.70
<i>ADF using 12 lags</i>	0.30	0.70	0.23	0.23	0.03	0.21	0.04	0.22	0.04	0.21
<i>ADF using 6 (optimum) lags</i>	0.26	0.79	0.07	0.68	0.05	0.63	0.05	0.54	0.05	0.56
<i>$\hat{\tau}_{SIMEX,OLS}$ using $a + b\delta + c/\delta$ as the extrapolating function.</i>	0.18	0.28	0.02	0.16	0.02	0.09	0.02	0.24	0.03	0.94
<i>$\hat{\tau}_{SIMEX,OLS}$ using $b\delta + c/\delta$ as the extrapolating function</i>	1.00	1.00	0.93	1.00	0.76	1.00	0.03	0.86	0.03	0.96
<i>Instrumental Variable (IV) approach</i>	0.44	0.67	0.21	0.70	0.14	0.80	0.72	1.00	0.08	1.00

The tendency of the SIMEX-based tests to have better power as σ_{η}^2 increases can be explained if we look at the way in which σ_{η}^2 affects δ in Table 1-31.

Table 1-31: Relationship between σ_{η}^2 and δ

σ_{η}^2	θ	δ
0.10	-0.87	0.10
0.50	-0.73	0.22
1	-0.64	0.30
10	-0.27	0.71
100	-0.04	0.95

As $\sigma_{\eta}^2 \rightarrow \infty$, $\theta \rightarrow 0$ and $\delta \rightarrow 1$. In other words, as $\sigma_{\eta}^2 \rightarrow \infty$, the regression model is fitted through the pairs $(\delta, \hat{\tau}_{OLS})$ where δ is increasingly closer to 1 (the point at which $\hat{\tau}_{OLS}$ is extrapolated). We would thus expect our regression model to perform well in extrapolating the value of $\hat{\tau}_{OLS}$ at $\delta = 1$.

On the other hand, as $\sigma_{\eta}^2 \rightarrow 0$, $\theta \rightarrow 1$ and $\delta \rightarrow 0$. In this case, the point $\delta = 1$ is away from the cluster of measurement error-induced OLS ‘t’ statistics obtained at small values of δ . Hence the estimated regression model $\hat{\tau}_{OLS} = a + b\delta + c/\delta$ is unreliable in tracing the true regression path across all the higher values of δ close to 1. This affects the extrapolation at $\delta = 1$ which shows up in the columns $P(\text{Rej } H_0 | H_0)$ and $P(\text{Rej } H_0 | \phi = 0.7)$.

Weighted Symmetric Estimator:

Under the model (1.8.2), Phillips (1987) also shows the following:

$$n^{-1} \sum_{t=1}^n y_{t-1}^2 \Rightarrow \sigma^2 \int_0^1 W^2(r) dr \quad (1.8.15)$$

$$n^{-1} \sum_{t=1}^n y_{t-1} (y_t - y_{t-1}) \Rightarrow \frac{\sigma^2}{2} \left[W^2(1) - \frac{\sigma_u^2}{\sigma^2} \right] \quad (1.8.16)$$

The WS statistic under the model (1.8.2) is given by

$$\hat{\tau}_{ws} = \frac{\left[\sum_{t=2}^n y_{t-1} u_t + y_1^2 - \frac{1}{n} \sum_{t=1}^n y_t^2 \right]}{\hat{\sigma}_{ws} \left(\sum_{t=2}^{n-1} y_t^2 + \frac{1}{n} \sum_{t=1}^n y_t^2 \right)^{1/2}} \quad (1.8.17)$$

where $\hat{\sigma}_{ws}$ is given by (1.2.8).

From Fuller (1996), we know that $\hat{\sigma}_{ws}$ converges to σ_u^2 in probability. Dividing the numerator and denominator of expression (1.8.17) by n and then applying the results (1.8.15) and (1.8.16), we obtain

$$\begin{aligned} \hat{\tau}_{ws} &\Rightarrow \frac{\frac{\sigma^2}{2} \left[W^2(1) - \frac{\sigma_u^2}{\sigma^2} \right] - \sigma^2 \int_0^1 W^2(r) dr}{\sigma_u \left[\sigma^2 \int_0^1 W^2(r) dr \right]^{1/2}} \\ &\Rightarrow \frac{\frac{\sigma}{2\sigma_u} \left[W^2(1) - \frac{\sigma_u^2}{\sigma^2} \right] - \frac{\sigma}{\sigma_u} \int_0^1 W^2(r) dr}{\left[\int_0^1 W^2(r) dr \right]^{1/2}} \\ &\Rightarrow \frac{\frac{(1+\theta)}{2\sqrt{1+\theta^2}} \left[W^2(1) - \frac{1+\theta^2}{(1+\theta)^2} \right] - \frac{(1+\theta)}{\sqrt{1+\theta^2}} \int_0^1 W^2(r) dr}{\left[\int_0^1 W^2(r) dr \right]^{1/2}} \end{aligned} \quad (1.8.18)$$

$$\therefore \hat{\tau}_{ws} \Rightarrow \frac{\frac{1}{2\delta} W^2(1) - \frac{\delta}{2} - \frac{1}{\delta} \int_0^1 W^2(r) dr}{\left[\int_0^1 W^2(r) dr \right]^{1/2}} \quad (1.8.19)$$

The above expression suggests that the WS statistic is influenced by δ in much the same way as the OLS statistic is influenced by δ .

Although the regression function $\hat{\tau}_{OLS} = a + b\delta + c/\delta$ supported by asymptotic theory is not very useful for small values of σ_η^2 , there is scope for improvement by investigating the sample plots (Figure 1-8) of $\hat{\tau}_{OLS}$ values against the values of δ and then selecting a suitable non-linear function of δ .

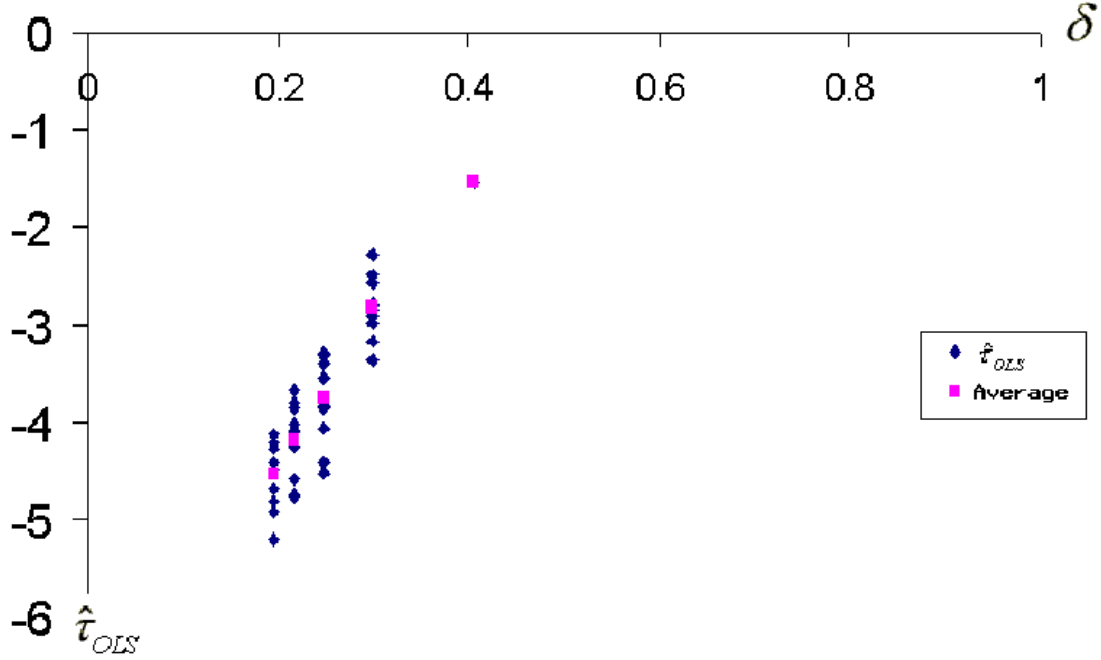


Figure 1-8: Plot of $\hat{\tau}_{OLS}$ against λ

In selecting a non-linear function, we again make use of the asymptotic theory. From equation (1.8.11), we see that as θ tends to -1 and hence δ tends to zero, the limiting distribution of $\hat{\tau}_{OLS}$ is mainly influenced by the term $\left[-\frac{1}{\delta}\right] / \left\{\int_0^1 W(r)^2 dr\right\}^{1/2}$. Hence we investigated the fit of the models

$$\hat{\tau}_{OLS} = \frac{1}{a + b\delta^c} + \varepsilon \quad (1.8.20)$$

and

$$\hat{\tau}_{OLS} = a + \frac{b}{\delta^c} + \varepsilon \quad (1.8.21)$$

to the simulated data sets. Use of any of the above non-linear extrapolants runs into problems because the convergence of the parameter estimates depends critically on the initial values of the estimates. We attempted to resolve this problem to some extent by forcing PROC NLIN in SAS (V8) to search through a grid of initial values and find the best starting value. The regression in (1.8.20) performed better than the one in (1.8.21), in terms of the error sums of squares and also in terms of maintaining the level of significance at the nominal level. So, we decide to focus on (1.8.20).

Seeking to overcome the initial value-problem for the convergence of the estimates, we reduced one parameter in the equation (1.8.20) by fixing a value for c . The choice of a value for c was narrowed down to a small range based on the good non-linear fits of (1.8.20). Simulation results were recorded for various values of c under various specifications of the SVM indexed by σ_η^2 . This eliminated the need to search for a different set of initial values for each simulation. If, a suitable initial value worked for any one of the simulated data sets generated by the SVM, then it also worked for all the simulated data sets. We further eliminated the need to specify an initial value by reducing the non-linear fitting problem to a linear regression of $1/\hat{\tau}_{OLS}$ on δ^c for a fixed value of c . Thus, the extrapolated value of t_ϕ at $\delta = 1$ using the model,

$$\frac{1}{\hat{\tau}_{OLS}} = a + b\delta^c + \varepsilon \quad (1.8.22)$$

gives the SIMEX estimate.

We ran the simulation study for different values of σ_η^2 . For each value of σ_η^2 , we ran 5000 replications of the testing procedure on a sample size of 100 observations under the null hypothesis ($\phi = 1$) and then under the alternative hypothesis ($\phi = 0.7$) using several extrapolation functions which are indexed by the parameter 'c'. The power under the null hypothesis and a particular alternative hypothesis are reported in the following tables.

Table 1-32: Power of SIMEX-based test using $a + b\delta^c$ as the extrapolating function
 $(\sigma_\eta^2 = 0.1, 0.2, 0.3), n = 100$

	$\sigma_\eta^2 = 0.1$		$\sigma_\eta^2 = 0.2$		$\sigma_\eta^2 = 0.3$	
	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$
$a + b\delta^1$	0.38	0.66	0.29	0.75	0.22	0.78
$a + b\delta^{1.1}$	0.31	0.55	0.23	0.66	0.18	0.69
$a + b\delta^{1.2}$	0.25	0.46	0.18	0.56	0.14	0.59
$a + b\delta^{1.3}$	0.20	0.38	0.15	0.48	0.11	0.51
$a + b\delta^{1.4}$	0.15	0.32	0.12	0.41	0.09	0.43
$a + b\delta^{1.5}$	0.12	0.26	0.09	0.34	0.07	0.36
$a + b\delta^{1.6}$	0.10	0.21	0.07	0.27	0.05	0.30

Table 1-33: Power of SIMEX-based test using $a + b\delta^c$ as the extrapolating function
 $(\sigma_\eta^2 = 0.4, 0.5, 0.6), n = 100$

	$\sigma_\eta^2 = 0.4$		$\sigma_\eta^2 = 0.5$		$\sigma_\eta^2 = 0.6$	
	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$
$a + b\delta^1$	0.18	0.79	0.15	0.79	0.13	0.78
$a + b\delta^{1.1}$	0.14	0.70	0.12	0.70	0.10	0.70
$a + b\delta^{1.2}$	0.11	0.61	0.09	0.61	0.08	0.61
$a + b\delta^{1.3}$	0.09	0.52	0.07	0.52	0.06	0.52
$a + b\delta^{1.4}$	0.07	0.44	0.06	0.44	0.05	0.44
$a + b\delta^{1.5}$	0.05	0.37	0.04	0.37	0.03	0.36
$a + b\delta^{1.6}$	0.04	0.30	0.03	0.30	0.02	0.30

Table 1-34: Power of SIMEX-based test using $a + b\delta^c$ as the extrapolating function
 $(\sigma_\eta^2 = 0.7, 0.8, 0.9), n = 100$

	$\sigma_\eta^2 = 0.7$		$\sigma_\eta^2 = 0.8$		$\sigma_\eta^2 = 0.9$	
	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$
$a + b\delta^1$	0.11	0.78	0.10	0.77	0.09	0.77
$a + b\delta^{1.1}$	0.09	0.70	0.08	0.69	0.07	0.69
$a + b\delta^{1.2}$	0.07	0.60	0.06	0.60	0.05	0.59
$a + b\delta^{1.3}$	0.05	0.51	0.04	0.51	0.04	0.50
$a + b\delta^{1.4}$	0.04	0.44	0.03	0.43	0.03	0.43
$a + b\delta^{1.5}$	0.03	0.36	0.02	0.36	0.02	0.35
$a + b\delta^{1.6}$	0.02	0.29	0.02	0.29	0.01	0.28

Table 1-35: Power of SIMEX-based test using $a + b\delta^c$ as the extrapolating function ($\sigma_\eta^2 = 1, 2, 3$), $n = 100$

	$\sigma_\eta^2 = 1$		$\sigma_\eta^2 = 2$		$\sigma_\eta^2 = 3$	
	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$
$a + b\delta^{0.1}$	0.38	1.00	0.18	0.99	0.11	0.99
$a + b\delta^{0.3}$	0.29	0.99	0.14	0.98	0.08	0.97
$a + b\delta^{0.5}$	0.21	0.98	0.10	0.95	0.06	0.92
$a + b\delta^{0.7}$	0.15	0.93	0.07	0.89	0.04	0.86
$a + b\delta^{0.9}$	0.10	0.84	0.04	0.79	0.02	0.75
$a + b\delta^1$	0.08	0.76	0.03	0.72	0.02	0.69
$a + b\delta^{1.2}$	0.05	0.59	0.02	0.56	0.01	0.55

Table 1-36: Power of SIMEX-based test using $a + b\delta^c$ as the extrapolating function ($\sigma_\eta^2 = 4, 5, 10$), $n = 100$

	$\sigma_\eta^2 = 4$		$\sigma_\eta^2 = 5$		$\sigma_\eta^2 = 10$	
	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$
$a + b\delta^{0.1}$	0.08	0.97	0.05	0.96	0.02	0.92
$a + b\delta^{0.3}$	0.06	0.95	0.04	0.93	0.02	0.89
$a + b\delta^{0.5}$	0.04	0.90	0.03	0.88	0.01	0.84
$a + b\delta^{0.7}$	0.02	0.83	0.02	0.81	0.01	0.79
$a + b\delta^{0.9}$	0.02	0.73	0.01	0.73	0.01	0.73
$a + b\delta^1$	0.01	0.68	0.01	0.68	0.01	0.70
$a + b\delta^{1.2}$	0.01	0.55	0.01	0.55	0.01	0.62

Table 1-37: Power of SIMEX-based test using $a + b\delta^c$ as the extrapolating function ($\sigma_\eta^2 = 50, 100$), $n = 100$

	$\sigma_\eta^2 = 50$		$\sigma_\eta^2 = 100$	
	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$
$a + b\delta^{0.1}$	0.03	0.93	0.03	0.96
$a + b\delta^{0.3}$	0.03	0.93	0.03	0.96
$a + b\delta^{0.5}$	0.03	0.92	0.03	0.96
$a + b\delta^{0.7}$	0.03	0.92	0.03	0.96
$a + b\delta^{0.9}$	0.03	0.91	0.03	0.96
$a + b\delta^1$	0.03	0.91	0.03	0.96

$a + b\delta^{1.2}$	0.02	0.91	0.03	0.95
---------------------	------	------	------	------

The objective of the above simulation study was to identify if possible, a single extrapolating function that would at the very least, maintain the size at 0.05 irrespective of the value of σ_{η}^2 . The study suggests that we don't have such an extrapolating function among the ones that we have considered. It suggests that we have to vary the parameter c in the function $a + b\delta^c$ according to value of σ_{η}^2 so that the size of the test is maintained at the nominal level of 0.05. However if we only require the estimated size of the test to be at or below 0.05, we can then narrow the set of appropriate extrapolating functions down to a few:

(1) $\sigma_{\eta}^2 \geq 1$:

The extrapolant $a + b\delta^{1.2}$ maintains the size of the test at or below the nominal size 0.05 and has at least 55% power against the alternative hypothesis.

Table 1-38: Performance of $a + b\delta^{1.2}$ for different values of $\sigma_{\eta}^2 \geq 1$, $n = 100$

σ_{η}^2		$\phi = 1$	$\phi = 0.7$
1	$a + b\delta^{1.2}$	0.05	0.59
2	$a + b\delta^{1.2}$	0.02	0.56
3	$a + b\delta^{1.2}$	0.01	0.55
4	$a + b\delta^{1.2}$	0.01	0.55
5	$a + b\delta^{1.2}$	0.01	0.55
10	$a + b\delta^{1.2}$	0.01	0.62
50	$a + b\delta^{1.2}$	0.02	0.91
100	$a + b\delta^{1.2}$	0.03	0.95

(2) $\sigma_{\eta}^2 < 1$

As σ_{η}^2 decreases, the extrapolant $a + b\delta^c$ requires increasing values of 'c' for the tests to deliver the right level. None of our extrapolating function work when σ_{η}^2 is as low as 0.2

or lower. If we use $a + b\delta^{1.2}$ as the extrapolating function irrespective of the value of σ_η^2 , then the size is overestimated by an amount that increases as σ_η^2 approaches zero. We collate the results for $a + b\delta^{1.2}$ in the following table:

Table 1-39: Performance of $a + b\delta^{1.2}$ for different values of $\sigma_\eta^2 < 1$, $n = 100$

σ_η^2		$\phi = 1$	$\phi = 0.7$
0.3	$a + b\delta^{1.2}$	0.14	0.59
0.4	$a + b\delta^{1.2}$	0.11	0.61
0.5	$a + b\delta^{1.2}$	0.09	0.61
0.6	$a + b\delta^{1.2}$	0.08	0.61
0.7	$a + b\delta^{1.2}$	0.07	0.60
0.8	$a + b\delta^{1.2}$	0.06	0.60
0.9	$a + b\delta^{1.2}$	0.05	0.59

The simulation results demonstrate the applicability of $a + b\delta^{1.2}$ as the extrapolating function over a wide range of σ_η^2 values. We could thus, in practice, perform the SIMEX test, using $a + b\delta^{1.2}$ as the extrapolating function irrespective of the value of σ_η^2 .

Section 9. Performance of SIMEX-based test when σ_η^2 is estimated

We outlined a procedure to conduct the SIMEX-based unit root test at the end of the last section. The performance of the extrapolating function $a + b\delta^{1.2}$ was evaluated under the assumption that the value of σ_η^2 would be known. In this section, we will relax this assumption and re-evaluate the performance of the various extrapolating functions, now based on an estimated value of σ_η^2 .

First, we describe the procedure of estimating σ_η^2 . Under the null hypothesis of a unit root in the h_t process, the first difference of the $\log(r_t^2)$ process has a moving average

(order 1) representation. The moving average parameter in terms of σ_η^2 is obtained by putting $\lambda = 0$ in equation (1.8.1). We fed the observed series $\log(r_t^2)$ as input to PROC ARIMA in SAS (V8) to get an estimate of the moving average parameter using the METHOD = ML (Maximum Likelihood) option in the ESTIMATE statement. The estimate $\hat{\theta}$ produced by SAS, and $\lambda = 0$ are put into equation (1.8.1) to back-solve for $\hat{\sigma}_\eta^2$. Next, the same equation uses the value of $\hat{\sigma}_\eta^2$ and successive values of λ to get $\hat{\theta}_\lambda$ and hence the corresponding $\hat{\delta}_\lambda$. Now that it is clear that δ takes on the value $\hat{\delta}_\lambda$ as λ varies, we will use δ and $\hat{\delta}_\lambda$ interchangeably in this section to simplify notations. We evaluated the performance of several extrapolating functions. The rejection rates are provided in the following tables:

Table 1-40: Power of SIMEX-based test using $a + b\hat{\delta}_\lambda^c$ as extrapolating function
 $(\sigma_\eta^2 = 0.1, 0.2, 0.3), n = 100$

	$\sigma_\eta^2 = 0.1$		$\sigma_\eta^2 = 0.2$		$\sigma_\eta^2 = 0.3$	
	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$
$a + b\delta^{1.1}$	0.28	0.23	0.21	0.23	0.16	0.34
$a + b\delta^{1.2}$	0.21	0.17	0.15	0.17	0.11	0.26
$a + b\delta^{1.3}$	0.15	0.13	0.10	0.13	0.07	0.19
$a + b\delta^{1.4}$	0.11	0.10	0.07	0.10	0.04	0.14
$a + b\delta^{1.5}$	0.09	0.07	0.04	0.07	0.03	0.09
$a + b\delta^{1.6}$	0.06	0.05	0.04	0.05	0.02	0.07
$a + b\delta^{1.7}$	0.04	0.04	0.03	0.04	0.01	0.05

Table 1-41: Power of SIMEX-based test using $a + b\hat{\delta}_\lambda^c$ as extrapolating function
 $(\sigma_\eta^2 = 0.4, 0.5, 0.6), n = 100$

	$\sigma_\eta^2 = 0.4$		$\sigma_\eta^2 = 0.5$		$\sigma_\eta^2 = 0.6$	
	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$
$a + b\delta^1$	0.12	0.38	0.10	0.41	0.09	0.44
$a + b\delta^{1.1}$	0.08	0.29	0.07	0.31	0.06	0.34
$a + b\delta^{1.2}$	0.05	0.21	0.04	0.23	0.03	0.25
$a + b\delta^{1.3}$	0.03	0.15	0.03	0.16	0.02	0.18
$a + b\delta^{1.4}$	0.02	0.10	0.02	0.11	0.01	0.12
$a + b\delta^{1.5}$	0.01	0.07	0.01	0.07	0.01	0.08
$a + b\delta^{1.6}$	0.01	0.04	0.01	0.05	0.01	0.05

Table 1-42: Power of SIMEX-based test using $a + b\hat{\delta}_\lambda^c$ as extrapolating function
 $(\sigma_\eta^2 = 0.7, 0.8, 0.9, 1), n = 100$

	$\sigma_\eta^2 = 0.7$		$\sigma_\eta^2 = 0.8$		$\sigma_\eta^2 = 0.9$		$\sigma_\eta^2 = 1$	
	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$
$a + b\delta^1$	0.07	0.47	0.06	0.49	0.06	0.51	0.05	0.53
$a + b\delta^{1.1}$	0.05	0.36	0.04	0.39	0.04	0.41	0.04	0.43
$a + b\delta^{1.2}$	0.03	0.27	0.03	0.29	0.03	0.31	0.02	0.33
$a + b\delta^{1.3}$	0.02	0.20	0.02	0.21	0.02	0.23	0.01	0.25
$a + b\delta^{1.4}$	0.01	0.14	0.01	0.15	0.01	0.17	0.01	0.18
$a + b\delta^{1.5}$	0.01	0.10	0.01	0.11	0.01	0.12	0.01	0.13
$a + b\delta^{1.6}$	0.01	0.07	0.00	0.07	0.00	0.08	0.00	0.09

Table 1-43: Power of SIMEX-based test using $a + b\hat{\delta}_\lambda^c$ as extrapolating function ($\sigma_\eta^2 = 2, 3$), $n = 100$

	$\sigma_\eta^2 = 2$		$\sigma_\eta^2 = 3$	
	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$
$a + b\delta^{0.1}$	0.22	1.00	0.14	1.00
$a + b\delta^{0.3}$	0.17	0.99	0.11	0.99
$a + b\delta^{0.5}$	0.12	0.96	0.09	0.96
$a + b\delta^{0.7}$	0.08	0.88	0.06	0.91
$a + b\delta^{0.9}$	0.05	0.76	0.04	0.82
$a + b\delta^1$	0.04	0.68	0.03	0.76
$a + b\delta^{1.2}$	0.02	0.52	0.02	0.62

Table 1-44: Power of SIMEX-based test using $a + b\hat{\delta}_\lambda^c$ as extrapolating function ($\sigma_\eta^2 = 4, 5, 8$), $n = 100$

	$\sigma_\eta^2 = 4$		$\sigma_\eta^2 = 5$		$\sigma_\eta^2 = 8$	
	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$
$a + b\delta^{0.1}$	0.11	1.00	0.09	0.99	0.06	0.98
$a + b\delta^{0.3}$	0.09	0.99	0.07	0.98	0.05	0.97
$a + b\delta^{0.5}$	0.07	0.96	0.06	0.96	0.05	0.96
$a + b\delta^{0.7}$	0.06	0.92	0.05	0.93	0.04	0.93
$a + b\delta^{0.9}$	0.04	0.85	0.04	0.87	0.04	0.90
$a + b\delta^1$	0.03	0.81	0.03	0.84	0.03	0.88
$a + b\delta^{1.2}$	0.03	0.70	0.03	0.76	0.03	0.82

Table 1-45: Power of SIMEX-based test using $a + b\hat{\delta}_\lambda^c$ as extrapolating function ($\sigma_\eta^2 = 10, 100$), $n = 100$

	$\sigma_\eta^2 = 10$		$\sigma_\eta^2 = 100$	
	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$
$a + b\delta^{0.1}$	0.05	0.98	0.03	0.95
$a + b\delta^{0.3}$	0.04	0.97	0.03	0.95
$a + b\delta^{0.5}$	0.04	0.95	0.03	0.94
$a + b\delta^{0.7}$	0.04	0.93	0.03	0.94
$a + b\delta^{0.9}$	0.03	0.90	0.03	0.93
$a + b\delta^1$	0.03	0.88	0.03	0.93
$a + b\delta^{1.2}$	0.03	0.84	0.03	0.92

When $\sigma_{\eta}^2 \geq 1$, the function $a + b\delta$ does remarkably well as the SIMEX extrapolating function. The estimated size of the test based on this extrapolating function is between 0.03 and 0.05 and the power of test increases from 0.53 when $\sigma_{\eta}^2 = 1$ to more than 90% when $\sigma_{\eta}^2 = 100$. If we use the same function in situations where $\sigma_{\eta}^2 < 1$, the size tends to be overestimated and the error in estimation increases as σ_{η}^2 approaches zero. Instead of $a + b\delta$, if we use $a + b\delta^{1.2}$ as the extrapolating function, there is some loss of power. However, the size of the test is controlled at below the 0.05 level for values of σ_{η}^2 as small as 0.4. The following tables compare the performance of $a + b\delta$ against $a + b\delta^{1.2}$ across various values of σ_{η}^2 .

Table 1-46: Comparing $a + b\hat{\delta}_{\lambda}$ and $a + b\hat{\delta}_{\lambda}^{1.2}$ as extrapolating functions, $n = 100$

	$a + b\hat{\delta}_{\lambda}$		$a + b\hat{\delta}_{\lambda}^{1.2}$	
σ_{η}^2	$\phi = 1$	$\phi = 0.7$	$\phi = 1$	$\phi = 0.7$
0.1	0.28	0.23	0.15	0.13
0.2	0.21	0.23	0.10	0.13
0.3	0.16	0.34	0.07	0.19
0.4	0.12	0.38	0.05	0.21
0.5	0.10	0.41	0.04	0.23
0.6	0.09	0.44	0.03	0.25
0.7	0.07	0.47	0.03	0.27
0.8	0.06	0.49	0.03	0.29
0.9	0.06	0.51	0.03	0.31
1.0	0.05	0.53	0.02	0.33
2.0	0.04	0.68	0.02	0.52
3.0	0.03	0.76	0.02	0.62
4.0		0.81		0.70
5.0	0.03	0.84	0.03	0.76
10.0	0.03	0.88	0.03	0.84
100.0	0.03	0.93	0.03	0.92

There is a clear indication that it is desirable to use $a + b\delta$ as the extrapolating function when $\sigma_{\eta}^2 \geq 1$ and to use $a + b\delta^{1.2}$ as the extrapolating function when $\sigma_{\eta}^2 < 1$. If we could selectively use one of these extrapolating functions on the basis of $\hat{\sigma}_{\eta}^2$, we would probably be better off than using exactly one of them exclusively.

Section 10. Combining the two promising extrapolating functions

In this section, we evaluate a host of rules, each of which states a criterion to choose one of the two extrapolating functions $a + b\delta$ or $a + b\delta^{1.2}$ based on the estimate $\hat{\sigma}_{\eta}^2$. Each rule is in essence a separate SIMEX-based test. The rules are laid out in the following table and are numbered from 1 through 20.

Table 1-47: Different rules for combining the extrapolating functions $a + b\hat{\delta}_\lambda$ and $a + b\hat{\delta}_\lambda^{1.2}$

Rule	Use Extrapolating Function
1	$(a + b\delta)I_{[0.1,\infty)}(\hat{\sigma}_\eta^2) + (a + b\delta^{1.1})I_{(-\infty,0.1)}(\hat{\sigma}_\eta^2)$
2	$(a + b\delta)I_{[0.2,\infty)}(\hat{\sigma}_\eta^2) + (a + b\delta^{1.1})I_{(-\infty,0.2)}(\hat{\sigma}_\eta^2)$
3	$(a + b\delta)I_{[0.3,\infty)}(\hat{\sigma}_\eta^2) + (a + b\delta^{1.1})I_{(-\infty,0.3)}(\hat{\sigma}_\eta^2)$
4	$(a + b\delta)I_{[0.4,\infty)}(\hat{\sigma}_\eta^2) + (a + b\delta^{1.1})I_{(-\infty,0.4)}(\hat{\sigma}_\eta^2)$
5	$(a + b\delta)I_{[0.5,\infty)}(\hat{\sigma}_\eta^2) + (a + b\delta^{1.1})I_{(-\infty,0.5)}(\hat{\sigma}_\eta^2)$
6	$(a + b\delta)I_{[0.6,\infty)}(\hat{\sigma}_\eta^2) + (a + b\delta^{1.1})I_{(-\infty,0.6)}(\hat{\sigma}_\eta^2)$
7	$(a + b\delta)I_{[0.7,\infty)}(\hat{\sigma}_\eta^2) + (a + b\delta^{1.1})I_{(-\infty,0.7)}(\hat{\sigma}_\eta^2)$
8	$(a + b\delta)I_{[0.8,\infty)}(\hat{\sigma}_\eta^2) + (a + b\delta^{1.1})I_{(-\infty,0.8)}(\hat{\sigma}_\eta^2)$
9	$(a + b\delta)I_{[0.9,\infty)}(\hat{\sigma}_\eta^2) + (a + b\delta^{1.1})I_{(-\infty,0.9)}(\hat{\sigma}_\eta^2)$
10	$(a + b\delta)I_{[0.1,\infty)}(\hat{\sigma}_\eta^2) + (a + b\delta^{1.2})I_{(-\infty,0.1)}(\hat{\sigma}_\eta^2)$
11	$(a + b\delta)I_{[0.2,\infty)}(\hat{\sigma}_\eta^2) + (a + b\delta^{1.2})I_{(-\infty,0.2)}(\hat{\sigma}_\eta^2)$
12	$(a + b\delta)I_{[0.3,\infty)}(\hat{\sigma}_\eta^2) + (a + b\delta^{1.2})I_{(-\infty,0.3)}(\hat{\sigma}_\eta^2)$
13	$(a + b\delta)I_{[0.4,\infty)}(\hat{\sigma}_\eta^2) + (a + b\delta^{1.2})I_{(-\infty,0.4)}(\hat{\sigma}_\eta^2)$
14	$(a + b\delta)I_{[0.5,\infty)}(\hat{\sigma}_\eta^2) + (a + b\delta^{1.2})I_{(-\infty,0.5)}(\hat{\sigma}_\eta^2)$
15	$(a + b\delta)I_{[0.6,\infty)}(\hat{\sigma}_\eta^2) + (a + b\delta^{1.2})I_{(-\infty,0.6)}(\hat{\sigma}_\eta^2)$
16	$(a + b\delta)I_{[0.7,\infty)}(\hat{\sigma}_\eta^2) + (a + b\delta^{1.2})I_{(-\infty,0.7)}(\hat{\sigma}_\eta^2)$
17	$(a + b\delta)I_{[0.8,\infty)}(\hat{\sigma}_\eta^2) + (a + b\delta^{1.2})I_{(-\infty,0.8)}(\hat{\sigma}_\eta^2)$
18	$(a + b\delta)I_{[0.9,\infty)}(\hat{\sigma}_\eta^2) + (a + b\delta^{1.2})I_{(-\infty,0.9)}(\hat{\sigma}_\eta^2)$

The estimated size and power of the each of these 18 tests based on 5000 replications are presented in the following tables. The power and size of the tests based entirely on $a + b\delta$ or $a + b\delta^{1.2}$ irrespective of the value of σ_η^2 are also provided in the same tables. We note that the power and the size of these 18 tests are virtually the same as that of the SIMEX test based entirely on $a + b\delta$ when $\sigma_\eta^2 \geq 1$. However, Rule 18 and Rule 19 beat the test based entirely on $a + b\delta$ in terms of maintaining the size of the test at 0.05 when σ_η^2 is as low as 0.5. Based on these simulated power comparisons, we would recommend using Rule 17 in the SIMEX procedure. Thus one would estimate σ_η^2 and then if $\hat{\sigma}_\eta^2$ is less than 0.8, one would use $a + b\delta^{1.2}$ as the extrapolating function, otherwise one would use $a + b\delta$ as the extrapolating function.

Table 1-48: Estimated size of the SIMEX test using different rules , $n = 100$

	σ_{η}^2											
Rule	0.1	0.3	0.5	0.7	0.9	1	2	3	4	5	10	100
1	0.24	0.15	0.1	0.07	0.06	0.05	0.04	0.03	0.03	0.03	0.03	0.03
2	0.22	0.14	0.1	0.07	0.06	0.05	0.04	0.03	0.03	0.03	0.03	0.03
3	0.21	0.13	0.09	0.07	0.06	0.05	0.04	0.03	0.03	0.03	0.03	0.03
4	0.21	0.12	0.09	0.07	0.06	0.05	0.04	0.03	0.03	0.03	0.03	0.03
5	0.21	0.12	0.08	0.06	0.05	0.05	0.04	0.03	0.03	0.03	0.03	0.03
6	0.21	0.11	0.08	0.06	0.05	0.05	0.04	0.03	0.03	0.03	0.03	0.03
7	0.21	0.11	0.07	0.06	0.05	0.05	0.04	0.03	0.03	0.03	0.03	0.03
8	0.21	0.11	0.07	0.06	0.05	0.05	0.04	0.03	0.03	0.03	0.03	0.03
9	0.21	0.11	0.07	0.06	0.05	0.05	0.04	0.03	0.03	0.03	0.03	0.03
10	0.21	0.15	0.1	0.07	0.06	0.05	0.04	0.03	0.03	0.03	0.03	0.03
11	0.17	0.13	0.09	0.07	0.06	0.05	0.04	0.03	0.03	0.03	0.03	0.03
12	0.16	0.11	0.08	0.06	0.06	0.05	0.04	0.03	0.03	0.03	0.03	0.03
13	0.15	0.1	0.08	0.06	0.05	0.05	0.04	0.03	0.03	0.03	0.03	0.03
14	0.15	0.09	0.07	0.06	0.05	0.05	0.04	0.03	0.03	0.03	0.03	0.03
15	0.15	0.08	0.06	0.05	0.05	0.05	0.04	0.03	0.03	0.03	0.03	0.03
16	0.15	0.08	0.06	0.05	0.05	0.05	0.04	0.03	0.03	0.03	0.03	0.03
17	0.15	0.08	0.05	0.05	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03
18	0.15	0.08	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03
$a + b\hat{\delta}_{\lambda}^{1,2}$	0.15	0.07	0.04	0.03	0.03	0.02	0.02	0.02	0.02	0.03	0.03	0.03
$a + b\hat{\delta}_{\lambda}$	0.28	0.16	0.1	0.07	0.06	0.05	0.04	0.03	0.03	0.03	0.03	0.03

Table 1-49: Estimated power of the SIMEX test using different rules at $H_1 : \phi = 0.7$, $n = 100$

Rule	σ_{η}^2											
	0.1	0.3	0.5	0.7	0.9	1	2	3	4	5	10	100
1	0.19	0.31	0.39	0.46	0.51	0.53	0.68	0.76	0.81	0.84	0.88	0.93
2	0.18	0.30	0.37	0.45	0.50	0.52	0.68	0.76	0.81	0.84	0.88	0.93
3	0.18	0.28	0.36	0.43	0.49	0.51	0.67	0.76	0.81	0.84	0.88	0.93
4	0.18	0.27	0.35	0.42	0.48	0.50	0.67	0.76	0.81	0.84	0.88	0.93
5	0.17	0.26	0.34	0.41	0.47	0.49	0.67	0.76	0.81	0.84	0.88	0.93
6	0.17	0.26	0.34	0.40	0.46	0.48	0.67	0.76	0.81	0.84	0.88	0.93
7	0.17	0.26	0.33	0.40	0.45	0.47	0.66	0.75	0.81	0.84	0.88	0.93
8	0.17	0.26	0.33	0.39	0.45	0.47	0.66	0.75	0.81	0.84	0.88	0.93
9	0.17	0.26	0.32	0.38	0.44	0.46	0.65	0.75	0.81	0.84	0.88	0.93
10	0.16	0.29	0.38	0.45	0.50	0.53	0.68	0.76	0.81	0.84	0.88	0.93
11	0.14	0.26	0.35	0.43	0.49	0.51	0.68	0.76	0.81	0.84	0.88	0.93
12	0.14	0.24	0.33	0.41	0.48	0.50	0.67	0.76	0.81	0.84	0.88	0.93
13	0.13	0.22	0.31	0.39	0.46	0.49	0.67	0.76	0.81	0.84	0.88	0.93
14	0.13	0.21	0.29	0.37	0.44	0.47	0.66	0.75	0.81	0.84	0.88	0.93
15	0.13	0.20	0.28	0.35	0.42	0.45	0.65	0.75	0.81	0.84	0.88	0.93
16	0.13	0.20	0.27	0.34	0.41	0.43	0.65	0.75	0.81	0.84	0.88	0.93
17	0.13	0.19	0.26	0.33	0.39	0.42	0.64	0.75	0.81	0.84	0.88	0.93
18	0.13	0.19	0.25	0.32	0.38	0.41	0.63	0.74	0.80	0.84	0.88	0.93
$a + b\hat{\delta}_{\lambda}^{1,2}$	0.13	0.19	0.23	0.27	0.31	0.33	0.51	0.62	0.70	0.76	0.84	0.92
$a + b\hat{\delta}_{\lambda}^{\infty}$	0.23	0.34	0.41	0.47	0.51	0.53	0.68	0.76	0.81	0.84	0.88	0.93

Rule 18 provides virtually 100% power against the alternative $\phi = 0.70$ when the sample size is 1000. It would be more useful to look at the power of the test given by Rule 18 for a sample size of 1000 against $\phi = 0.99$ when σ_{η}^2 is 0.1 or lower. In the following table, we report the power and size of all the rules enlisted in Table 1-36 on the basis of 1000 replications.

Table 1-50: Power and Size of various rules when Sample Size $n = 1000$ and $0.01 \leq \sigma_\eta^2 \leq 0.10$

	σ_η^2											
	0.01		0.03		0.05		0.07		0.09		0.1	
	ϕ		ϕ		ϕ		ϕ		ϕ		ϕ	
	0.99	1	0.99	1	0.99	1	0.99	1	0.99	1	0.99	1
Rule												
1	0.87	0.36	0.82	0.19	0.73	0.13	0.66	0.11	0.66	0.10	0.67	0.10
2	0.87	0.36	0.82	0.19	0.73	0.13	0.64	0.10	0.58	0.09	0.56	0.07
3	0.87	0.36	0.82	0.19	0.73	0.13	0.64	0.10	0.58	0.09	0.56	0.07
4	0.87	0.36	0.82	0.19	0.73	0.13	0.64	0.10	0.58	0.09	0.56	0.07
5	0.87	0.36	0.82	0.19	0.73	0.13	0.64	0.10	0.58	0.09	0.56	0.07
6	0.87	0.36	0.82	0.19	0.73	0.13	0.64	0.10	0.58	0.09	0.56	0.07
7	0.87	0.36	0.82	0.19	0.73	0.13	0.64	0.10	0.58	0.09	0.56	0.07
8	0.87	0.36	0.82	0.19	0.73	0.13	0.64	0.10	0.58	0.09	0.56	0.07
9	0.87	0.36	0.82	0.19	0.73	0.13	0.64	0.10	0.58	0.09	0.56	0.07
10	0.67	0.20	0.57	0.09	0.47	0.06	0.44	0.05	0.51	0.06	0.57	0.08
11	0.67	0.20	0.57	0.09	0.47	0.06	0.39	0.04	0.35	0.03	0.34	0.03
12	0.67	0.20	0.57	0.09	0.47	0.06	0.39	0.04	0.35	0.03	0.34	0.03
13	0.67	0.20	0.57	0.09	0.47	0.06	0.39	0.04	0.35	0.03	0.34	0.03
14	0.67	0.20	0.57	0.09	0.47	0.06	0.39	0.04	0.35	0.03	0.34	0.03
15	0.67	0.20	0.57	0.09	0.47	0.06	0.39	0.04	0.35	0.03	0.34	0.03
16	0.67	0.20	0.57	0.09	0.47	0.06	0.39	0.04	0.35	0.03	0.34	0.03
17	0.67	0.20	0.57	0.09	0.47	0.06	0.39	0.04	0.35	0.03	0.34	0.03
18	0.67	0.20	0.57	0.09	0.47	0.06	0.39	0.04	0.35	0.03	0.34	0.03
$a + b\hat{\delta}_\lambda^{1.2}$	0.67	0.20	0.57	0.09	0.47	0.06	0.39	0.04	0.35	0.03	0.34	0.03
$a + b\hat{\delta}_\lambda$	0.95	0.54	0.94	0.32	0.90	0.23	0.85	0.18	0.80	0.16	0.78	0.14

Rules 10 through 18 and $a + b\hat{\delta}_\lambda^{1.2}$ have the same performance across the board. They maintain the size at or below 0.05 (nominal level) for $\sigma_\eta^2 = 0.05, 0.07, 0.09$ and 0.10 and deliver around 35% to 40% power against $\phi = 0.99$.

Section 11. Conclusion

In this chapter, we have derived a unit root test for the SVM using the SIMEX method—a technique used to assess the effect of measurement error on parameter estimates. Drawing from asymptotic theory and representative plots of the statistics against artificially induced measurement error variance, we were able to determine the nature of the appropriate extrapolating function. The SIMEX-based test is unable to control the nominal size if σ_{η}^2 is too small for a given sample size. Such a feature is common to all the unit root tests for SVM available in the literature. The simulation studies provide evidence of superior performance of the SIMEX-based test compared to the Instrumental Variables approach. Our test also compares favorably with the ADF approach in certain aspects. When using the ADF, choice of the appropriate number of lags for a data set depends critically upon the estimated σ_{η}^2 . Too many lags would result in loss of power and too few lags would result in over-estimation of the nominal size. Our method provides a single extrapolating function that can be used for any data set. Also, we showed that the SIMEX-based test has better power than the ADF test under many alternative hypotheses when the choice of lags for the ADF test is based on certain sample size-based rules.

Chapter 2 A unit root test based on the maximum order statistic

Section 1. Introduction

Fuller (1976), and Dickey and Fuller (1979) introduced the Dickey-Fuller t test (DF test) and the augmented DF (ADF) test for testing the unit root hypothesis in the autoregressive moving average (ARMA) representation of a time series. Phillips (1987), and Phillips and Perron (1988) showed that these tests based on the ordinary least squares (OLS) estimator have some desirable robustness properties. The asymptotic properties of the OLS estimator in the non-stationary time series setting proved to be quite robust to considerable amount of heterogeneity and dependence among the innovations. However, least squares-based tests may suffer a loss of power relative to other methods when the innovations are not normal. In the recent past, researchers have attempted to develop robust tests that are designed to have good power for different error distributions. Hasan and Koenkar (1997) proposed a family of rank tests based on the regression rank score process to test the unit root hypothesis in economic time series. Simulations in Lucas (1995) showed that the tests based on M-estimators were more powerful than OLS-based tests if the innovations were thick-tailed. Rothenberg and Stock (1997) developed robust tests and estimators based on non-normal quasi-likelihood functions for autoregressive models with near unit root. They demonstrated the sensitivity of the power function to the extent of non-normality in the innovations.

The present paper proposes a test statistic that standardizes the maximum of the squared series by the OLS-based error variance estimate, to test for the unit root hypothesis in the AR(1) model. Cavaliere (2001) used the difference in the maximum and the minimum order statistic to test for a unit root. He generalized the rescaled range statistic introduced by Hurst (1951) and used it to test for a unit root in time series models. Testing procedures based on the generalized rescaled range statistics were found to be consistent against $I(0)$, $I(2)$ as well as fractional integration and misspecification of the

deterministic trend. Our proposed statistic seeks to differentiate between a unit-root series and a stationary series by just picking up the most extreme value in the series. One would expect such a test statistic to be robust to departures from Normality in the innovations process. At the same time, one would expect it to be strongly affected by the presence of outliers. However, the standardization of the maximum order statistic by the OLS estimate of the error variance might deliver better power than the OLS ‘t’ statistic under some non-normal specifications of the error distributions. We attempt to investigate such possibilities in this paper. The next section introduces the test statistic. The finite sample percentiles and the approximate asymptotic percentiles are tabulated Section 3. In Section 4, we compare the size and power properties of the proposed test and the DF test under different error specifications including skewed, symmetric and thick-tailed-distributions. The current study is concluded in Section 5.

Section 2. The test statistic

Consider the AR(1) model for y_t :

$$\begin{aligned} y_t &= \phi y_{t-1} + \varepsilon_t, & t = 1, 2, \dots, n \\ y_0 &= 0, \\ \varepsilon_t &\sim NID(0, \sigma^2) \end{aligned} \quad (2.2.1)$$

For the hypothesis testing problem:

$$H_0 : \phi = 1 \quad \text{v/s} \quad H_1 : |\phi| < 1 \quad (2.2.2)$$

we consider the following statistic:

$$\tau = \frac{\left(\max_{1 \leq t \leq n} |y_t| \right)^2}{\sum_{t=2}^n \left(y_t - \hat{\phi}_{OLS} y_{t-1} \right)^2} \quad (2.2.3)$$

where,

$$\hat{\phi}_{OLS} = \left(\sum_{t=2}^n y_{t-1}^2 \right)^{-1} \sum_{t=2}^n y_{t-1} y_t$$

Under the null hypothesis, one would expect the numerator in the expression for τ to assume larger values than under the alternative stationary hypothesis and hence one would like to reject the unit root hypothesis for “small” values of τ . Our decision criterion would be to reject H_0 if $\tau < c$ where c is such that $P(\tau < c | H_0) = \alpha$, the nominal significance level.

Section 3. Finite Sample and Asymptotic Percentiles

To determine c for a given value of α , we simulated the distribution of τ under the null hypothesis for various sample sizes. Using the RANNOR function in SAS (V8), we generated the standard normal variates ε_t ($\sigma^2 = 1$). Here are the 5th and the 95th percentiles of the distribution of τ under the unit root hypothesis. The percentiles are based on 10,000 Monte-Carlo replications.

Table 2-1: Percentiles from the Simulated distribution of τ

Statistic	5 th percentile	95 th percentile
$n = 25$	0.33	4.72
$n = 100$	0.33	4.89
$n = 1000$	0.37	4.88

Based on the above percentiles, we evaluated the power performance of τ under the alternatives $\phi = 0.95, 0.98, 0.99$. The power of the DF ‘t’ test $\hat{\tau}_{DF}$ (without an intercept) under these alternatives is also listed below to facilitate a quick comparison of the two test statistics. The 5th percentile of the finite sample distribution of $\hat{\tau}_{DF}$ has been computed to be -1.95 in Fuller (1996).

Table 2-2: Power Comparison for $n = 25$

Power	ϕ			
	0.95	0.98	0.99	1
$P(\hat{\tau}_{DF} < -1.95)$	0.092	0.067	0.059	0.050
$P(\tau < 0.33)$	0.091	0.064	0.058	0.050

Table 2-3: Power Comparison for $n = 100$

Power	ϕ			
	0.95	0.98	0.99	1
$P(\hat{\tau}_{DF} < -1.95)$	0.313	0.112	0.067	0.050
$P(\tau < 0.33)$	0.280	0.112	0.076	0.050

Table 2-4: Power Comparison for $n = 1000$

Power	ϕ			
	0.95	0.98	0.99	1
$P(\hat{\tau}_{DF} < -1.95)$	1.000	0.992	0.762	0.050
$P(\tau < 0.37)$	1.000	0.977	0.688	0.050

The first observation that we can make from the above simulation study is that the percentiles for τ seem to be converging as the sample size increases without bound. We can therefore expect to be able to derive the limiting distribution of the statistic. Secondly, as one would expect of any useful test statistic, the power of the test statistic under any alternative does seem to converge to 1 as the sample size increases without bound. Finally, the simulation study shows that the DF test is slightly more powerful than τ when the innovations are normal. In order to assess the usefulness of τ in situations where the innovations are not normal, we also need to compute the asymptotic percentiles of τ so that we may use these percentiles instead of the finite sample percentiles based on normal innovations.

We now derive the asymptotic distribution of the statistic τ under the random-walk model:

$$\begin{aligned} y_t &= \phi y_{t-1} + e_t, \quad t = 1, 2, \dots, n. \\ \phi &= 1, \quad y_0 = 0. \\ e_t &\sim i.i.d.(0, \sigma^2). \end{aligned} \tag{2.3.1}$$

Theorem: Under the model given by (2.3.1),

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\left(\text{Max}_{1 \leq t \leq n} |y_t| / \sigma \sqrt{n} \right)^2}{\left(\frac{1}{n} \sum_{t=2}^n \left(y_t - \hat{\phi}_{OLS} y_{t-1} \right)^2 \right) / \sigma^2} \leq a \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{a}}^{\sqrt{a}} \sum_{k=-\infty}^{+\infty} (-1)^k \exp \left[-\frac{(u - 2k\sqrt{a})^2}{2} \right] du \tag{2.3.2}$$

Proof: From (2.2.3),

$$\tau = \frac{\left(\text{Max}_{1 \leq t \leq n} |y_t| / \sigma \sqrt{n} \right)^2}{\left(\frac{1}{n} \sum_{t=2}^n \left(y_t - \hat{\phi}_{OLS} y_{t-1} \right)^2 \right) / \sigma^2} \tag{2.3.3}$$

We know that $\left(\frac{1}{n} \sum_{t=2}^n \left(y_t - \hat{\phi}_{OLS} y_{t-1} \right)^2 \right) / \sigma^2$ converges in probability to 1. Hence the distribution of τ converges to the asymptotic distribution to $\left(\text{Max}_{1 \leq t \leq n} |y_t| / \sigma \sqrt{n} \right)^2$.

Now, we need the following theorem:

Donsker's Theorem (Billingsley, 1999)

Let $C = C[0,1]$ be the space of continuous functions on the unit interval. The distance between two points x and y (two continuous functions $x(\cdot)$ and $y(\cdot)$ on $[0,1]$) is defined as

$$\rho(x, y) = \|x - y\| = \sup_t |x(t) - y(t)| \tag{2.3.4}$$

Let ξ_1, ξ_2, \dots be a sequence of independently and identically distributed random variables having mean 0 and variance σ^2 . Let $S_n = \xi_1 + \dots + \xi_n$ ($S_0 = 0$), and let $X^n(\omega)$ be the element of C having the value

$$X_t^n(\omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}(\omega) + (nt - [nt]) \frac{1}{\sigma\sqrt{n}} \xi_{[nt]+1}(\omega) \quad (2.3.5)$$

at t . Thus $X^n(\omega)$ is the function defined by the linear interpolation between its values

$$X_{i/n}^n(\omega) = S_i(\omega) / \sigma\sqrt{n} \text{ at the points } i/n.$$

Donsker's Theorem states that under the above assumptions, $X^n \Rightarrow W$, a random function having Wiener measure as its distribution over C .

In our case, $e_t = \xi_t$ and $y_t = S_t$.

Let

$$M_n = \max_{0 \leq t \leq n} |y_t| \quad (2.3.6)$$

Since $h(x) = \sup_t |x(t)|$ is a continuous function on C , it follows from Donsker's

Theorem and the Continuous Mapping Theorem that

$$\sup_t |X_t^n| \Rightarrow \sup_t |W_t| \quad (2.3.7)$$

Now, $\sup_t |X_t^n| = M_n / \sigma\sqrt{n}$ and hence,

$$\frac{M_n}{\sigma\sqrt{n}} \Rightarrow \sup_t |W_t| \quad (2.3.8)$$

To get the distribution of $\sup_{0 \leq t \leq 1} |W_t|$, we use Theorem 2 of Gikhman & Skorokhod (1996,

Pg 286) which states,

$$\begin{aligned}
& P \left[\sup_{0 \leq t \leq T} |W_t| < a, W(T) \in [c, d] \right] \\
&= \frac{1}{\sqrt{2\pi T}} \int_c^d \sum_{k=-\infty}^{+\infty} (-1)^k \exp \left[-\frac{(u-2ka)^2}{2T} \right] du
\end{aligned} \tag{2.3.9}$$

where $[c, d] \subseteq [-a, a]$.

Putting $T = 1$ and $c = -a, d = a$ in equation (2.3.9), we get

$$\begin{aligned}
& P \left[\sup_{0 \leq t \leq 1} |W_t| < a \right] \\
&= \frac{1}{\sqrt{2\pi}} \int_{-a}^a \sum_{k=-\infty}^{+\infty} (-1)^k \exp \left[-\frac{(u-2ka)^2}{2} \right] du
\end{aligned} \tag{2.3.10}$$

$$\begin{aligned}
& \therefore \lim_{n \rightarrow \infty} P \left\{ \frac{\max_{1 \leq t \leq n} |y_t|}{\sigma \sqrt{n}} \leq a \right\} \\
&= \frac{1}{\sqrt{2\pi}} \int_{-a}^a \sum_{k=-\infty}^{+\infty} (-1)^k \exp \left[-\frac{(u-2ka)^2}{2} \right] du
\end{aligned} \tag{2.3.11}$$

Thus,

$$\begin{aligned}
& \lim_{n \rightarrow \infty} P \left\{ \left(\frac{\max_{1 \leq t \leq n} |y_t|}{\sigma \sqrt{n}} \right)^2 \leq a \right\} \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{a}}^{\sqrt{a}} \sum_{k=-\infty}^{+\infty} (-1)^k \exp \left[-\frac{(u-2k\sqrt{a})^2}{2} \right] du
\end{aligned} \tag{2.3.12}$$

$$\begin{aligned}
& \therefore \lim_{n \rightarrow \infty} P \left\{ \frac{\left(\max_{1 \leq t \leq n} |y_t| / \sigma \sqrt{n} \right)^2}{\left(\frac{1}{n} \sum_{t=1}^n \left(y_t - \hat{\phi}_{OLS} y_{t-1} \right)^2 \right) / \sigma^2} \leq a \right\} \\
& = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{a}}^{\sqrt{a}} \sum_{k=-\infty}^{+\infty} (-1)^k \exp \left[-\frac{(u - 2k\sqrt{a})^2}{2} \right] du
\end{aligned} \tag{2.3.13}$$

Thus we have not only shown that the distribution of the statistic τ converges to the asymptotic distribution of $\left(\max_{1 \leq t \leq n} |y_t| / \sigma \sqrt{n} \right)^2$ but we have also found an expression for limiting distribution function. We can further exploit the expression in (2.3.13) to approximate the limiting percentiles if we are allowed to change the sequence of application of the summation and integral operators. Such an interchange is justified by Dominated Convergence Theorem (Billingsley, 1991). To apply this theorem, we first

need to show that $\sum_{k=-\infty}^{\infty} a_k = \sum_{k=-\infty}^{+\infty} (-1)^k \exp \left[-\frac{(u - 2k\sqrt{a})^2}{2} \right]$ exists. If we can show

that $\sum_{k=-\infty}^{\infty} |a_k|$ is convergent, then we can conclude that $\sum_{k=-\infty}^{\infty} a_k$ is convergent.

Since $\sum_{k=-\infty}^{\infty} |a_n|$ is a series of positive terms, we can always write

$$\sum_{k=-\infty}^{\infty} |a_n| = \sum_{k=1}^{\infty} \exp \left[-\frac{(u - 2k\sqrt{a})^2}{2} \right] + \sum_{k=1}^{\infty} \exp \left[-\frac{(u + 2k\sqrt{a})^2}{2} \right] + \exp \left[-\frac{u^2}{2} \right]$$

To show that $\sum_{k=1}^{\infty} \exp \left[-\frac{(u - 2k\sqrt{a})^2}{2} \right]$ and $\sum_{k=1}^{\infty} \exp \left[-\frac{(u + 2k\sqrt{a})^2}{2} \right]$ are convergent, we will apply the Ratio Test (Theorem 8.25 in Apostol, 1997). Consider

$$\begin{aligned}
R &= \limsup_{n \rightarrow \infty} \frac{\exp\left[-\frac{(u-2(n+1)\sqrt{a})^2}{2}\right]}{\exp\left[-\frac{(u-2n\sqrt{a})^2}{2}\right]} \\
&= \limsup_{n \rightarrow \infty} \exp\left[-\frac{(u-2(n+1)\sqrt{a})^2}{2} + \frac{(u-2n\sqrt{a})^2}{2}\right] \\
&= \limsup_{n \rightarrow \infty} \exp\left[-\frac{1}{2}\left[(u-2(n+1)\sqrt{a})^2 - (u-2n\sqrt{a})^2\right]\right] \\
&= \limsup_{n \rightarrow \infty} \exp\left[-\frac{1}{2}(-2\sqrt{a})(2u-2\sqrt{a}-4n\sqrt{a})\right] \\
&= \limsup_{n \rightarrow \infty} \exp\left[\sqrt{a}(2u-2\sqrt{a}-4n\sqrt{a})\right] \\
&= 0 \\
&< 1
\end{aligned}$$

Hence by the Ration Test, $\sum_{k=1}^{\infty} \exp\left[-\frac{(u-2k\sqrt{a})^2}{2}\right]$ converges and similarly,

$\sum_{k=1}^{\infty} \exp\left[-\frac{(u+2k\sqrt{a})^2}{2}\right]$ also converges. Thus $\sum_{k=-\infty}^{\infty} |a_k|$ converges and hence $\sum_{k=-\infty}^{\infty} a_k$ converges.

Now, since $u \in (-\sqrt{a}, \sqrt{a})$,

$$\exp\left[-\frac{(u-2k\sqrt{a})^2}{2}\right] \leq \exp\left[-\frac{1}{2}u^2\right] \quad \text{and} \quad \exp\left[-\frac{(u+2k\sqrt{a})^2}{2}\right] \leq \exp\left[-\frac{1}{2}u^2\right]. \quad \text{Hence,}$$

$\left|\sum_{k=-n}^n a_k\right| \leq (2n+1)\exp\left[-\frac{1}{2}u^2\right]$ which is integrable. We can now apply Theorem 16.7 (Billingsley, 1991) to obtain,

$$\begin{aligned}
& \lim_{n \rightarrow \infty} P \left\{ \frac{\left(\max_{1 \leq t \leq n} |y_t| / \sigma \sqrt{n} \right)^2}{\left(\frac{1}{n} \sum_{t=1}^n \left(y_t - \hat{\phi}_{OLS} y_{t-1} \right)^2 \right) / \sigma^2} \leq a \right\} \\
&= \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \int_{-\sqrt{a}}^{\sqrt{a}} (-1)^k \exp \left[-\frac{(u - 2k\sqrt{a})^2}{2} \right] du \\
&= \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} (-1)^k \int_{-\sqrt{a}}^{\sqrt{a}} \exp \left[-\frac{(u - 2k\sqrt{a})^2}{2} \right] du \\
&= \sum_{k=-\infty}^{\infty} (-1)^k \int_{-\sqrt{a}-2k\sqrt{a}}^{\sqrt{a}-2k\sqrt{a}} \phi(v) dv ; \\
&\quad \phi(v) \text{ denotes the Standard Normal Density} \\
&= \sum_{k=-\infty}^{\infty} (-1)^k \left[\Phi(\sqrt{a} - 2k\sqrt{a}) - \Phi(-\sqrt{a} - 2k\sqrt{a}) \right]
\end{aligned} \tag{2.3.14}$$

We approximated the infinite sum in the above expression by

$$\sum_{k=-1000}^{1000} (-1)^k \left[\Phi(\sqrt{a} - 2k\sqrt{a}) - \Phi(-\sqrt{a} - 2k\sqrt{a}) \right] \tag{2.3.15}$$

The value of a , starting at 0 was incremented in steps of 0.0001 up to $a = 20$ and at each increment, the cumulative probability $P(\tau \leq a)$ was approximated by (2.3.15). The PROBNORM function in SAS was used to compute the Normal CDF Φ . This procedure yielded the following percentiles (rounded to two decimal places) that were accurate up to three decimal places and the corresponding probabilities that were accurate up to at least four decimal places.

Table 2-5: Asymptotic Percentiles of τ

	$p = P(\tau \leq a)$						
	0.01	0.05	0.10	0.50	0.90	0.95	0.99
a	0.25	0.38	0.48	1.32	3.84	5.02	7.87

Section 4. Comparative Power Study

In this section we evaluate the performance of the proposed test under various error specifications. First, we introduce the error distributions that differ from the Normal distribution in terms of the skewness or the peakedness of the density curve. We will then tabulate the rejection probabilities of the τ test and the DF tests under the unit root null hypothesis and a stationary alternative specified by $\phi = 0.95$.

We considered a variety of error distributions. The following graphs show the nature of the corresponding density functions.

(a) Skew-Normal Density and Skew-T Density

Ferenandez and Steel (1998) showed that one can introduce skewness into any symmetrical and unimodal distribution by changing a single scale parameter. If $f(\cdot)$ denotes the symmetric probability density function (p.d.f) of a random variable X , then the p.d.f $f(\cdot)$ can be skewed by scaling the density with factors $\frac{1}{\lambda}$ and λ in the positive and negative orthant. The procedure generates the following class of skewed distributions, indexed by λ :

$$p(x) = \frac{2\lambda}{1+\lambda^2} f(x/\lambda) I_{[0,\infty)} + \frac{2\lambda}{1+\lambda^2} f(x\lambda) I_{[-\infty,0)} \quad (2.4.1)$$

So if we let $f(x)$ to be the standard normal density function, then $p(x)$ will belong to the Skew-Normal family. Here are two plots of the Skew-Normal density curve corresponding to $\lambda = 0.5$ and $\lambda = 1.5$. While $0 < \lambda < 1$ skews the standard normal density to the right, $\lambda > 1$ skews the distribution to the left. We have the standard normal curve when $\lambda = 1$.

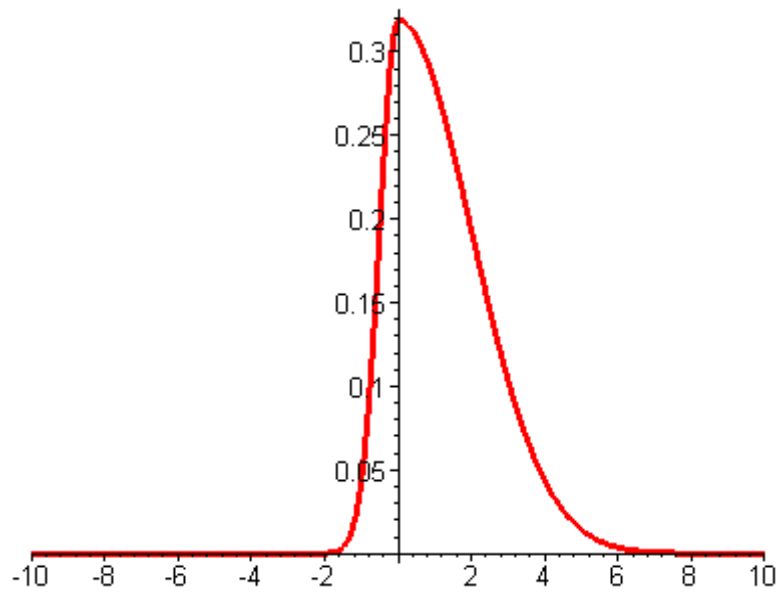


Figure 2-1:Skew-Normal Density Curve with $\lambda = 0.5$

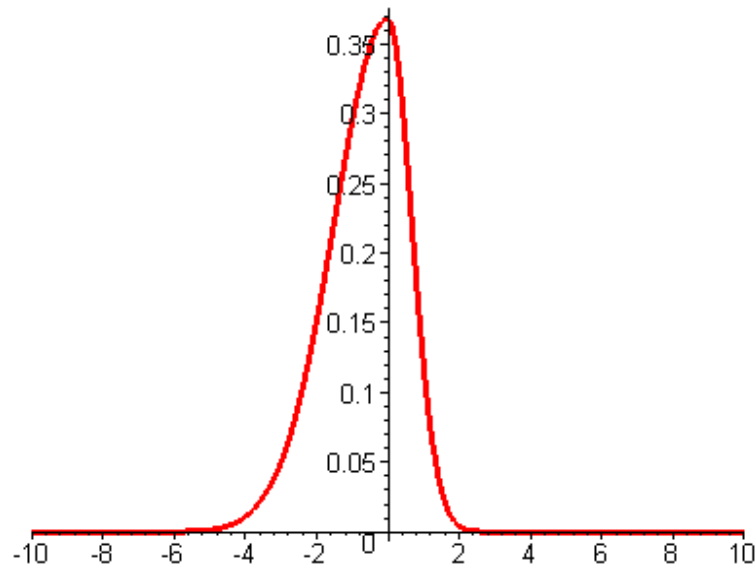


Figure 2-2: Skew-Normal Density with $\lambda = 1.5$

In a similar way, we can derive Skew-t distributions with mean zero. If we use the t density function as our $f(x)$, we will then obtain a family of thick-tailed skewed distributions. The following two plots display a t-density curve with 10 degrees of freedom (df) and a Skew-t density curve with 10 df.

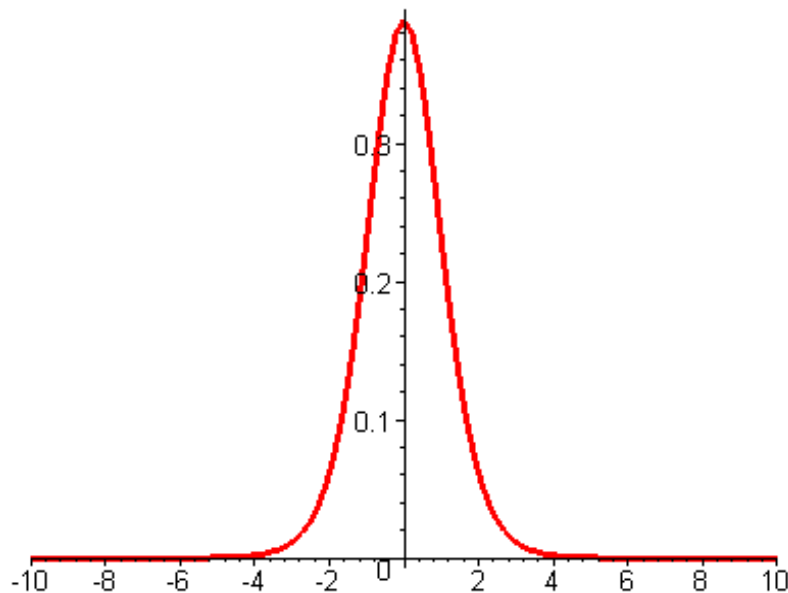


Figure 2-3: Skew t-distribution with 10 d.f. and $\lambda = 1$

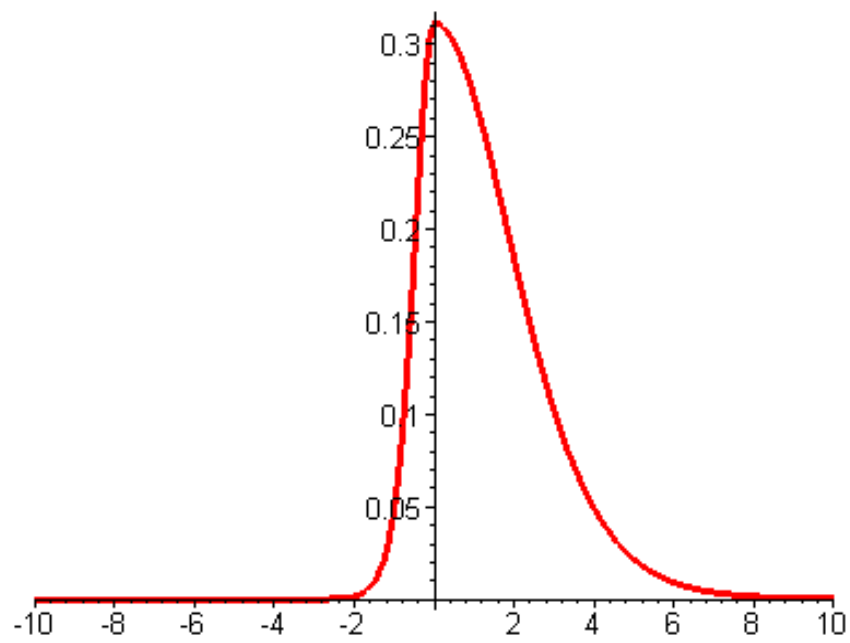


Figure 2-4: Skew-t distribution with 10 d.f. and $\lambda = 0.5$

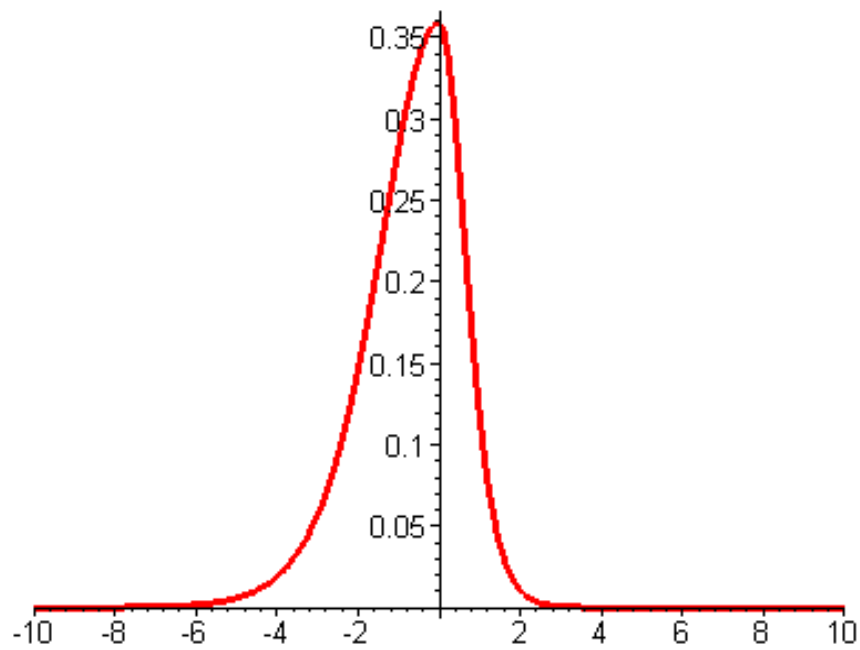
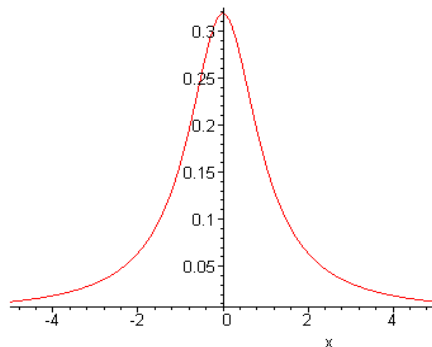


Figure 2-5: Skew-t distribution with 10 d.f. and $\lambda = 1.5$

(b) Cauchy Density:

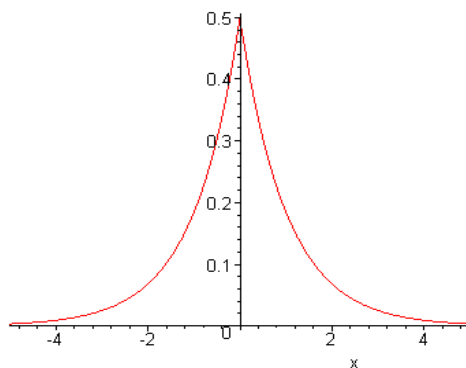
If $X \sim \text{Cauchy}$, then $f(x) = \frac{1}{\pi(1+x^2)}$; $-\infty \leq x \leq \infty$ is the standard Cauchy density function.

The Cauchy density curve has much fatter tails than the normal density curve as shown in the figure below.



(c) Double-Exponential Density:

The standard double-exponential density function is given by $f(x) = \frac{1}{2} \exp(-|x|)$; $-\infty \leq x \leq \infty$. The density curve has tails that are thicker than the tails of the standard normal curve but are thinner than those of the standard Cauchy density curve.



We now report the power and size properties of the proposed statistic τ and the DF statistic based on a Monte Carlo study. In testing the hypothesis stated by (2.2.2) under the model (2.3.1), we set the sample size $n = 100$ for the following study. The level of significance (LOS) was set to 0.05. The DF test rejected the unit root hypothesis if the computed statistic value was less than -1.95 . As tabulated in Fuller (1996), -1.95 happens to be the 5th percentile for the finite sample distribution as well as for the asymptotic distribution of the DF test statistic. Our proposed test rejected the null hypothesis at 5% LOS if the computed statistic value was less than 0.33 (when using the finite sample distribution of τ). When the error distribution is non-normal, we might want to use the asymptotic percentiles. So, we also report the size and power if we used 0.38 (the 5th percentile from the asymptotic distribution of τ) as the cut-off value for the rejection region. The rejection probabilities are based on 1000 replications.

The power of the test based on the finite sample percentile of τ is less than the power of the DF test when the innovations follow either the Normal distribution or t-distribution or Cauchy distribution or the double-exponential distribution. Relative to the DF test, the worst performance of τ is in the case of the Cauchy distribution. When the errors are sampled from either the Skew-Normal or the Skew-T distribution, the power of the τ test is almost as good as that of the DF test.

Note that as the skewness parameter λ goes farther away from 1 on either side, the power of both the τ test and the DF test falls under both, the null hypothesis and the alternative hypothesis. The τ test seems to perform almost as well as the DF test when either the normal density or the t-density gets skewed to the left or right.

Table 2-6: Size and Power based on Finite Sample Percentiles under various specifications for the Error distribution

Error Distribution	Rejection Probabilities based on finite sample percentiles			
	$\phi = 1$		$\phi = 0.95$	
	$P(\tau < 0.33)$	$P(DF < -1.95)$	$P(\tau < 0.33)$	$P(DF < -1.95)$
Skew-Normal- $\lambda = 0.90$	0.015	0.019	0.093	0.105
Skew-Normal- $\lambda = 0.92$	0.020	0.019	0.155	0.156
Skew-Normal- $\lambda = 0.94$	0.035	0.032	0.213	0.216
Skew-Normal- $\lambda = 0.96$	0.040	0.040	0.254	0.270
Skew-Normal- $\lambda = 0.98$	0.039	0.044	0.266	0.306
Skew-Normal- $\lambda = 1.00$	0.048	0.051	0.280	0.326
Skew-Normal- $\lambda = 1.02$	0.057	0.059	0.272	0.316
Skew-Normal- $\lambda = 1.04$	0.042	0.045	0.261	0.281
Skew-Normal- $\lambda = 1.06$	0.034	0.030	0.213	0.233
Skew-Normal- $\lambda = 1.08$	0.020	0.018	0.172	0.176
Skew-Normal- $\lambda = 1.10$	0.016	0.016	0.114	0.134
Cauchy	0.046	0.032	0.078	0.187
Double-Exponential	0.046	0.049	0.283	0.323
Skew-T: $\lambda = 1.00$, $10d.f.$	0.046	0.051	0.275	0.319
Skew-T: $\lambda = 0.90$, $10d.f.$	0.015	0.019	0.101	0.112

We also compared the τ test and the DF test based on the asymptotic percentiles. The DF asymptotic 5th percentile being the same as the DF 5th percentile based on a sample size of 100, there is no change in the size and power of the DF test based on the asymptotic percentiles.

However, the 5th percentile of the asymptotic distribution of τ was approximated to be 0.38 which is slightly higher than 0.33—the 5th percentile based on a sample size of 100. As expected, we now see a clear dominance of the τ test over the DF test (except in the case of the Cauchy distribution) in terms of the probability of rejecting the unit root hypothesis in favor of the stationary hypothesis identified by $\phi = 0.95$. But at the same time, the τ test over-estimates the nominal size of the test in many cases. Skew-Normal ($\lambda = 0.94, 0.92, 0.90, 1.06, 1.08, 1.10$) and Skew-T ($\lambda = 0.90$, $10d.f.$) distributions are the only cases where the τ test performs better than the DF test in terms of maintaining the size at or below the nominal level and in terms of having better power.

Table 2-7: Size and Power based on Asymptotic Percentiles under various specifications for the Error distribution

Error Distribution	Rejection Probabilities based on Asymptotic percentiles			
	$\phi = 1$		$\phi = 0.95$	
	$P(\tau < 0.38)$	$P(DF < -1.95)$	$P(\tau < 0.38)$	$P(DF < -1.95)$
Skew-Normal- $\lambda = 0.90$	0.020	0.019	0.154	0.105
Skew-Normal- $\lambda = 0.92$	0.032	0.019	0.222	0.156
Skew-Normal- $\lambda = 0.94$	0.050	0.032	0.285	0.216
Skew-Normal- $\lambda = 0.96$	0.065	0.040	0.335	0.270
Skew-Normal- $\lambda = 0.98$	0.064	0.044	0.374	0.306
Skew-Normal- $\lambda = 1.00$	0.069	0.051	0.387	0.326
Skew-Normal- $\lambda = 1.02$	0.080	0.059	0.393	0.316
Skew-Normal- $\lambda = 1.04$	0.075	0.045	0.363	0.281
Skew-Normal- $\lambda = 1.06$	0.051	0.030	0.314	0.233
Skew-Normal- $\lambda = 1.08$	0.035	0.018	0.250	0.176
Skew-Normal- $\lambda = 1.10$	0.021	0.016	0.202	0.134
Cauchy	0.076	0.032	0.126	0.187
Double-Exponential	0.070	0.049	0.390	0.323
Skew-T: $\lambda = 1.00, 10d.f.$	0.077	0.051	0.389	0.319
Skew-T: $\lambda = 0.90, 10d.f.$	0.025	0.019	0.160	0.112

Centered Skew-Normal, Centered Skew-t, Centered Uniform and Centered Exponential distributions.

Note that the mean of the Skew-Normal variable or the Skew-t variable is not zero. The Theorem (2.3.2) is based on the assumption that the errors are i.i.d. with zero mean and variance σ^2 . So, if $X \sim \text{Skew-Normal}$, then we need to generate $Y = X - E(X)$ as the i.i.d. errors so that the use of DF percentiles or the τ percentiles can be justified. We also have to make a similar adjustment to the Skew-t random variable. The expectation of a Skew-Normal random variable and a Skew-t random variable are given by

$$\begin{aligned}
 E(X) &= \frac{\lambda^2 - 1}{\lambda} \sqrt{\frac{2}{\pi}} & ; X \sim \text{Skew-Normal} \\
 E(X) &= \frac{\lambda^2 - 1}{\lambda} \sqrt{\frac{v}{\pi}} \left| \frac{v-1}{2} \right| / \left| \frac{v}{2} \right| & ; X \sim \text{Skew-t}
 \end{aligned}
 \tag{2.4.2}$$

From the simulation results in Table 2-8, we see that the power of the DF test dominates the power of the τ -test (based on finite sample percentiles). It seems that the power of the DF test is around 5% points higher than that of the τ -test when the innovations come from a centered right-skewed Normal Distribution. When the errors arise from a centered and heavily left-skewed Normal distribution, the powers of the two tests are almost the same.

Errors were also generated from Uniform (-0.5, 0.5) and $\{Exponential(1)-1\}$. The power of the DF test again dominated the power of the test based on τ .

Table 2-8: Size and Power based on Finite Sample Percentiles under centered Skew-Normal, Skew-t, Uniform and Exponential distributions.

Error Distribution	Rejection Probabilities based on finite sample percentiles			
	$\phi = 1$		$\phi = 0.95$	
	$P(\tau < 0.33)$	$P(DF < -1.95)$	$P(\tau < 0.33)$	$P(DF < -1.95)$
Skew-Normal- $\lambda = 0.90$	0.043	0.049	0.287	0.328
Skew-Normal- $\lambda = 0.70$	0.037	0.053	0.287	0.314
Skew-Normal- $\lambda = 0.50$	0.046	0.053	0.289	0.304
Skew-Normal- $\lambda = 0.30$	0.045	0.047	0.288	0.304
Skew-Normal- $\lambda = 0.10$	0.046	0.049	0.293	0.308
Skew-Normal- $\lambda = 1.00$	0.048	0.051	0.280	0.326
Skew-Normal- $\lambda = 1.10$	0.045	0.049	0.278	0.328
Skew-Normal- $\lambda = 1.30$	0.053	0.054	0.276	0.330
Skew-Normal- $\lambda = 1.50$	0.060	0.054	0.278	0.333
Skew-Normal- $\lambda = 1.70$	0.063	0.056	0.282	0.334
Skew-Normal- $\lambda = 1.90$	0.063	0.055	0.285	0.338
Skew-T: $\lambda = 1.00$, $10d.f.$	0.046	0.051	0.275	0.319
Skew-T: $\lambda = 0.10$, $10d.f.$	0.048	0.049	0.282	0.308
Centered Exponential	0.039	0.034	0.274	0.318
Centered Uniform	0.041	0.046	0.283	0.319

If we use the asymptotic percentiles to conduct the τ -test, then its power is about 5-7% points higher than that of the DF test. However, the τ -test overestimates the nominal size (5%) by almost 2-4% points whereas the DF estimates the size very accurately. The estimated size of the τ -test, when the innovations follow right skew-Normal distributions

are within the 3σ confidence interval (0.029, 0.0706) for the nominal size in most cases. In such cases, the power of the DF test is found to be significantly higher than the τ test based on the 3σ confidence interval for the difference in proportions ($P(\tau < 0.38) - P(DF < -1.95)$). The size distortion is more pronounced when the errors come from a left skew-Normal distribution.

Table 2-9: Size and Power based on Asymptotic Percentiles under centered Skew-Normal, Skew-t, Uniform and Exponential distributions.

Error Distribution	Rejection Probabilities based on Asymptotic percentiles			
	$\phi = 1$		$\phi = 0.95$	
	$P(\tau < 0.38)$	$P(DF < -1.95)$	$P(\tau < 0.38)$	$P(DF < -1.95)$
Skew-Normal- $\lambda = 0.90$	0.070	0.049	0.387	0.328
Skew-Normal- $\lambda = 0.70$	0.065	0.053	0.384	0.314
Skew-Normal- $\lambda = 0.50$	0.071	0.053	0.385	0.304
Skew-Normal- $\lambda = 0.30$	0.068	0.047	0.389	0.304
Skew-Normal- $\lambda = 0.10$	0.069	0.049	0.384	0.308
Skew-Normal- $\lambda = 1.00$	0.069	0.051	0.387	0.326
Skew-Normal- $\lambda = 1.10$	0.070	0.049	0.371	0.328
Skew-Normal- $\lambda = 1.30$	0.083	0.054	0.390	0.330
Skew-Normal- $\lambda = 1.50$	0.080	0.054	0.394	0.333
Skew-Normal- $\lambda = 1.70$	0.086	0.056	0.392	0.334
Skew-Normal- $\lambda = 1.90$	0.087	0.055	0.394	0.338
Skew-T: $\lambda = 1.00, 10d.f.$	0.077	0.051	0.389	0.319
Skew-T: $\lambda = 0.10, 10d.f.$	0.071	0.049	0.390	0.308
Centered Exponential	0.070	0.034	0.387	0.318
Centered Uniform	0.071	0.046	0.399	0.319

Table 2-10 through Table 2-13 compares the DF test and the τ test for small sample sizes. When the error distribution is Centered Exponential, the τ test based on the finite sample percentiles has slightly better power (though statistically insignificant) than the DF test for $n = 25$ as well as $n = 50$. Simulations indicate that use of asymptotic percentiles would be inappropriate when the error distribution is Centered Exponential. When the error distribution is Centered Uniform, the τ test based on asymptotic percentiles, delivers higher rejection rates under the alternatives than the DF test.

However, the estimated size of the τ test is almost on the border of the 3σ limits of the nominal size.

Table 2-10: Rejection Probabilities when $e_t \sim \{Exponential(1)-1\}$ or $Uniform(-0.5,0.5)$, $n = 25$ based on asymptotic percentiles

		ϕ							
		0.95		0.98		0.99		1	
		τ	DF	τ	DF	τ	DF	τ	DF
Error Distribution									
Centered Exponential	0.139	0.091		0.114	0.077	0.108	0.07	0.095	0.061
Centered Uniform	0.126	0.089		0.078	0.061	0.074	0.051	0.069	0.046

Table 2-11: Rejection Probabilities when $e_t \sim \{Exponential(1)-1\}$ or $Uniform(-0.5,0.5)$, based on finite sample ($n = 25$) percentiles.

		ϕ							
		0.95		0.98		0.99		1	
		τ	DF	τ	DF	τ	DF	τ	DF
Error Distribution									
Centered Exponential	0.092	0.091		0.078	0.077	0.072	0.07	0.065	0.061
Centered Uniform	0.082	0.089		0.057	0.061	0.051	0.051	0.048	0.046

Table 2-12: Rejection Probabilities when $e_t \sim \{Exponential(1)-1\}$ or $Uniform(-0.5,0.5)$, $n = 50$ based on asymptotic percentiles

		ϕ							
		0.95		0.98		0.99		1	
		τ	DF	τ	DF	τ	DF	τ	DF
Error Distribution									
Centered Exponential	0.187	0.121		0.121	0.076	0.094	0.064	0.081	0.051
Centered Uniform	0.184	0.137		0.104	0.071	0.095	0.056	0.067	0.041

Table 2-13: Rejection Probabilities when $e_t \sim \{Exponential(1)-1\}$ or $Uniform(-0.5,0.5)$, based on finite sample ($n = 50$) percentiles.

	ϕ							
	0.95		0.98		0.99		1	
	τ	DF	τ	DF	τ	DF	τ	DF
Error Distribution								
Centered Exponential	0.131	0.121	0.08	0.076	0.071	0.064	0.056	0.051
Centered Uniform	0.129	0.137	0.072	0.071	0.057	0.056	0.04	0.041

We also generated the errors from another symmetric random variable X whose density function is given by

$$f(x) = \begin{cases} -x & -1 \leq x \leq 0 \\ x & 0 \leq x \leq 1 \end{cases} \quad (2.5.1)$$

The corresponding density curve is plotted in the figure below:

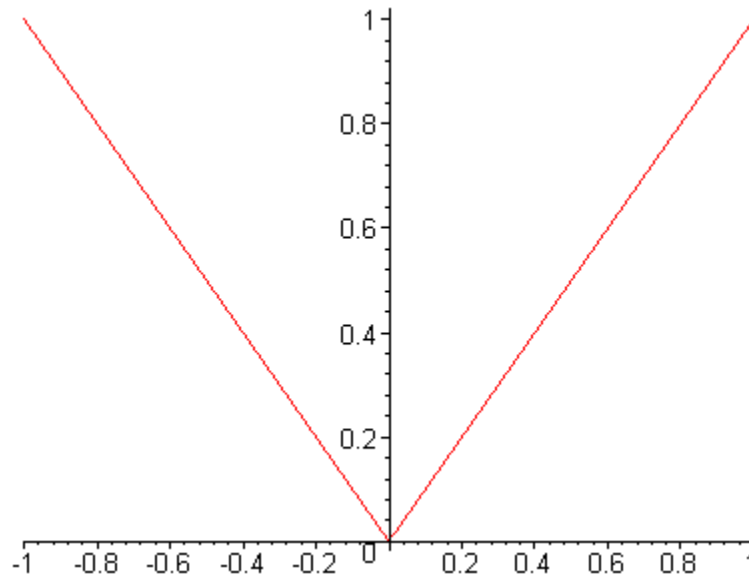


Figure 2-6: Density Curve corresponding to (2.5.1)

The size and power of the τ test and the DF test based on finite sample percentiles and asymptotic percentiles are compared in Table 2-14 and Table 2-15. Again we observe that the DF provides better power than the τ test based on finite sample percentiles. The

τ test based on asymptotic percentiles, overestimates the nominal size as was noted in the case of most of the other error distributions.

Table 2-14: Rejection probabilities when the error distribution is given by the cone-shaped distribution (2.5.1) and the sample size $n = 50$

ϕ						
	0.95		0.98		1	
	τ	DF	τ	DF	τ	DF
Based on finite sample percentiles	0.120	0.134	0.065	0.075	0.044	0.042
Based on Asymptotic percentiles	0.190	0.134	0.101	0.075	0.065	0.042

Table 2-15: Rejection probabilities when the error distribution is given by the cone-shaped distribution (2.5.1) and the sample size $n = 25$

ϕ						
	0.95		0.98		1	
	τ	DF	τ	DF	τ	DF
Based on finite sample percentiles	0.087	0.093	0.057	0.067	0.047	0.054
Based on Asymptotic percentiles	0.124	0.093	0.099	0.067	0.077	0.054

Section 5. Conclusions

In this chapter we have proposed a test statistic for the unit-root null hypothesis in a simple AR (1) model. The test statistic was proven to have a non-degenerate asymptotic distribution. The asymptotic percentiles were easily obtained by using SAS (v8) to truncate the infinite sum expression to a finite sum with sufficiently large number of terms. It would be easy to obtain the p-value for the τ test since the asymptotic distribution function has a simple expression. Simulation studies indicate that the proposed tends to perform better when the errors originate from skewed distributions (skew Normal or Exponential) than when the errors originate from symmetric distributions.

Chapter 3 Testing the Null hypothesis of Stationarity using Intersection-Union Principle

Section 1. Introduction

Knowing whether the time series data is stationary around a deterministic time trend or it has a unit root is an interesting question for macroeconomists. Answering this question would help us gauge the impact of transitory shocks on the levels of the series. While the presence of a unit root results in a permanent effect of a shock on the level of the series, the effect of a shock dies down in the case of a trend stationary series. Specifically, one would like to know, whether economic recessions have permanent consequences for the level of future GNP, or instead represent temporary decline with the lost output eventually made up when the economy picks up. A similar question about the level of economic productivity could also be framed -- Do economic booms raise the level of productivity permanently or instead productivity rises temporarily only to fall back to lower levels during the economic contraction.

A trend stationary (TS) series can be modeled as the sum of a linear polynomial in time and a stationary and invertible autoregressive moving average (ARMA) process. For instance, $L_t = C + \mu t + u_t$, where u_t is an ARMA(p,q) process. A DS process, on the other hand, describes time series whose first difference $L_t - L_{t-1}$ is assumed to be a stationary and invertible ARMA process with mean μ . The debate on trend stationarity (TS) vs. difference stationarity (DS) gained momentum after Nelson and Plosser (1982) investigated the behavior of annual U.S. time series data on output, spending, money, prices and interest rates. The influential article of Nelson and Plosser (1982) argued that most economic time series are characterized better by DS processes than by TS processes. On the basis of Dickey-Fuller type tests, they failed to reject the null hypothesis of difference stationarity (or a unit root): H_{DS} in favor of the alternative hypothesis of trend stationary: H_{TS} in all but one of the 14 annual U.S. time series. These

results are maintained even after allowing for autocorrelated errors using the augmented tests of Said and Dickey (1984) or the test statistics of Phillips (1987) and Phillips and Perron (1988).

However, in the above empirical studies, the burden of proof laid on the alternative hypothesis of stationarity. Hence it could be argued that the tests failed to reject the unit root hypothesis because the standard unit root tests are not powerful enough. Such arguments are supported by DeJong et al.(1989) and Diebold and Rudebusch (1990). It would thus be useful to test the null hypothesis H_{TS} against the alternative H_{DS} .

While the ARMA representation for the levels L_t of a difference stationary process has a unit root in its autoregressive (AR) polynomial, the ARMA representation for the first differences $L_t - L_{t-1}$ of a trend stationary process has a unit root in its moving average polynomial. Plosser and Schwert(1977), Campbell and Mankiw(1987), Fuller(1988) and Chang(1989) have considered testing the null hypothesis H_{TS} against the alternative H_{DS} , assuming that the first difference process is an ARMA of known orders. Cochrane (1988) presented a nonparametric criterion for testing H_{TS} versus H_{DS} which does not require the specification of the orders. Tanaka (1990) and Saikkonen and Luukkonen (1993) provided test procedures motivated by locally best invariant and unbiased (LBIU) arguments. Tsay (1993) used nonstationary but invertible processes derived from a stationary but noninvertible ARMA model to test for a moving average (MA) unit root in the original process. Arellano and Pantula(1995) also used the idea of testing for noninvertibility of a ARMA model and proposed test statistics based on one-step Gauss-Newton estimators of the moving average (MA) parameters.

Kwiatkowski et al. (1992) considered a different approach to testing the null hypothesis of stationarity. They represented the time series under study as a sum of a deterministic trend, a random walk and a stationary error. Thus, the null hypothesis of trend stationarity corresponds to the hypothesis that the variance of the random walk equals zero. They derived the limiting distributions of the one-sided LM statistic for the trend stationarity

hypothesis under general conditions on the stationary error. The results of application of the KPSS test to Nelson-Plosser data depends on the way the deterministic trend is accommodated. For all most all series, we can reject the hypothesis of level stationarity, but for many of the series we cannot reject the hypothesis of trend stationarity.

In this paper, we consider the problem of testing a slightly different null hypothesis against the same unit root alternative hypothesis considered in Kwiatkowski et al.

We set up our hypotheses as follows:

$$H_0 : \beta < 1 \text{ or } \beta > 1 \text{ against } H_1 : \beta = 1 \quad (3.1.1)$$

in the model $Y_t = \beta Y_{t-1} + \varepsilon_t$, $Y_0 = 0$, where ε_t is a sequence of i.i.d $N(0, \sigma^2)$. Note that our H_0 includes stationary as well as explosive series. We propose a test based on the principle of Intersection-Union tests (IUT). Berger (1982) proposed the use of IUT in a quality control context closely related to bioequivalence testing. The IUT method allows the combination of size- α tests to form an overall size- α test without making any adjustment to the sizes of the individual tests. Westlake (1981) and Schuirmann(1981) proposed the “two one sided test”(TOST) which turns out to be a simple example of an IUT. As a first attempt to apply IUT to the unit root-testing problem, we investigate the performance of a TOST in testing the above hypothesis.

The rest of this paper is organized into six sections. Section 2 presents the relevant theory underlying IUT and the appropriate formulation of our original hypotheses so that IUT is applicable. Section 3 discusses the asymptotic properties of the proposed test in Section 2. Section 4 provides a Monte Carlo study of the power of the test. Section 5 considers a variant of the formulation presented in Section 2. Finally, in Section 6, we lay out the future course of work on the current study.

Section 2. Intersection Union Test for near Unit Root hypothesis

Before we present the test statistics, we introduce some results on IUT that will be used to construct the test. The IUT method is useful for the following type of hypothesis testing problem. Let θ denote the unknown parameter (θ can be vector valued) in the distribution of the data X . Let Θ denote the parameter space. Let $\Theta_1, \dots, \Theta_k$ denote subsets of Θ . Suppose we wish to test

$$H_0 : \theta \in \bigcup_{i=1}^k \Theta_i \text{ versus } H_a : \theta \in \bigcap_{i=1}^k \Theta_i^c \quad (3.2.1)$$

where Θ_i^c denotes the complement of Θ_i . For $i=1, \dots, k$, let R_i denote a rejection region for a test of $H_{0i} : \theta \in \Theta_i$ versus $H_{ai} : \theta \in \Theta_i^c$. Then an IUT of (3.2.1) is the test that rejects H_0 if and only if $X \in \bigcap_{i=1}^k R_i$. Simply stated, the overall null hypothesis can be rejected only if each of the individual hypotheses can be rejected.

Berger (1982) proved that if R_i is a level- α test of H_{0i} , for $i=1, \dots, k$, then the intersection union test with rejection region $R = \bigcap_{i=1}^k R_i$ is a level- α test of H_0 versus H_a in (3.2.1).

Our approach to study the hypothesis-testing problem (3.1.1) is to approximate (3.1.1) by the following sequence of sub-problems.

For $t = 1, 2, \dots, n$, suppose $Y_t(n)$ satisfies the AR(1) model,

$$Y_t(n) = b_n Y_{t-1}(n) + e_t \quad (3.2.2)$$

$Y_0(n) = 0$, for all n . And e_t is a sequence of i.i.d. normal random variables $(0, \sigma^2)$.

Let

$$H_{0n} : \beta_n \leq 1 - \frac{\gamma}{n} \text{ or } \beta_n \geq 1 + \frac{\gamma}{n} \text{ versus } H_{1n} : 1 - \frac{\gamma}{n} < \beta_n < 1 + \frac{\gamma}{n}, \quad (3.2.3)$$

where γ is a fixed positive constant with respect to n . Then as $n \rightarrow \infty$, H_{0n} and H_{1n} tend to H_0 and H_1 respectively.

The above forms of H_{0n} and H_{1n} suggest that we can construct an IUT for the problem at hand. To see this, let $\Theta_1 = \{\beta_n : \beta_n \leq 1 - \frac{\gamma}{n}\}$ and $\Theta_2 = \{\beta_n : \beta_n \geq 1 + \frac{\gamma}{n}\}$.

Then our H_{0n} is $\Theta_1 \cup \Theta_2$ and H_{1n} is $\Theta_1^c \cap \Theta_2^c$. According to the IUT theory, if R_i (rejection region) is a level α test of $\Theta_i, i=1,2$. Then, the IUT with rejection region $R = R_1 \cap R_2$ is a level α test of H_{0n} versus H_{1n} .

In this paper, R_1 and R_2 will be constructed using (a) the pivotal statistic associated with the LSE of β_n and (b) the Weighted Symmetric Estimator (WS).

(a) Least squares Estimator:

The model we consider is given by (3.2.2). Let $\hat{\beta}_{OLS,n} = \frac{\sum_{t=2}^n Y_{t-1} Y_t}{\sum_{t=2}^n Y_{t-1}^2}$ denote the ordinary least squares estimator of β .

$$\text{Let } \tau_{OLS,n}^{(1)} = \left(\frac{\hat{\beta}_{OLS,n} - \left(1 - \frac{\gamma}{n}\right)}{s} \right) \left(\sum_{t=2}^n Y_{t-1}^2 \right)^{\frac{1}{2}} \text{ and } \tau_{OLS,n}^{(2)} = \left(\frac{\hat{\beta}_{OLS,n} - \left(1 + \frac{\gamma}{n}\right)}{s} \right) \left(\sum_{t=2}^n Y_{t-1}^2 \right)^{\frac{1}{2}}$$

where, $s^2 = \frac{I}{n-2} \left(\sum_{t=2}^n (Y_t - \hat{\beta}_{OLS,n} Y_{t-1})^2 \right)$ is the estimator of σ^2 . First of all, we need to determine R_1 , the rejection region for testing $H: \beta_n \leq 1 - \frac{\gamma}{n}$ against $K: \beta_n > 1 - \frac{\gamma}{n}$ and R_2 , the rejection region for testing $H: \beta_n \geq 1 + \frac{\gamma}{n}$ against $K: \beta_n < 1 + \frac{\gamma}{n}$. The two individual rejection regions can be constructed as $R_1 = \{ \tau_{OLS,n}^{(1)} : \tau_{OLS,n}^{(1)} \geq c_{OLS,n}^{(1)} \}$ where $c_{OLS,n}^{(1)}$ is such that $\sup_{\beta_n \in \Theta_1} P(R_1 | \beta_n) \leq \alpha$ and $R_2 = \{ \tau_{OLS,n}^{(2)} : \tau_{OLS,n}^{(2)} \leq c_{OLS,n}^{(2)} \}$ where $c_{OLS,n}^{(2)}$ is such that $\sup_{\beta_n \in \Theta_2} P(R_2 | \beta_n) \leq \alpha$. At a given level- α , the critical value $c_{OLS,n}^{(1)}$ can be computed from the distribution of $\tau_{OLS,n}^{(1)}$ generated under $\beta_n = 1 - \frac{\gamma}{n}$. Similarly $c_{OLS,n}^{(2)}$ can be computed from the distribution of $\tau_{OLS,n}^{(2)}$ generated under $\beta_n = 1 + \frac{\gamma}{n}$. Then the overall rejection region for the IUT will be $R = \{ (\tau_{OLS,n}^{(1)}, \tau_{OLS,n}^{(2)}) : \tau_{OLS,n}^{(1)} \geq c_{OLS,n}^{(1)} \text{ and } \tau_{OLS,n}^{(2)} \leq c_{OLS,n}^{(2)} \}$ and will have level- α .

(b) Weighted Symmetric Estimator (WS):

The WS estimator is a member of the class of symmetric estimators—the symmetry referring to the fact that if a normal stationary AR process satisfies $y_{t+1} = \beta y_t + \eta_t$, $\eta_t \sim NID(0, \sigma_\eta^2)$, then it also satisfies the equation $y_t = \beta y_{t+1} + \eta_t^*$, $\eta_t^* \sim NID(0, \sigma_\eta^2)$. The WS estimator of β , given by

$$\hat{\beta}_{WS,n} = \frac{\sum_{t=2}^n y_{t-1} y_t}{\sum_{t=2}^{n-1} y_t^2 + n^{-1} \sum_{t=1}^n y_t^2}, \text{ was studied by Park and Fuller (1993). It is obtained by}$$

minimizing $Q_w(\beta) = \sum_{t=2}^n w_t (y_t - \beta y_{t-1})^2 + \sum (1 - w_{t+1})(y_t - \beta y_{t+1})^2$, where $w_t = n^{-1}(t-1)$,

$t = 2, 3, \dots, n$. The pivotal statistic corresponding to $\hat{\beta}_{WS,n}$ is given by

$$\tau_{WS} = \left[\hat{\beta}_{WS,n} - \beta \right] \left[\sum_{t=2}^{n-1} y_t^2 + n^{-1} \sum_{t=1}^n y_t^2 \right]^{1/2} \hat{\sigma}_{WS}^{-1}, \text{ where } \hat{\sigma}_{WS}^2 = (n-2)^{-1} Q_w(\hat{\beta}_{WS,n}).$$

Here again, we construct the overall rejection region by taking the intersection of the two individual rejection regions. We define

$$\tau_{WS,n}^{(1)} = \left[\hat{\beta}_{WS,n} - \left(1 - \frac{\gamma}{n} \right) \right] \left[\sum_{t=2}^{n-1} y_t^2 + n^{-1} \sum_{t=1}^n y_t^2 \right]^{1/2} \hat{\sigma}_{WS}^{-1}$$

and

$$\tau_{WS,n}^{(2)} = \left[\hat{\beta}_{WS,n} - \left(1 + \frac{\gamma}{n} \right) \right] \left[\sum_{t=2}^{n-1} y_t^2 + n^{-1} \sum_{t=1}^n y_t^2 \right]^{1/2} \hat{\sigma}_{WS}^{-1}$$

It is given by $R = \left\{ (\tau_{WS,n}^{(1)}, \tau_{WS,n}^{(2)}) : \tau_{WS,n}^{(1)} \geq c_{WS,n}^{(1)} \text{ and } \tau_{WS,n}^{(2)} \leq c_{WS,n}^{(2)} \right\}$ where $c_{WS,n}^{(1)}$ and $c_{WS,n}^{(2)}$ are such that $\sup_{\beta_n \in \Theta_1} P \left\{ \tau_{WS,n}^{(1)} : \tau_{WS,n}^{(1)} \geq c_{WS,n}^{(1)} \mid \beta_n \right\} \leq \alpha$ and $\sup_{\beta_n \in \Theta_2} P \left\{ \tau_{WS,n}^{(2)} : \tau_{WS,n}^{(2)} \leq c_{WS,n}^{(2)} \mid \beta_n \right\} \leq \alpha$.

Section 3. Asymptotic Theory

The $Y_t(n)$ process in (3.2.2) when $\beta_n = \left(1 - \frac{\gamma}{n} \right)$ is the so called Nearly Non-stationary (NNS) autoregressive process. It has been investigated among others by Ahtola and Tiao(1984), Chan and Wei(1987). Under the assumption that $\{\varepsilon_t\}$ is a martingale difference sequence with respect to an increasing sequence of sigma fields $\{\mathfrak{F}_t\}$ such that $(as\ n \rightarrow \infty)$,

$$\begin{aligned}
n^{-1} \sum_{t=1}^n E(\varepsilon_t^2 | \mathfrak{T}_{t-1}) &\xrightarrow{p} 1 \\
\forall \alpha > 0, \quad n^{-1} \sum_{t=1}^n E(\varepsilon_t^2 I(|\varepsilon_t| > \sqrt{n\alpha}) | \mathfrak{T}_{t-1}) &\xrightarrow{p} 0
\end{aligned} \tag{3.3.1}$$

Chan and Wei (1987) derived the limiting distribution of the statistic

$$\tau_n = \sigma \left(\hat{\beta}_n - \left(1 - \frac{\gamma}{n}\right) \right) \left(\sum_{t=2}^n Y_{t-1}^2 \right)^{\frac{1}{2}} \text{ where } \hat{\beta}_n = \frac{\sum_{t=1}^n Y_t Y_{t-1}}{\sum_{t=1}^n Y_{t-1}^2} \text{ is the least squares estimator (LSE)}$$

of β_n . If $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with zero means and finite variances σ^2 , then $\{\varepsilon_t\}$ satisfy the above conditions. Note that the distribution of τ_n

does not depend on σ and for convenience, the authors assumed $\sigma^2 = 1$. In practice, we

may not know the value of σ . We can then use $s = \left\{ \frac{I}{n-2} \left(\sum_{t=1}^n (Y_t - \hat{\beta}_n Y_{t-1})^2 \right) \right\}^{\frac{1}{2}}$.

Simulation studies of Ahtola and Tiao(1984) suggest that the distribution of τ_n is not significantly altered when σ is replaced by s . Using Theorem 1 in Chan and Wei (1987), if $\gamma = \gamma_0$, then $\tau_n^{(1)} \rightarrow_d \mathcal{L}(\gamma_0)$ where

$$\mathcal{L}(\gamma_0) = \frac{\int_0^1 (1+bt)^{-1} W(t) dW(t)}{\left\{ \int_0^1 (1+bt)^{-2} W^2(t) dt \right\}^{\frac{1}{2}}}, b = e^{2\gamma_0} - 1 \tag{3.3.2}$$

and $\{W(t): 0 \leq t \leq 1\}$ is a standard Brownian motion. Using the same result,

$$\tau_n^{(2)} \rightarrow_d \mathcal{L}(-\gamma_0).$$

Chan(1988) provided a simple and efficient means to tabulate the percentiles of the distribution of $\mathcal{L}(\gamma)$. The percentiles for $\mathcal{L}(\gamma)$ at a few specific values of γ were presented and three algorithms for the construction of the percentiles at any other value of γ were presented in Chan (1988). Using these percentiles, we can construct the asymptotic rejection region to perform the IUT test.

Section 4. Monte Carlo Study

(a) Empirical Percentiles

Our model is

$$y_t(n) = \beta_n y_{t-1}(n) + \varepsilon_t, \quad t \geq 2 \quad (3.4.1)$$

where $\varepsilon_t \sim NID(0, \sigma^2)$ and n is the sample size. To construct the percentiles, we let $y_1 = 0$, and generate the ε_t 's as independent standard normal random variables. The RANNOR function in SAS(V8) is used to generate the ε_t 's. For a given sample size n , a given value of γ and for $\beta_n = 1 - \frac{\gamma}{n}$, we generated 10,000 samples of size n and computed the OLS statistic $\tau_{OLS,n}^{(1)}$ and the WS statistic $\tau_{WS,n}^{(1)}$. The statistics, $\tau_{OLS,n}^{(2)}$ and $\tau_{WS,n}^{(2)}$ were computed by putting $\beta_n = 1 + \frac{\gamma}{n}$. The empirical critical values $c_{OLS,n}^{(1)}$ and $c_{OLS,n}^{(2)}$, for a 5%-level IUT based on the OLS statistic, were computed as the 95th percentile of the distribution of $\tau_{OLS,n}^{(1)}$ and the 5th percentile of the distribution of $\tau_{OLS,n}^{(2)}$ respectively. The following table reports $c_{OLS,n}^{(1)}$ and $c_{OLS,n}^{(2)}$ for $n=100, 500$ and 1000 and $\gamma = 1, 2, 3, 4, 5$.

Table 3-1: 95th percentile based on OLS statistic $\tau_{OLS,n}^{(2)}$

Test statistic	γ	n		
		100	500	1000
$\tau_{OLS,n}^{(1)}$	1	1.27	1.27	1.25
	2	1.27	1.28	1.26
	3	1.30	1.30	1.25
	4	1.37	1.32	1.26
	5	1.35	1.37	1.34

Table 3-2: 5th percentile based on OLS statistic $\tau_{OLS,n}^{(2)}$

Test statistic	γ	n		
		100	500	1000
$\tau_{OLS,n}^{(2)}$	1	-1.97	-2.00	-2.00
	2	-2.09	-2.13	-2.11
	3	-2.33	-2.31	-2.34
	4	-2.23	-2.15	-2.18
	5	-2.03	-1.94	-2.04

Similarly, the empirical critical values $c_{WS,n}^{(1)}$ and $c_{WS,n}^{(2)}$, for a 5%-level IUT based on the WS statistic, were computed as the 95th percentile of the distribution of $\tau_{WS,n}^{(1)}$ and the 5th percentile of the distribution of $\tau_{WS,n}^{(2)}$ respectively.

Table 3-3: 95th percentile based on OLS statistic $\tau_{WS,n}^{(1)}$

Test statistic	γ	n		
		100	500	1000
$\tau_{WS,n}^{(1)}$	1	0.41	0.42	0.44
	2	0.57	0.55	0.56
	3	0.71	0.68	0.68
	4	0.84	0.78	0.77
	5	0.94	0.88	0.84

Table 3-4: 5th percentile based on OLS statistic $\tau_{WS,n}^{(2)}$

Test statistic	γ	n		
		100	500	1000
$\tau_{WS,n}^{(2)}$	1	-2.38	-2.36	-2.35
	2	-3.24	-3.34	-3.36
	3	-4.18	-5.46	-5.78
	4	-4.30	-7.31	-8.90
	5	-3.95	-7.53	-10.08

(b) Power Study of the OLS statistic

Table 3-5 through Table 3-7 are based on 10,000 replications. In each row of the tables below, $P(R|\Theta_1)$ gives the probability of rejection of the IUT when $\beta_n = 1 - \frac{\gamma}{n}$. $P(R|\Theta_1)$ gives the probability of rejection of the IUT when $\beta_n = 1 + \frac{\gamma}{n}$. The maximum of the columns $P(R|\Theta_1)$ and $P(R|\Theta_2)$ represent the empirical type I error associated with the IUT proposed in Section 2. $P(R_1|\beta_n = 1)$ and $P(R_2|\beta_n = 1)$ represent respectively, the power of the level- α tests given by the rejection regions R_1 and R_2 (in Section 2 above) at $\beta_n = 1$ for testing $H: \beta_n \leq 1 - \frac{\gamma}{n}$ against $K: \beta_n > 1 - \frac{\gamma}{n}$ and that for testing $H: \beta_n \geq 1 + \frac{\gamma}{n}$ against $K: \beta_n < 1 + \frac{\gamma}{n}$. The power of the IUT at $\beta_n = 1$ is given by $P(R|\beta_n = 1)$. For a fixed value of n , as we move from $\gamma = 1$ to $\gamma = 5$, our null hypothesis changes from $H_{0n}: \beta_n \leq 1 - \frac{1}{n}$ or $\beta_n \geq 1 + \frac{1}{n}$ to $H_{0n}: \beta_n \leq 1 - \frac{5}{n}$ or $\beta_n \geq 1 + \frac{5}{n}$ i.e., the parameter space under the null hypothesis moves away from the unit root alternative which is a part of the composite alternative hypothesis. On the other hand, if we fix γ , and increase n from 100 to 1000, our null hypothesis changes

from $H_{0n} : \beta_n \leq 1 - \frac{\gamma}{100}$ or $\beta_n \geq 1 + \frac{\gamma}{100}$ to $H_{0n} : \beta_n \leq 1 - \frac{\gamma}{1000}$ or $\beta_n \geq 1 + \frac{\gamma}{1000}$, i.e., the parameter space under the null hypothesis moves closer towards the unit root alternative.

For a particular value of γ , the empirical size of the IUT is maintained at approximately the same level when the sample size n increases from 100 to 500 to 1000. The IUT rejection region R has a consistently higher probability under Θ_1 than under Θ_2 . While the IUT fails to maintain the size close to 0.05 for $\gamma = 1, 2$ and 3, the test performs much better in terms of size for $\gamma = 4$ and 5. This observation holds true for $n=100, 500$ and 1000.

The failure to maintain the size of the test at $\gamma = 1, 2$ and 3 adversely affects the power of the IUT. For instance, at $\gamma = 1$, the power of the test under the unit root alternative stays at around 0.045 as n increases from 100 to 1000. But as the size of the test improves at $\gamma = 4$, the power of the test jumps to approximately 0.47. And at $\gamma = 5$, the power increases to around 0.62.

We can further look at the performance of the individual rejection regions R_1 and R_2 that constitute R . For any fixed value of γ , $P(R_1 | \beta_n = 1)$ and $P(R_2 | \beta_n = 1)$ do not change significantly as n increases from 100 to 1000. For a fixed value of n , $P(R_1 | \beta_n = 1)$ increases from around 0.2 for $\gamma = 1$ to around 0.65 for $\gamma = 5$ and $P(R_2 | \beta_n = 1)$ increases from 0 for $\gamma = 1$ to more than 0.95 for $\gamma = 5$.

Table 3-5: Size and Power of the IUT based on OLS, $n = 100$

γ	$1-\gamma/n$	$1+\gamma/n$	$P(R \Theta_1)$	$P(R \Theta_2)$	$P(R_1 \beta_n=1)$	$P(R_2 \beta_n=1)$	$P(R \beta_n=1)$
1	0.99	1.01	0.000	0.000	0.196	0.084	0.000
2	0.98	1.02	0.005	0.003	0.354	0.203	0.043
3	0.97	1.03	0.020	0.013	0.466	0.520	0.222
4	0.96	1.04	0.040	0.024	0.575	0.850	0.470
5	0.95	1.05	0.046	0.036	0.649	0.967	0.620

Table 3-6: Size and Power of IUT based on OLS, $n = 500$

γ	$1-\gamma/n$	$1+\gamma/n$	$P(R \Theta_1)$	$P(R \Theta_2)$	$P(R_1 \beta_n=1)$	$P(R_2 \beta_n=1)$	$P(R \beta_n=1)$
1	0.998	1.002	0.000	0.000	0.200	0.082	0.000
2	0.996	1.004	0.004	0.004	0.350	0.180	0.044
3	0.994	1.006	0.021	0.013	0.474	0.539	0.233
4	0.992	1.008	0.040	0.027	0.576	0.874	0.483
5	0.990	1.010	0.045	0.039	0.650	0.978	0.630

Table 3-7: Size and Power based of IUT based on OLS, $n = 1000$

γ	$1-\gamma/n$	$1+\gamma/n$	$P(R \Theta_1)$	$P(R \Theta_2)$	$P(R_1 \beta_n=1)$	$P(R_2 \beta_n=1)$	$P(R \beta_n=1)$
1	0.999	1.001	0.000	0.000	0.204	0.082	0.000
2	0.998	1.002	0.005	0.004	0.358	0.191	0.049
3	0.997	1.003	0.020	0.011	0.485	0.512	0.233
4	0.996	1.004	0.047	0.028	0.592	0.867	0.491
5	0.995	1.005	0.044	0.034	0.661	0.968	0.632

(c) Power Study of the WS statistic

The empirical size and the power of the IUT based on the WS statistic is reported in the following tables for various sample sizes. The size and the power for a fixed value of γ do not change appreciably as the sample size increases from 100 to 1000.

The power at $\beta_n = 1$ obtained from the IUT based on the WS statistic is well below 0.1 for all the combinations of γ and n considered below except for $\gamma = 4, 5$ when $n = 100$. However the rejection region R_1 has a consistently higher probability mass than R_2 when the data is generated from a unit root process. Moreover, for a given sample size n , $P(R_1 | \beta_n = 1)$ increases steadily as γ increases from 1 to 5. But $P(R_2 | \beta_n = 1)$ turns out to be very low and does not exhibit any trend with increasing γ . These contrasting features of the two separate regions that constitute R explain the extremely low power of the test in detecting a unit root.

Table 3-8: Size and Power based on WS statistic, $n = 100$

γ	$1 - \gamma/n$	$1 + \gamma/n$	$P(R \Theta_1)$	$P(R \Theta_2)$	$P(R_1 \beta_n = 1)$	$P(R_2 \beta_n = 1)$	$P(R \beta_n = 1)$
1	0.99	1.01	0.001	0.009	0.192	0.075	0.006
2	0.98	1.02	0.002	0.047	0.356	0.055	0.046
3	0.97	1.03	0.000	0.050	0.491	0.069	0.069
4	0.96	1.04	0.001	0.050	0.588	0.180	0.173
5	0.95	1.05	0.014	0.049	0.672	0.235	0.228

Table 3-9: Size and Power based on WS statistic, $n = 500$

γ	$1-\gamma/n$	$1+\gamma/n$	$P(R \Theta_1)$	$P(R \Theta_2)$	$P(R_1 \beta_n=1)$	$P(R_2 \beta_n=1)$	$P(R \beta_n=1)$
1	0.998	1.002	0.000	0.008	0.191	0.079	0.006
2	0.996	1.004	0.001	0.047	0.360	0.042	0.036
3	0.994	1.006	0.000	0.050	0.495	0.010	0.010
4	0.992	1.008	0.000	0.050	0.592	0.007	0.007
5	0.990	1.010	0.000	0.050	0.672	0.027	0.027

Table 3-10: Size and Power based on WS statistic, $n = 1000$

γ	$1-\gamma/n$	$1+\gamma/n$	$P(R \Theta_1)$	$P(R \Theta_2)$	$P(R_1 \beta_n=1)$	$P(R_2 \beta_n=1)$	$P(R \beta_n=1)$
1	0.999	1.001	0.000	0.007	0.1878	0.0813	0.0048
2	0.998	1.002	0.001	0.048	0.3665	0.0000	0.0000
3	0.997	1.003	0.000	0.051	0.4991	0.0053	0.0053
4	0.996	1.004	0.000	0.050	0.5984	0.0005	0.0005
5	0.995	1.005	0.000	0.050	0.6789	0.0015	0.0015

As we have seen from simulation studies, our test based on the OLS will have the right size if we use the correct hypothesis based on γ and n . For a fixed γ , as the sample size increases, the appropriate null hypothesis contains increasingly more stationary β values that are increasingly closer to 1. Depending upon the value of γ , we get a power that stays at the same level as we increase n from 100 to 1000. For instance, when $\gamma = 5$, the power to detect a unit root stays at about 63% as we keep increasing the sample size. This does not imply that the test does not do better as the sample size increases. On the contrary, at $\gamma = 5$, the range of stationary or explosive processes that the test can differentiate from a unit root process widens from $(-\infty, 0.95) \cup (1.05, \infty)$ at $n = 100$ to $(-\infty, 0.995) \cup (1.005, \infty)$ at $n = 1000$. But in practice, the use of the test will be limited if

we have to re-compute the percentiles for a $\gamma = 5$ whenever we have a different sample size. So if we can tabulate the asymptotic percentiles of the test statistics for $\gamma = 5$ and verify its applicability for increasing sample sizes, we will have a useful test for stationarity. Note that it is not necessary to tabulate the percentiles for every value of γ . Since one would always be interested in having a test that has reasonable power of distinguishing between near-unit root processes and the unit root process, it would be quite helpful if we had the asymptotic percentiles tabulated for $\gamma = 5, 4$ and 3 . Larger values of γ would be unnecessary because they will be farther away from the unit root process. On the other hand, smaller values of γ (2 or less) will be useless since simulations indicate that our tests based on such low values of γ have essentially no power at all to distinguish between such processes that have their AR root extremely close to 1 and the unit root AR process. To obtain the asymptotic percentiles for our IUT test, we will use the results of Chan (1988).

Chan (1988) provided three different representations of the limiting distribution $\mathcal{L}(\gamma)$. Of the three representations, Chan (1988) advocated the use of the representation given by

$$\mathcal{L}(\gamma) \stackrel{D}{=} \frac{\int_0^1 \tilde{X}(t) d\tilde{W}(t)}{\left\{ \int_0^1 \tilde{X}^2(t) dt \right\}^{1/2}}, \quad (3.4.2)$$

where $\tilde{X}(t)$ is the Ornstein-Uhlenbeck process satisfying

$$d\tilde{X}(t) = -\gamma \tilde{X}(t) dt + d\tilde{W}(t), \quad \tilde{X}(0) = 0, \quad (3.4.3)$$

where $\tilde{W}(t)$ is a standard Brownian motion, and $\stackrel{D}{=}$ indicates equality in distribution.

The expression in (3.4.2) is approximated by the finite sum,

$$\mathcal{L}(\gamma) = \left(\sum_{i=1}^n \sum_{k=1}^i e^{-\gamma(i-k)/n} \varepsilon_k \varepsilon_{i+1} \right) \div \left\{ \sum_{i=1}^n \left(\sum_{k=1}^i \left(e^{-\gamma(i-k)/n} \varepsilon_k \right)^2 \right) \right\}^{1/2}, \quad (3.4.4)$$

where $\{\varepsilon_k\}$ is a sequence of random variables satisfying (3.3.1).

Putting $n = 1000$, we simulated the asymptotic percentiles for various values of γ , each based on 10,000 replications. These percentiles are listed in the following table.

Table 3-11: Asymptotic Percentiles corresponding to various values of γ

	Probability					
	1%	5%	10%	90%	95%	99%
γ						
1	-2.55	-1.92	-1.59	0.85	1.21	1.96
2	-2.52	-1.89	-1.55	0.87	1.21	1.92
3	-2.55	-1.87	-1.54	0.9	1.23	1.91
4	-2.52	-1.85	-1.53	0.92	1.28	1.93
5	-2.51	-1.84	-1.52	0.95	1.3	1.95
100	-2.42	-1.75	-1.35	1.21	1.56	2.2
0.01	-2.59	-1.94	-1.63	0.88	1.25	1.99
-1	-2.6	-2.01	-1.68	0.92	1.3	2.02
-2	-2.69	-2.13	-1.92	1	1.35	2.07
-3	-2.8	-2.31	-1.9	1.06	1.42	2.13
-4	-2.89	-2.23	-1.73	1.12	1.49	2.18
-5	-2.97	-1.97	-1.53	1.19	1.54	2.25
-100	-2.38	-1.63	-1.27	1.29	1.65	2.29
-0.01	-2.59	-1.94	-1.63	0.88	1.25	1.99

Note that the percentiles in the above table for $\gamma > 0$ are the asymptotic percentiles for $\tau_{OLS,n}^{(1)}$ and the percentiles for $\gamma < 0$ are the asymptotic percentiles for $\tau_{OLS,n}^{(2)}$.

Table 3-12 below shows that we can indeed use the asymptotic percentiles obtained above to conduct the IUT test for finite sample sizes. The size of the test is maintained at the nominal level of 5% and the power is very close to that obtained using finite sample percentiles.

Table 3-12: Rejection probabilities under the null and alternative hypothesis based on asymptotic percentiles

	β								
	$1-\gamma/n$			$1+\gamma/n$			1		
	γ			γ			γ		
	3	4	5	3	4	5	3	4	5
n									
100	0.03	0.04	0.06	0.02	0.04	0.05	0.22	0.45	0.62
500	0.02	0.04	0.04	0.02	0.03	0.05	0.23	0.45	0.66
1000	0.02	0.03	0.05	0.02	0.02	0.05	0.22	0.47	0.64

It will also be useful to analyze the performance of the IUT test in the following manner: Fix the null hypothesis and the corresponding test statistic, use the asymptotic percentiles based on some γ for the rejection rule. Obtain the rejection probabilities under the null and the alternative hypotheses for various sample sizes. From such an analysis, we would expect to see an increase in the rejection probabilities as the sample size increases for a given γ and a given null hypothesis and a fixed test statistic. That would translate into an increase in power to detect a unit root, accompanied by an overestimation of the nominal size of the test. However, from such a study, we also expect to find out the extent of overestimation of the nominal size at different sample sizes and also at different values of γ . A few interesting null hypotheses would be:

- a) $H_0 : \beta < 0.95$ or $\beta > 1.05$
- b) $H_0 : \beta < 0.96$ or $\beta > 1.04$
- c) $H_0 : \beta < 0.97$ or $\beta > 1.03$
- d) $H_0 : \beta < 0.98$ or $\beta > 1.02$
- e) $H_0 : \beta < 0.99$ or $\beta > 1.01$

The five pairs of test statistics appropriate for each of the above null hypotheses are defined as:

$$\begin{aligned}
\text{TS1: } \tau_{OLS,n}^{(1)} &= \left(\frac{\hat{\beta}_{OLS,n} - (0.95)}{s} \right) \left(\sum_{t=2}^n Y_{t-1}^2 \right)^{\frac{1}{2}} \text{ and } \tau_{OLS,n}^{(2)} = \left(\frac{\hat{\beta}_{OLS,n} - (1.05)}{s} \right) \left(\sum_{t=2}^n Y_{t-1}^2 \right)^{\frac{1}{2}} \\
\text{TS2: } \tau_{OLS,n}^{(1)} &= \left(\frac{\hat{\beta}_{OLS,n} - (0.96)}{s} \right) \left(\sum_{t=2}^n Y_{t-1}^2 \right)^{\frac{1}{2}} \text{ and } \tau_{OLS,n}^{(2)} = \left(\frac{\hat{\beta}_{OLS,n} - (1.04)}{s} \right) \left(\sum_{t=2}^n Y_{t-1}^2 \right)^{\frac{1}{2}} \\
\text{TS3: } \tau_{OLS,n}^{(1)} &= \left(\frac{\hat{\beta}_{OLS,n} - (0.97)}{s} \right) \left(\sum_{t=2}^n Y_{t-1}^2 \right)^{\frac{1}{2}} \text{ and } \tau_{OLS,n}^{(2)} = \left(\frac{\hat{\beta}_{OLS,n} - (1.03)}{s} \right) \left(\sum_{t=2}^n Y_{t-1}^2 \right)^{\frac{1}{2}} \\
\text{TS4: } \tau_{OLS,n}^{(1)} &= \left(\frac{\hat{\beta}_{OLS,n} - (0.98)}{s} \right) \left(\sum_{t=2}^n Y_{t-1}^2 \right)^{\frac{1}{2}} \text{ and } \tau_{OLS,n}^{(2)} = \left(\frac{\hat{\beta}_{OLS,n} - (1.02)}{s} \right) \left(\sum_{t=2}^n Y_{t-1}^2 \right)^{\frac{1}{2}} \\
\text{TS5: } \tau_{OLS,n}^{(1)} &= \left(\frac{\hat{\beta}_{OLS,n} - (0.99)}{s} \right) \left(\sum_{t=2}^n Y_{t-1}^2 \right)^{\frac{1}{2}} \text{ and } \tau_{OLS,n}^{(2)} = \left(\frac{\hat{\beta}_{OLS,n} - (1.01)}{s} \right) \left(\sum_{t=2}^n Y_{t-1}^2 \right)^{\frac{1}{2}}
\end{aligned}$$

The seven different rejection criteria based on the seven different pairs of γ can be stated as follows:

RC1: Based of $\gamma = \pm 0.01$: Reject H_0 if $\tau_{OLS,n}^{(1)} > 1.2474839$ and $\tau_{OLS,n}^{(2)} < -1.940104$.

RC2: Based of $\gamma = \pm 1$: Reject H_0 if $\tau_{OLS,n}^{(1)} > 1.2125599$ and $\tau_{OLS,n}^{(2)} < -2.599016$.

RC3: Based of $\gamma = \pm 2$: Reject H_0 if $\tau_{OLS,n}^{(1)} > 1.2093295$ and $\tau_{OLS,n}^{(2)} < -2.133968$.

RC4: Based of $\gamma = \pm 3$: Reject H_0 if $\tau_{OLS,n}^{(1)} > 1.2324658$ and $\tau_{OLS,n}^{(2)} < -2.311191$.

RC5: Based of $\gamma = \pm 4$: Reject H_0 if $\tau_{OLS,n}^{(1)} > 1.2796976$ and $\tau_{OLS,n}^{(2)} < -2.891987$.

RC6: Based of $\gamma = \pm 5$: Reject H_0 if $\tau_{OLS,n}^{(1)} > 1.3048007$ and $\tau_{OLS,n}^{(2)} < -2.969209$.

RC7: Based of $\gamma = \pm 100$: Reject H_0 if $\tau_{OLS,n}^{(1)} > 1.5598589$ and $\tau_{OLS,n}^{(2)} < -1.628943$.

Table 3-13 through Table 3-17 show the performance of the various test statistics TS1 through TS5 when combined with the different rejection rules RC1 through RC7 for sample sizes ranging from $n = 100$ to $n = 1000$. Every rejection frequency in these tables is based on 1000 replications.

The probabilities in the column for $\beta = 1$ give the power of the test to reject the null hypothesis in favor of the unit root alternative hypothesis. The maximum of the other two columns for β (for a fixed n and a fixed γ) gives the level of test.

All the seven RC (except RC 7) have approximately the same power for a fixed sample size and a fixed null hypothesis. RC 7 has slightly lesser power than the other rejection criteria for all the different cases. Power of identifying a unit root being almost the same across all γ 's, it is essential that we compare the estimated sizes of these RC's across several hypotheses and sample sizes. As the sample size increases, the estimated size of any particular test based on a particular γ increases and overshoots the nominal size. This implies that as the sample size increases, we should consider null hypotheses that are closer to the unit root alternative to control the size of the test. Such an implication is confirmed by the decreasing trend in the size of the test for a particular n for successively wider null hypothesis (Table 3-13 through Table 3-17).

The simulation results advise against using a rejection criterion based on a very large $\gamma (=100)$ to test any of the five hypotheses listed above when the sample size ranges from 100 to 1000. The rejection criterion RC 6 (based on $\gamma = \pm 5$) seems to be an appropriate rejection criterion to use under any of the null hypotheses considered here. It delivers a power that is virtually as high as any other rejection criterion, and it also maintains the size of the test closer to the nominal size of the test than the other rejection criteria under most of the scenarios (combination of a null hypothesis and a sample size). Over all the different combinations of the null hypothesis and the sample size, the rejection criterion RC 6 overestimates the nominal level of the test to be 0.08 in the worst cases. The overestimation would have only worsened if we had considered even higher sample sizes. However it is important to note that the simulation study demonstrates that

even when the null hypothesis ($H_0: 0.95 \leq \beta \leq 1.05$) and the rejection criteria (for instance, RC 6) are inappropriately fixed regardless of the sample size, the extent of overestimation of the nominal size is not very high.

Table 3-13: Size and Power of TS1 for seven different DC and three different sample sizes

	β																				
	0.95							1							1.05						
	RC							RC							RC						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
n																					
100	0.06	0.07	0.07	0.06	0.06	0.06	0.03	0.62	0.63	0.62	0.59	0.59	0.62	0.59	0.05	0.05	0.05	0.03	0.04	0.05	0.08
500	0.08	0.09	0.09	0.08	0.07	0.07	0.03	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.04	0.03	0.02	0.02	0.02	0.03	0.06
1000	0.08	0.09	0.09	0.09	0.08	0.08	0.04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.06	0.06	0.06	0.06	0.06	0.06	0.06

Table 3-14: Size and Power of TS2 for seven different DC and three different sample sizes

	β																				
	0.96							1							1.04						
	RC							RC							RC						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
n																					
100	0.06	0.05	0.05	0.05	0.04	0.05	0.03	0.50	0.49	0.47	0.44	0.45	0.49	0.48	0.06	0.05	0.04	0.03	0.04	0.05	0.09
500	0.07	0.08	0.08	0.08	0.07	0.06	0.03	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.04	0.03	0.02	0.01	0.02	0.03	0.06
1000	0.09	0.09	0.09	0.09	0.08	0.08	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.08	0.08	0.08	0.08	0.08	0.08	0.08

Table 3-15: Size and Power of TS3 for seven different DC and three different sample sizes

	β																				
	0.97							1							1.03						
	RC							RC							RC						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
n																					
100	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.31	0.30	0.27	0.22	0.23	0.30	0.32	0.04	0.03	0.03	0.02	0.02	0.03	0.06
500	0.06	0.07	0.07	0.06	0.06	0.05	0.03	0.96	0.96	0.96	0.96	0.96	0.96	0.95	0.03	0.03	0.02	0.01	0.02	0.03	0.06
1000	0.08	0.09	0.09	0.08	0.08	0.07	0.04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.03	0.03	0.02	0.02	0.02	0.03	0.06

Table 3-16: Size and Power of TS4 for seven different DC and three different sample sizes

	β																				
	0.98							1							1.02						
	RC							RC							RC						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
n																					
100	0.01	0.01	0.01	0.00	0.00	0.01	0.01	0.05	0.05	0.03	0.02	0.02	0.05	0.09	0.01	0.01	0.01	0.00	0.00	0.01	0.03
500	0.05	0.06	0.06	0.05	0.05	0.05	0.03	0.89	0.89	0.89	0.89	0.89	0.89	0.86	0.04	0.03	0.02	0.01	0.02	0.03	0.07
1000	0.07	0.08	0.08	0.07	0.07	0.07	0.04	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.03	0.03	0.02	0.01	0.01	0.03	0.06

Table 3-17: Size and Power of TS5 for seven different DC and three different sample sizes

	β																				
	0.99							1							1.01						
	RC							RC							RC						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
n																					
100	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
500	0.05	0.05	0.05	0.05	0.04	0.04	0.03	0.67	0.67	0.66	0.63	0.64	0.66	0.61	0.05	0.04	0.03	0.02	0.03	0.05	0.09
1000	0.07	0.07	0.07	0.07	0.06	0.06	0.03	0.90	0.91	0.91	0.91	0.90	0.90	0.88	0.04	0.03	0.02	0.02	0.02	0.04	0.08

Section 5. A different sequence of Hypothesis

Instead of using the sequence of sub-problems given by (3.2.3) we can also use the following sequence.

$$H_{0n} : \beta_n \leq 1 - \frac{\gamma}{\sqrt{n}} \text{ or } \beta_n \geq 1 + \frac{\gamma}{\sqrt{n}} \text{ versus } H_{1n} : 1 - \frac{\gamma}{\sqrt{n}} < \beta_n < 1 + \frac{\gamma}{\sqrt{n}} \quad (3.4.5)$$

This sequence also approaches (3.1.1) but at a slower rate compared to (3.2.3). For a fixed sample size n and a fixed γ , the null hypothesis in (3.4.5) is farther away than the null hypothesis in (3.2.3) from the unit root alternative. One would thus expect to see better power when testing (3.4.5) compared to the power obtained when testing (3.2.3) for the same sample size n and the same parameter γ . For small sample sizes, the null hypothesis in (3.4.5) will be a significantly smaller subset of the null hypothesis in (3.2.3) and hence the gain in power will not be very valuable. However, the null hypothesis in (3.4.5) can be driven as close to the unit root alternative as we desire by choosing a suitably large sample size n . Hence the gain in power, if substantial, can prove to be very useful for large sample sizes.

Corresponding to (3.4.5), the OLS statistics are $\tau_{OLS, \sqrt{n}}^{(1)} = \left(\frac{\hat{\beta}_n - \left(1 - \frac{\gamma}{\sqrt{n}}\right)}{s} \right) \left(\sum_{t=2}^n Y_{t-1}^2 \right)^{\frac{1}{2}}$ and

$$\tau_{OLS, \sqrt{n}}^{(2)} = \left(\frac{\hat{\beta}_n - \left(1 + \frac{\gamma}{\sqrt{n}}\right)}{s} \right) \left(\sum_{t=2}^n Y_{t-1}^2 \right)^{\frac{1}{2}}. \text{ The WS statistics modify to}$$

$$\tau_{ws, \sqrt{n}}^{(1)} = \left[\hat{\beta}_{ws} - \left(1 - \frac{\gamma}{\sqrt{n}}\right) \right] \left[\sum_{t=2}^{n-1} y_t^2 + n^{-1} \sum_{t=1}^n y_t^2 \right]^{1/2} \hat{\sigma}_{ws}^{-1}$$

and

$$\tau_{ws, \sqrt{n}}^{(2)} = \left[\hat{\beta}_{ws} - \left(1 + \frac{\gamma}{\sqrt{n}}\right) \right] \left[\sum_{t=2}^{n-1} y_t^2 + n^{-1} \sum_{t=1}^n y_t^2 \right]^{1/2} \hat{\sigma}_{ws}^{-1}$$

We first present the simulated percentiles for a 5%-level IUT based on the OLS and the WS statistics.

All the percentiles presented below were obtained using SAS (V8). The percentiles presented in Table 3-19 raise about the computational accuracy provided by SAS when handling explosive time series. The 5th percentile of the distribution of $\tau_{OLS,\sqrt{n}}^{(2)}$ for $n=500$ and $n=1000$ show marked difference from what we would expect from the established theory of explosive time series. Except for $\gamma = 1$, the percentiles suggest that the distribution of the OLS statistic is not converging to that of a standard normal variable as the sample size increases, as has been proved in Fuller (1996).

The percentiles for $\tau_{ws,\sqrt{n}}^{(2)}$ presented in Table 3-21 also exhibit an unusual behavior. Although the values seem to stabilize as the sample size increases, the simulated distribution of $\tau_{ws,\sqrt{n}}^{(2)}$ has zero variance for all the three sample sizes that we have considered. We are not certain as to whether this behavior can be explained theoretically or whether such a behavior is a result of the computational limitations of SAS.

Table 3-18: 95th percentile based on the distribution of $\tau_{OLS,\sqrt{n}}^{(1)}$

Test statistic	γ	n		
		100	500	1000
$\tau_{OLS,\sqrt{n}}^{(1)}$	1	1.57	1.56	1.60
	2	1.54	1.55	1.58
	3	1.54	1.58	1.55
	4	1.46	1.57	1.59
	5	1.41	1.50	1.51

Table 3-19: 5th percentile based on the distribution of $\tau_{OLS,\sqrt{n}}^{(2)}$

Test	γ	n		
		100	500	1000
$\tau_{OLS,\sqrt{n}}^{(2)}$	1	-1.64	-1.66	-1.64
	2	-1.64	-23.08	-32.51
	3	-1.64	-23.32	-32.71
	4	-1.64	-23.77	-32.00
	5	-11.16	-24.02	-33.43

Table 3-20: 95th percentile based on the distribution of $\tau_{ws,\sqrt{n}}^{(1)}$

Test statistic	γ	n		
		100	500	1000
$\tau_{ws,\sqrt{n}}^{(1)}$	1	1.56	1.53	1.54
	2	1.54	1.50	1.52
	3	1.47	1.45	1.50
	4	1.37	1.43	1.45
	5	1.19	1.31	1.33

Table 3-21: 5th percentile based on the distribution of $\tau_{ws,\sqrt{n}}^{(2)}$

Test statistic	γ	n		
		100	500	1000
$\tau_{ws,\sqrt{n}}^{(2)}$	1	-2.71	-3.65	-4.23
	2	-2.19	-2.76	-3.14
	3	-2.04	-2.41	-2.69
	4	-2.00	-2.23	-2.45
	5	-2.01	-2.13	-2.30

Due to reasons cited above, it does not make sense to study the simulated power properties of the IUT based on the OLS or WS for the hypothesis formulated in (3.4.5).

However, we studied yet another sequence of hypotheses that converges to (3.1.1). In the following sequence of hypotheses,

$$H_{0n} : \beta_n \leq 1 - \frac{\gamma}{\sqrt{n}} \text{ or } \beta_n \geq 1 + \frac{\gamma}{\sqrt{n}} \text{ versus } H_{1n} : 1 - \frac{\gamma}{\sqrt{n}} < \beta_n < 1 + \frac{\gamma}{\sqrt{n}}, \quad (3.4.6)$$

the interval around the unit root in H_{1n} is asymmetric in nature. For a fixed value of γ , the upper limit $1 + \frac{\gamma}{\sqrt{n}}$ and the lower limit $1 - \frac{\gamma}{\sqrt{n}}$ of the interval around $\beta_n = 1$ converge to 1 at different rates.

Now, the OLS statistics are $\tau_{OLS, \sqrt{n}}^{(1)} = \left(\frac{\hat{\beta}_n - \left(1 - \frac{\gamma}{\sqrt{n}}\right)}{s} \right) \left(\sum_{t=2}^n Y_{t-1}^2 \right)^{\frac{1}{2}}$ and

$$\tau_{OLS, n}^{(2)} = \left(\frac{\hat{\beta}_n - \left(1 + \frac{\gamma}{n}\right)}{s} \right) \left(\sum_{t=2}^n Y_{t-1}^2 \right)^{\frac{1}{2}}. \text{ The WS statistics modify to}$$

$$\tau_{WS, \sqrt{n}}^{(1)} = \left[\hat{\beta}_{WS} - \left(1 - \frac{\gamma}{\sqrt{n}}\right) \right] \left[\sum_{t=2}^{n-1} y_t^2 + n^{-1} \sum_{t=1}^n y_t^2 \right]^{1/2} \hat{\sigma}_{WS}^{-1}$$

and

$$\tau_{WS, n}^{(2)} = \left[\hat{\beta}_{WS} - \left(1 + \frac{\gamma}{n}\right) \right] \left[\sum_{t=2}^{n-1} y_t^2 + n^{-1} \sum_{t=1}^n y_t^2 \right]^{1/2} \hat{\sigma}_{WS}^{-1}$$

We have already simulated the percentiles for $\tau_{OLS, \sqrt{n}}^{(1)}, \tau_{OLS, n}^{(2)}, \tau_{WS, \sqrt{n}}^{(1)}, \tau_{WS, n}^{(2)}$ Table 3-18, Table 3-2, Table 3-20 and Table 3-4 respectively. Based on these percentiles, we simulated the size and power of the IUT. We report the size and power of the IUT based on OLS in Table 3-22, Table 3-23, and Table 3-24. The WS-based IUT was not investigated any further because of the poor power performance demonstrated in Table 3-8, Table 3-9 and Table 3-10.

Table 3-22: Size and Power of the IUT based on $\tau_{OLS,\sqrt{n}}^{(1)}$ and $\tau_{OLS,n}^{(2)}$, $n = 100$

γ	$1 - \frac{\gamma}{\sqrt{n}}$	$1 + \gamma/n$	$P(R \Theta_1)$	$P(R \Theta_2)$	$P(R_1 \beta_n=1)$	$P(R_2 \beta_n=1)$	$P(R \beta_n=1)$
1	0.90	1.01	0.00	0.00	0.88	0.08	0.00
2	0.80	1.02	0.05	0.04	0.98	0.19	0.18
3	0.70	1.03	0.05	0.05	0.99	0.51	0.51
4	0.60	1.04	0.05	0.04	0.99	0.84	0.84
5	0.50	1.05	0.05	0.05	1.00	0.97	0.97

Table 3-23: Size and Power of the IUT based on $\tau_{OLS,\sqrt{n}}^{(1)}$ and $\tau_{OLS,n}^{(2)}$, $n = 500$

γ	$1 - \frac{\gamma}{\sqrt{n}}$	$1 + \gamma/n$	$P(R \Theta_1)$	$P(R \Theta_2)$	$P(R_1 \beta_n=1)$	$P(R_2 \beta_n=1)$	$P(R \beta_n=1)$
1	0.96	1.002	0.00	0.00	0.99	0.08	0.07
2	0.91	1.004	0.05	0.04	1.00	0.19	0.19
3	0.87	1.006	0.05	0.05	1.00	0.53	0.53
4	0.82	1.008	0.05	0.04	1.00	0.87	0.87
5	0.77	1.010	0.05	0.05	1.00	0.97	0.97

Table 3-24: Size and Power of the IUT based on $\tau_{OLS,\sqrt{n}}^{(1)}$ and $\tau_{OLS,n}^{(2)}$, $n = 1000$

γ	$1 - \frac{\gamma}{\sqrt{n}}$	$1 + \gamma/n$	$P(R \Theta_1)$	$P(R \Theta_2)$	$P(R_1 \beta_n=1)$	$P(R_2 \beta_n=1)$	$P(R \beta_n=1)$
1	0.97	1.001	0.00	0.00	0.99	0.08	0.07
2	0.94	1.002	0.05	0.04	1.00	0.19	0.19
3	0.91	1.003	0.05	0.05	1.00	0.53	0.53
4	0.87	1.004	0.05	0.04	1.00	0.87	0.87
5	0.84	1.001	0.05	0.05	1.00	0.97	0.97

Section 6. Conclusions

We have presented an application of the IUT theory to the testing of the null hypothesis of stationarity or explosiveness in AR (1) process. This null hypothesis is viewed as the limit of a particular sequence of sub-hypotheses. The structure of these sub-hypotheses lends itself to the construction of a test criterion using the IUT method. Simulation studies indicate that for all practical purposes, we may use the asymptotic percentiles based on $\gamma = \pm 5$ to distinguish between near unit root processes and a unit root process. However, with increasing sample size, our null hypothesis needs to move in closer to the unit root alternative in order to ensure that the nominal size of the test is not over-estimated by the IUT. This requirement is not restrictive at all, because a researcher would always want to distinguish between the unit-root process and another process, which is as close to the unit-root process as possible. Even in situations where this requirement is not met and the null hypothesis is kept fixed while the sample size increases ten-fold, our simulation studies show that the size distortion is not too bad. One recommendation would be to set $\gamma = \pm 5$ in the null hypothesis (3.2.3) and use the test statistics $\tau_{OLS,n}^{(1)}$ and $\tau_{OLS,n}^{(2)}$ to perform the IUT. Thus, we will always have a fixed set of percentiles under the null hypothesis whereas the test statistics will be appropriately determined by the sample size.

Bibliography

Abramowitz, M and Stegun, N.C. (1970), *Handbook of Mathematical Functions* (New York: Dover Publications Inc.).

Ahtola, J. A., and Tiao, G.C. (1984), "Parametric inference for a Nearly Non-stationary First Order autoregressive Model", *Biometrika*, 71, 263-272.

Apostol, Tom M. (1973), "Mathematical Analysis", 2nd edition, Indian Student Edition, Narosa Publishing House.

Arellano, C., and Pantula, S. G., (1995), Testing for trend stationarity versus difference stationarity. *Journal of Time Series Analysis*. 16, 147-164.

Berger, Roger L., and Hsu, Jason C. (1996). "Bioequivalence trials, intersection-union tests, and equivalence confidence sets." With discussion. *Statistical Science*, 11, 283-319.

Billingsley, P. (1991), "Probability and Measure", 2nd edition, New York, Wiley.

Billingsley, P. (1999), "Convergence of Probability Measures", 2nd edition, New York, Wiley.

Bollerslev, Tim. (1986), "Generalized Autoregressive Conditional Heteroscedasticity." *Journal of Econometrics*, 31, 307-27.

Campbell, J.Y. and Mankiw, N.G. (1987) Are output fluctuations transitory? *Q. J. Econ.* 102, 857-80.

Cavaliere, G. (2001), "Testing the unit root hypothesis using the generalized range statistics", *Econometrics Journal* 4, 70-88.

Chan, N.H. (1988), "The Parameter Inference for Nearly Nonstationary Time Series", *Journal of the American Statistical Association*, 83, 857-862.

Chan, N.H., and Wei, C. Z. (1987), "Asymptotic Inference for Nearly Nonstationary AR(1) Processes", *The Annals of Statistics*, 15, 1050-1063.

Chou, R.Y. (1988), "Volatility Persistence and Stock Valuations: Some empirical evidence using GARCH," *Journal of Applied Econometrics*, 3, 279-294.

Cochrane, J. H. (1988) How big is the random walk in GNP? *J. Poliyi. Econ.* 96, 893-920.

Cook, J.R, Stefanski, L.A. (1994), "Simulation-Extrapolation estimation in Parsimonious Measurement Error Models", *Journal of the American Statistical Association*, 89, 1314-1328.

DeJong, D. N., Nankervis, J.C. and Whiteman, C.H. (1992), "Integration Versus Trend-Stationary in Macroeconomic Time Series", *Econometrica*, 60, 423-434.

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

Dickey, D.A., and Fuller, W.A. (1981), Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica* 49, 1057-1072.

Diebold, F.X. and G.D. Rudebusch (1991), "On the power of Dickey-Fuller tests against fractional alternatives", *Economic Letters* 35, 155-160.

Elliott, G., Rothenberg, T.J., and Stock, J.H. (1996), Efficient tests for an autoregressive unit root. *Econometrica*, 64 :813-836.

Engle, R.F. (1982), "Autoregressive Conditional Heteroscedasticity with estimates of U.K. inflation," *Econometrica*, 50, 987-1007.

Fernandez, C., Steel, M. F. J. (1998), "On Bayesian modeling of fat tails and skewness", *Journal of the American Statistical Association* 93, 359-371.

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

Fuller, W.A. (1996), *Introduction to Statistical Time Series*, 2nd edition, Wiley Series.

Gikhman, I. I., and Skorokhod, A. V. (1996), "Introduction to the theory of random processes", New York, Dover Publications. English Translation of Original work in 1969.

Hall, Alastair (1989), " Testing for Unit root in the presence of moving average errors, " *Biometrika*, 76, 49-56.

Hamilton, James D., *Time Series Analysis*, 1994(Princeton University Press).

Harvey, A.C., Ruiz, E., and Shephard, N. (1994), "Multivariate Stochastic Variance Models." *Review of Economic Studies*, 61, 247-264.

Hasan, M. A., and Koenkar, R. W. (1997), "Robust Rank tests for the Unit Root Hypothesis", *Econometrica* 65(1), 133-161.

Hurst, H. (1951), "Long-term storage capacity of reservoirs.", Transactions of the American Society of Civil Engineers 116, 770–99.

King, M.L. (1987), Towards a Theory of Point Optimal Testing, Econometric Reviews, 6, 169-218.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992): Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root, Journal of Econometrics, 54, pp. 159-178, North-Holland.

Lehmann, E.L. (1994), Testing Statistical Hypotheses, (2nd Edition), Chapman and Hall: New York.

Lucas, A. (1995), "Unit Root tests based on m-estimators", Econometric Theory 11, 331-346.

Nelson, C.R. and C.I. Plosser, 1982, Trends versus random Walks in Macroeconomic Time Series: Some Evidence and Implications, Journal of Monetary Economics, 10, 139-162 .

Pindyck, R. S. (1984), "Risk, Inflation, and the Stock Market," American Economic Review, 74, 335-51.

Pantula, S.G. (1989), "Optimal Instrumental Variable Estimator of the AR parameter of an ARMA (1,1) process," Problem 89.1.2, Econometric Theory, 5, 173.

Pantula, S.G. (1991), "Asymptotic Distributions of Unit root tests when the process is nearly stationary," Journal of Business and Economic Statistics, 9, 63-71.

Park, H.J, and Fuller. W.A. (1993), "Alternative Estimators for Parameters of the Autoregressive Process, Unpublished manuscript, Iowa State University, Department of Statistics.

Perron, P. and S. Ng (1996), "Useful Modifications to some unit root tests with dependent errors and their local asymptotic properties," Review of Economic Studies, 63, 435-463.

Phillips, P. C. B. (1987), "Towards a unified asymptotic theory for autoregression", Biometrika, 74, 535-547.

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

Phillips, P.C.B. (1987), "Time Series Regression with a unit root, " Econometrica, 55, 277-301.

Poon, S.H., and Taylor, S.J. (1992), "Stock Returns and Volatility: An empirical study of UK Stock Market," *Journal of Banking and Finance*, 16, 37-59.

Poterba, J.M. and Summers, L.H. (1986), "The Persistence of Volatility and Stock Market Returns," *American Economic Review*, 76, 1142-1151.

Rothenberg, T. J. and Stock, J. H. (1997), "Inference in nearly integrated autoregressive model with nonnormal innovations", *Journal of Econometrics* 80, 269-286.

Said, S.E, and Dicekey, D.A. (1984), "Testing for unit roots in Autoregressive-Moving Average Models of unknown order, " *Biometrika*, 71, 599-608.

Said, S.E, and Dicekey, D.A. (1985), "Hypothesis testing in ARIMA(p,1,q) models, *Journal of American Statistical Association*, 80, 369-74.

Saikkonen, P. and Luukkonen, R. (1993), Testing for a moving average unit root in autoregressive inetegrated moving average models. *Journal of American Statistical Association*. 88, 596-601.

So, M.K.P, Lam,K. and Li, W.K. (1997), "An Empirical study of volatility in seven Southeast Asian Stock Markets using ARV Models," *Journal of Business Finance and Accounting*, 24, 261-275.

Tanaka, K. (1990), Testing for a moving average unit root. *Econometric Theory* 6, 433-444.

Taylor, S.J. (1986), "Modeling Financial Time Series," John Wiley.

Tsay, R.S. (1993), Testing for noninvertible models with applications. *Journal of Business and Economic Statistics*. 11, 225-233.

Wright, J.H. (1999), "Testing for a Unit root in the Volatility of Asset Returns, *Journal of Applied Econometrics*, 14, 309-318.