ABSTRACT

KIM, SEONG-TAE. A New Approach to Unit Root Tests in Univariate Time Series Robust to Structural Changes. (Under the direction of Dr. David A. Dickey.)

Using methodology in panel unit root tests we propose a new approach to univariate unit root tests. Our method leads to an asymptotically normal distribution of the least squares estimator and is robust to contaminated data having structural changes or outliers while the power of the test does not drastically worsen.

The main idea is that under the assumption that the process has a unit root we transform an AR(1) process $\{y_t : 1 \leq t \leq T\}$ to a double-index process $\{y_{ij} : 1 \leq i \leq m, 1 \leq j \leq n, mn = T\}$ in such a way that the segments are independent for $i = 1, 2, \cdots, m$. For this transformed data, we apply the same sequential limit as in Levin and Lin (1992, 2002). First, as $n \to \infty$ we obtain asymptotic results for each i. These have the same form as in conventional univariate unit root tests. Second, as $m \to \infty$ we obtain an asymptotically normal distribution for the OLS estimator by the Lindeberg-Feller CLT. An advantage of this technique is that an undetected break has a relatively minor effect which, in fact, disappears as m increases. We also show that for a general ARMA (p,q) model we still obtain the asymptotic normality of the unit root statistics under the sequential limit assumption.

A New Approach to Unit Root Tests in Univariate Time Series Robust to Structural Changes

by

Seong-Tae Kim

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

2006

APPROVED BY:

Professor David A. Dickey Chair of Advisory Committee	Professor Bibhuti B. Bhattacharyya
Professor Sastry G. Pantula	Professor Alastair R. Hall

To my dad,

Jeong-Cheol Kim,

who I regret will never be able to share my happiness at my great moment in person

but whose spirit is there at every step of my life.

Biography

Seong-Tae Kim was born in Hongdo, Jeollanamdo, the Republic of Korea on July 17, 1971. He obtained his B.S. in Economics from Hanyang University at Ansan in 1997, and his M.S. in Economics from Korea University in 1999. He entered the Department of Statistics, North Carolina State University for the Ph.D. degree in January, 2002 and obtained his M.S. in Statistics in December, 2003.

Acknowledgements

I would like to express my deepest gratitude and biggest appreciation to the people supported to complete this dissertation. First of all, my deepest appreciation goes out to my advisor Dr. David A. Dickey for his guidance, patience, and encouragement during my research. I also appreciate my committee members Dr. Sastry G. Pantula, Dr. Bibhuti B. Bhattacharyya and Dr. Alastair R. Hall. They not only provided me with valuable information and instruction through my research but also gave me a chance to upgrade my knowledge through their impressive classes.

I must acknowledge as well my friends, colleagues, and faculty members. I will never forget the moment of Dr. William Swallow visiting the ER at Rex Hospital when I passed out. It goes without saying that his sudden visiting deeply touched my heart. I wholeheartedly thank my friends, Man Sik Park and Jung Wook Park and their family. They have been the greatest supporter for my studies as well as they have shared my joy and sorrow.

Last, but not least, I would like to appreciate my mother, Sarye Park, and five brothers for always being at my side and supporting me. Without their support, I would not have been possible to complete my dissertation.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Statement of the Problem

Since Dickey and Fuller's (1979) seminal work, the unit root test has become standard in time series analysis, especially, in macroeconomic and financial data. Despite a number of studies on the unit root test, it has been agreed that the test has some critical problems such as low power of tests and vulnerability to structural changes. The low testing power problem has been improved by various approaches such as the weighted symmetric estimator (Pantula, Gonzalez-Farias and Fuller, 1994 and Park and Fuller, 1995) and the Dickey-Fuller GLS estimator (Elliot, Rothenberg and Stock, 1996). Since Perron (1989) addressed the unit root test with structural change models, researchers have proposed numerous approaches to this topic. In spite of voluminous work, however, this issue has led to quite different, controversal results partially due to difficulty of detecting a change point, which means that this topic still deserves attention.

It is well known that the limiting distributions of standard unit root tests are non-normal. These distributions are called "Dickey-Fuller distributions." Under the null hypothesis of a unit root, the limiting distributions are usually represented by functionals of standard Brownian motion. Although the asymptotic non-normality is not as serious as low testing power and sensitivity to model specification, it is inconvenient to use nonstandard tables as in Dickey (1976) or Fuller (1996). Hence, our research focuses on robustness to structural change and asymptotic normality. In this paper, we obtain the asymptotic normality in univariate unit root tests under local detrending using the methodology in panel unit root tests. Our approach is robust under the situation where there are structural changes or outliers while the power of the test does not drastically worsen.

The estimation models usually considered in standard unit root tests are:

$$\text{Model 1:}\quad Y_t = \rho Y_{t-1} + \epsilon_t \tag{1.1}$$

$$\text{Model 2:}\quad Y_t = \mu + \rho Y_{t-1} + \epsilon_t \tag{1.2}$$

$$\text{Model 3:}\quad Y_t = \alpha + \beta t + \rho Y_{t-1} + \epsilon_t \tag{1.3}$$

The corresponding estimators and their asymptotics are:

$$\text{Model 1:}\qquad \hat{\rho} = \frac{\sum_{t=2}^{n} Y_t Y_{t-1}}{\sum_{t=2}^{n} Y_{t-1}^2}$$

$$n(\hat{\rho}-1) \;\Rightarrow\; \frac{\frac{1}{2}\{W^2(1)-1\}}{\int_0^1 W^2(r)dr} \quad \text{and} \quad \hat{t} \;\Rightarrow\; \frac{\frac{1}{2}\{W^2(1)-1\}}{\left(\int_0^1 W^2(r)dr\right)^{1/2}}$$

$$\text{Model 2:}\qquad \hat{\rho}_\mu = \frac{\sum_{t=2}^{n} (Y_t - \bar{y}_{(0)})(Y_{t-1} - \bar{y}_{(-1)})}{\sum_{t=2}^{n} (Y_{t-1} - \bar{y}_{(-1)})^2}$$

$$n(\hat{\rho}_\mu - 1) \;\Rightarrow\; \frac{\int_0^1 \Psi(r)dW}{\int \Psi^2 dr} \quad \text{and} \quad \hat{t}_\mu \;\Rightarrow\; \frac{\int_0^1 \Psi(r)dW}{\left(\int \Psi^2 dr\right)^{1/2}}$$

where $\Psi = W(r) - \int W(r)dr$.

$$\text{Model 3:} \qquad \hat{\rho}_\tau = \frac{\sum_{t=2}^{T} \hat{y}_t \hat{y}_{t-1}}{\sum_{t=2}^{T} \hat{y}_{t-1}^2}, \qquad \hat{y}_t = y_t - \hat{\alpha} - \hat{\beta} t$$

$$n(\hat{\rho}_\tau - 1) \quad \Rightarrow \quad \frac{\int W(r)dW + N}{D} \quad \text{and} \quad \hat{t}_\tau \quad \Rightarrow \quad \frac{\int W(r)dW + N}{D^{1/2}}$$

where

$$N = 12\left[\int rW(r)dr - \frac{1}{2}\int W(r)dr\right]\left[\int W(r)dr - \frac{1}{2}W(1)\right] - W(1)\int W(r)dr$$

$$D = \int W^2(r)dr - 12\left(\int rW(r)dr\right)^2 + 12\int W(r)dr \int rW(r)dr - 4\left(\int W(r)dr\right)^2.$$

Researchers have studied structural change in the variance of the error process and structural changes in the parameters, μ, α, and β in Model 2 and Model 3. In the unit root test, the structural change leads to a different estimator of ρ and different asymptotic behavior where the estimator and limiting distribution depend on the break point and/or break size.

The main idea of the paper is that under the assumption that the process has a unit root we transform a univariate time series process $\{Y_t : 1 \leq t \leq T\}$ to a double-index process $\{y_{ij} : 1 \leq i \leq m, 1 \leq j \leq n, mn = T\}$ in such a way that the segments are independent for $i = 1, 2, \cdots, m$. For this transformed data, we apply the same sequential limit as in Levin and Lin (1992, 1993). First, as $n \to \infty$ we obtain asymptotic results for each i. These have the same form as in conventional univariate unit root tests. Second, as $m \to \infty$ we obtain an asymptotically normal distribution for the OLS estimator by the Lindeberg-Feller CLT. An advantage of this

technique is that an undetected break has a relatively minor effect on the asymptotic results which, in fact, disappears as m increases.

1.2 Literature Review

Literature review concentrates on three topics. They are unit root tests in panel data, univariate unit root tests with asymptotic normality, and unit root tests with structural changes since we adopt the methodology of panel unit root tests to achieve asymptotic normality, especially, under the situation of structural change. See Maddala and Kim (1998) and references therein for various issues in unit root tests.

1.2.1 Asymptotic normality in unit root tests

Lai and Siegmund (1983)

Lai and Siegmund (1983) showed that for a simple non-explosive AR (1) model if data follow a particular stopping rule, the OLS estimator of ρ is asymptotically normally distributed uniformly in ρ. Consider the simple non-explosive AR(1) model with unknown parameter $\rho \in [-1, 1]$

$$y_t = \rho y_{t-1} \ + \ \epsilon_t, \qquad t = 1, 2, \cdots$$

where ϵ_t are $i.i.d.$ with $\mathrm{E}\epsilon_1 = 0$ and $\mathrm{E}\epsilon_1^2 = \sigma^2$. For $c > 0$ let the optimal stopping rule be

$$T_c = \inf\{t : \sum_{i=1}^{t} y_{i-1}^2 \geq c\}. \tag{1.4}$$

Then uniformly in ρ and $t \in (-\infty, \infty)$,

$$P_\rho\left\{c^{-1/2}\sum_{i=1}^{T_c} y_{i-1}\epsilon_i \le t\right\} \to \Phi(t) \qquad \text{as} \;\; c \to \infty. \tag{1.5}$$

Now define

$$N_c = \text{first } n \ge 1 \text{ such that } I_n \ge c\sigma^2. \tag{1.6}$$

Then for the OLS estimator of ρ,

$$I_{N_c}^{1/2}(\rho_c - \rho) \xrightarrow{d} N(0, \sigma^2) \text{ uniformly in } \rho \in [-1, 1] \tag{1.7}$$

where $I_n = -\frac{d^2}{d\rho^2}\left(\rho\sum_{i=1}^n y_{i-1}y_i - \frac{1}{2}\rho^2\sum_{i=1}^n y_{i-1}^2\right) = \sum_{i=1}^n y_{i-1}^2$, which is the observed Fisher information about ρ.

Although this approach has the nice asymptotic normality, it has two main drawbacks: (1) it is hard to extend to a complicated model,i.e., linear trend model. Recently, Chang and Park (2004) extended the results to the local-to-unity hypothesis and AR(1) model with intercept; (2) According to the optimal stopping rule in their paper, the choice of c is totally arbitrary. This means that the value of c must be chosen a priori regardless of the pattern of a data set.

Breitung and Hassler (2002)

Breitung and Hassler (2002) proposed an asymptotically normally distributed $t-$statistic for $H_0 : z_t \sim I(1)$ against $H_1 : z_t$ is fractionally integrated. Letting $y_t = \Delta z_t = z_t - z_{t-1}$, the estimation regression is given by

$$y_t = \rho y_{t-1}^* + \epsilon_t,$$

where $y_{t-1}^* = \sum_{j=1}^{t-1}\frac{y_{t-j}}{j} = \sum_{j=1}^{t-1}\frac{\Delta z_{t-j}}{j}$ and ϵ_t are $i.i.d.$ with $\mathrm{E}(\epsilon_t) = 0$ and $\mathrm{E}(\epsilon_t^2) = \sigma^2$. Then the null hypothesis reduces to $H_0 : \rho = 0$. Finally, they obtained the asymptotic

results as $T \to \infty$

$$t_\rho \xrightarrow{d} N(0,\ 1)$$

$$T \cdot R^2 \xrightarrow{d} \chi^2(1)$$

where R^2 is the coefficient of determination. Although this approach achieves the asymptotic normality, the alternative is different from our framework.

Phillips, Moon, Xiao (2001)

Phillips *et al.* suggested a method of blocking a univariate time series to estimate the local to unity parameter. They first consider the following AR(1) model with the local to unity parameter:

$$y_t = ay_{t-1} + u_t, \quad a = 1 + \frac{c}{n}, \quad u_t \ \sim \ iid(0, \sigma^2), \quad t = 1, 2, \cdots, n, \ c < 0.$$

The block local to unity model they proposed is defined as follows:

$$\begin{aligned} y_{k,t} &= ay_{k,t-1} + u_{k,t}, \quad t = \{1, \cdots, m\}; \ k = 1, \cdots, M, \\ y_{k,0} &= y_{k-1,m}, \\ a &= e^{c/m} \sim 1 + \tfrac{c}{m}. \end{aligned} \tag{1.8}$$

In this transformation, the total number of observations, n, is divided into M blocks of m observations. Another feature is that the initial value of each interval is the last value of the previous interval. Using this transformation, they found the limiting distributions of $\hat{a}$ depending on the local to unity parameter c.

$$(1)\ c < 0 : \ \sqrt{M}m(\hat{a} - a) \Rightarrow N(0, -2c)$$

$$(2)\ c = 0 : \ Mm(\hat{a} - a) \Rightarrow \Big(\int W^2(r)dr \Big)^{-1} \int W(r)dW.$$

In order to obtain the above asymptotic results, they used the sequential limit method in Phillips and Moon (1999). Although they achieved the asymptotic normality in

the case of $c < 0$, they still had the same limiting distribution as the Dickey-Fuller test for the unit root case. We reference this method since our method also uses the blocking method in (1.8).

Kunst (1997)

The circular χ^2 family in Kunst (1997) is another double-index transformation of a univariate time series. Kunst focused on the univariate process $\{Y_t\}$ expressed as

$$(Y_1,Y_2,\cdots,Y_n,Y_{n+1},\cdots) = (X_{1,1},X_{2,1},\cdots,X_{n,1},X_{1,2},\cdots)$$

or equivalently, with $[x]$ for the integer part of x,

$$Y_t = X_{t-n[(t-1)/n],[(t-1)/n]+1}, \quad t = 1,\cdots,T;$$

where n is an arbitrary finite integer, and $(X_{1,t},\cdots,X_{n,t})$ is a finite sequence of n mutually independent random walks. This kind of transformation is quite similar to the seasonality model although the value of n is not restricted to the specific seasonal values such as $n = 2, 4, 12$. However, since n does not grow the asymptotic results are affected only by T. The circular χ^2 distribution with periodicity n is given by:

$$\beta_n(T) = \hat{\sigma}^{-2}\mathbf{h}_n(T)'\mathbf{H}_n(T)^{-1}\mathbf{h}_n(T)$$

where

$$\mathbf{H}_n(T) = (H_{ij}(T))_{i,j=1,\cdots,n} = \left(\left(\sum_{t=\max(i,j)+1}^{T} Y_{t-i}Y_{t-j}\right)/T^2\right)$$

and with $\Delta_n Y_t = Y_t - Y_{t-n}$,

$$\mathbf{h}_n(T) = \left(T^{-1}\sum_{t=2}^{T} Y_{t-1}\Delta_n Y_t,\cdots,T^{-1}\sum_{t=n+1}^{T} Y_{t-n}\Delta_n Y_t\right)'.$$

To obtain the asymptotic result for $\beta_n(T)$, he proved the following results:

$$T^{-2}\sum_{t=j+1}^{T} Y_{t-i}Y_{t-j} \Rightarrow \sigma^2 n^{-1}\sum_{l=1}^{n}\int_0^1 W_l(r)W_{l+j-i-n[(l+j-i-1)/n]}(r)dr,$$

$$T^{-2}\sum_{t=1}^{T} Y_t^2 \Rightarrow \sigma^2 n^{-1}\sum_{l=1}^{n}\int_0^1 W_l^2(r)dr,$$

and

$$T^{-1}\sum_{t=1}^{T} Y_{t-i}\Delta_n Y_t \Rightarrow \sigma^2 n^{-1}\sum_{l=1}^{n}\int_0^1 W_l dW_{l+i-[l+i-1]/n}(r).$$

As $T \to \infty$, $\mathbf{H}_n(T) \Rightarrow \mathbf{H}_n$ and $\mathbf{h}_n(T) \Rightarrow \mathbf{h}_n$ in which $\mathbf{h}_n$ is identically distributed as a transform of χ_n^2, but elements of $\mathbf{h}_n$ are not independent, and $\mathbf{H}_n$ is a symmetric circular Toeplitz matrix, which is defined as $\mathbf{V}_n\mathbf{R}_n^{-1}\mathbf{V}_n$ where the matrix of eigenvectors for $\mathbf{H}_n$ $\mathbf{V}_n = (v_1, \cdots, v_n)$ is symmetric, $\mathbf{V}_n^2 = I_n$ and the matrix of eigenvalues is $\mathbf{R}_n = \text{diag}\{r_1, \cdots, r_n\}$. Refer to Kunst (1997) for the values of v_i and r_i for $i = 1, \cdots, n$.

1.2.2 Panel Unit Root Tests

Through the paper we count mainly on the methods of panel data unit root tests, which have been developed relatively recently. Hence, we overview the development of unit root tests in panel data. Since the seminal works of Levin and Lin (1992, 1993) and Quah (1994), the study of unit root tests has played an increasingly important role in empirical analysis of panel data. Combining the cross-sectional dimension with the time dimension is very important in the context of nonstationary series because this provides a solution to the low power problem of the conventional unit root tests.

Levin and Lin (1992)

Levin and Lin (1992) construct an exhaustive study in unit root tests for the panel data model. They introduce a general form of dynamic panel data model

$$\Delta y_{it} = \rho y_{i,t-1} + \alpha + \delta t + \alpha_i + \theta_t + \epsilon_{it}$$

where $i = 1, 2, \cdots, N; t = 1, 2, \cdots, T$. They assumed that ϵ_{it} are $i.i.d.$ with $\mathrm{E}(\epsilon_{it}) = 0$ and $\mathrm{E}(\epsilon_{it}^2) = \sigma^2$ for all i and t. In this model they consider aggregate effects as well as individual-specific effects as follows:

$$\text{Model 1: } \Delta y_{it} = \rho y_{i,t-1} + \epsilon_{it}$$

$$\text{Model 2: } \Delta y_{it} = \rho y_{i,t-1} + \mu + \epsilon_{it}$$

$$\text{Model 3: } \Delta y_{it} = \rho y_{i,t-1} + \mu + \delta t + \epsilon_{it}$$

$$\text{Model 4: } \Delta y_{it} = \rho y_{i,t-1} + \theta_t + \epsilon_{it}$$

$$\text{Model 5: } \Delta y_{it} = \rho y_{i,t-1} + \mu_i + \epsilon_{it}$$

$$\text{Model 6: } \Delta y_{it} = \rho y_{i,t-1} + \mu_i + \delta_i t + \epsilon_{it}$$

Model 2 and 3 are aggregate effect models and Model 5 and 6 are individual-specific effect models. As $T \to \infty$ and $N \to \infty$ sequentially, for model 1-4, the Pooled OLS (POLS) estimator of ρ has the following asymptotic results:

$$T\sqrt{N}(\hat{\rho} - 1) \xrightarrow{d} N(0, 2) \quad \text{and} \quad \tau \xrightarrow{d} N(0, 1).$$

However, Model 5 and 6 have some bias in the POLS estimator due to individual-specific effects (or incidental parameters) so they have different limiting distributions:

Model 5:

$$T\sqrt{N}(\hat{\rho} - 1) + 3\sqrt{N} \xrightarrow{d} N(0, 10.2) \quad \text{and} \quad \sqrt{1.25}\tau + \sqrt{1.875N} \xrightarrow{d} N(0, 1)$$

Model 6:

$$T\sqrt{N}\{(\hat{\rho}-1)+7.5\} \xrightarrow{d} N\left(0, \frac{2895}{112}\right) \quad \text{and} \quad \sqrt{\frac{448}{277}}\{\tau+\sqrt{3.75N}\} \xrightarrow{d} N(0,1)$$

These asymptotic results depend sequentially on the functional CLT and Lindeberg-Feller CLT. First, as the number of observations in an individual (T) goes to infinity, we obtain well-defined random variables as functionals of standard Brownian motion via the functional central limit theorem. Those random variables are independent of each other. Second, as the number of individuals (N) goes to infinity, we obtain the asymptotically normal distribution via the Lindeberg-Feller central limit theorem.

Levin, Lin (1993) and Levin, Lin and Chu (2002)

These papers extend Levin and Lin (1992) to the more general case where the model allows the error process to have correlated and heteroscedastic structures while the independence across individuals still holds. The testing model is given by

$$\Delta y_{it} = \rho y_{i,t-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{it-L} + \alpha_{mi} d_{mt} + \epsilon_{it}$$

where d_{mt} for $m = 1, 2, 3$ indicate the deterministic parts, i.e., $d_{1t} = \varnothing$, $d_{2t} = \{1\}$ and $d_{3t} = \{1, t\}$ and α_{mi} are corresponding coefficients. The estimation process is as follows:

Step 1: After determining the lag order p_i using Hall (1994), perform ADF regressions and generate orthogonalized residuals for $L = 1, 2, \cdots, p_i$:

$$\text{Regress} \Delta y_{it} \text{ on } \Delta y_{i,t-L} \text{ and } d_{mt} \Rightarrow \text{residuals } \widehat{e}_{it}$$

$$\text{Regress} y_{i,t-1} \text{ on } \Delta y_{i,t-L} \text{ and } d_{mt} \Rightarrow \text{residuals } \widehat{v}_{i,t-1}$$

Step 2: Calculate the ratio of long-run to short-run standard deviations.

$$\widetilde{e}_{it} = \widehat{e}_{it}/\widehat{\sigma}_{\epsilon i} \text{ and } \widetilde{v}_{i,t-1} = \widehat{v}_{it-1}/\widehat{\sigma}_{\epsilon i}$$

where $\widehat{\sigma}_{\epsilon i}$ is the standard error from each ADF regression, for $i = 1, 2, \cdots, N$. Under the null hypothesis of a unit root, the long-run variance for the testing model is given by

$$\widehat{\sigma}^2_{yi} = \frac{1}{T-1}\sum_{t=2}^{T}\Delta y^2_{it} + 2\sum_{L=1}^{\bar{K}} w_{\bar{K}L}\left[\frac{1}{T-1}\sum_{t=2+L}^{T}\Delta y_{it}\Delta y_{i,t-L}\right]$$

where $\bar{K}$ is a truncation lag that can be chosen in a manner that ensures the consistency of $\widehat{\sigma}^2_{yi}$ and $w_{\bar{K}L}$ is the sample covariance weights, which depends on the choice of kernel, *e.g.* the Bartlett kernel is used, $w_{\bar{K}L} = 1-(L/(\bar{K}+1))$. For each individual i, the ratio of standard deviations is estimated by $\widehat{s}_i = \widehat{\sigma}_{yi}/\widehat{\sigma}_{\epsilon i}$ and their average is $\widehat{s}_N = N^{-1}\sum_{i=1}^{N}\widehat{s}_i$.

Step 3: Compute the panel test statistics. Running the following pooled regression

$$\widetilde{e}_{it} = \rho\widetilde{v}_{i,t-1} + \widetilde{\epsilon}_{it},$$

obtain the conventional t-statistic for $H_0 : \rho = 0$. This is $t = \frac{\hat{\rho}}{\widehat{\sigma}(\hat{\rho})}$ where

$$\hat{\rho} = \frac{\sum_{i=1}^{N}\sum_{t=2+p_i}^{T}\tilde{v}_{i,t-1}\tilde{e}_{it}}{\sum_{i=1}^{N}\sum_{t=2+p_i}^{T}\tilde{v}^2_{i,t-1}},$$

$$\widehat{\sigma}(\hat{\rho}) = \frac{\hat{\sigma}_{\tilde{\epsilon}}}{\left[\sum_{i=1}^{N}\sum_{t=2+p_i}^{T}\tilde{v}^2_{i,t-1}\right]^{1/2}}$$

and

$$\hat{\sigma}^2_{\tilde{\epsilon}} = \frac{1}{N\tilde{T}}\sum_{i=1}^{N}\sum_{t=2+p_i}^{T}(\tilde{e}_{it} - \hat{\rho}\tilde{v}_{i,t-1})^2$$

is the estimated variance of $\tilde{e}_{it}$. Finally, the adjusted $t-$ statistic is

$$t^* = \frac{t - N\widetilde{T}\hat{S}_N\hat{\sigma}_{\tilde{\epsilon}}^{-2}\hat{\sigma}\hat{\rho}\mu^*_{m\tilde{T}}}{\sigma^*_{m\tilde{T}}} \xrightarrow{d} N(0,1) \text{ as} N, T \to \infty$$

where $\mu^*_{m\tilde{T}}$ and $\sigma^*_{m\tilde{T}}$ are mean and standard deviation adjustments provided by Levin, Lin and Chu (2002).

Levin and Lin (and Chu) papers had a huge influence on subsequent work in panel unit root tests, but framework has a restriction that ρ is homogeneous across cross-section units. This restriction is not an obstacle in our modeling because our data generating process has only one value of ρ.

Im, Pesaran, and Shin (2003)

Im, Pesaran and Shin (2003) relax the restriction of homogeneous ρ, and then they consider heterogeneous ρ across i. The estimation model is given by

$$\Delta y_{it} = \alpha_i + \rho_i y_{i,t-1} + \xi_{it} \qquad i = 1, 2, \cdots, N;\ t = 1, 2, \cdots, T,$$

where the errors ξ_{it} are serially correlated with different correlation properties across units.

$$H_0 : \rho_i = 0 \text{ for all } i$$
$$H_1 : \begin{cases} \rho_i < 0 & \text{for } i = 1, \cdots, N_1 \\ \rho_i = 0 & \text{for } i = N_1 + 1, \cdots, N \end{cases}$$

The IPS $t-$statistic is defined as the average of the individual ADF statistics:

$$\bar{t} = \frac{1}{N}\sum_{i=1}^{N} t_{\rho_i}$$

where t_{ρ_i} is the individual $t-$statistic for testing $H_0 : \rho_i = 0$ for all i. That is,

$$t_{\rho_i} \Rightarrow \frac{\int_0^1 W_i(r)dW_i(r)}{\left(\int_0^1 W_i^2(r)dr\right)^{1/2}} = t_{iT} \text{ as } T \to \infty.$$

As a key point, Im, Pesaran and Shin showed that t_{iT} are *i.i.d.* with finite mean and variance with Nabeya's (1999) results. Then, by the Lindeberg-Levy CLT,

$$\frac{\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} t_{iT} - \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}[t_{iT}|\rho_i = 0]\right)}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} \mathrm{Var}[t_{iT}|\rho_i = 0]}} \xrightarrow{d} N(0,1) \quad \text{as } N \to \infty.$$

They use the simulated values for $\mathrm{E}[t_{iT}|\rho_i = 0]$ and $\mathrm{Var}[t_{iT}|\rho_i = 0]$ which depend on T and lag length used in the ADF test.

Many other researchers have addressed the panel unit root test under the assumption of independence across individuals. For example, Quah (1994) used the diagonal path limit to prove the asymptotic normality of the coefficient estimate of the unit root, Breitung and Meyer (1994) considered the subtraction of initial value to remove the effect of the individual intercept in the panel unit root, , Harris and Tzavalis (1998) provided the asymptotic distribution for the finite time case, Maddala and Wu (1999) used Fisher's P-value statistic, and Phillips and Moon (1999, 2000) provided the various theoretical properties used in the panel unit root. As the second generation in panel unit root tests, the cross-section defendency issues follow.

1.2.3 Structural change

Using the Augmented Dickey-Fuller (ADF) test Nelson and Plosser (1982) analyzed whether or not several U.S. macroeconomic series are *difference-stationary*, i.e., the data have a unit root. They could not reject the null hypothesis of unit root against the *trend-stationary* alternative for 13 out of 14 annual data series. After Nelson and Plosser's findings of a unit root in U.S. macro aggregates, a lot of studies have been subsequently performed to detect nonstationarity or stationarity of those

time series data.

However, Perron (1989) argued that macroeconomic time series data might be found to be stationary if a structural change were modeled in the deterministic part where the structural change is a known single change in intercept and/or slope. With this framework of structural change, he rejected the null hypothesis of a unit root against the trend-stationary alternative for 11 out of 14 series used in Nelson and Plosser (1982). After Perron (1989), there were several approaches in which the breaking point is estimated rather than fixed. There are two types of methods endogenizing the detection of a breaking point into a test procedure, i.e., recursive methods and sequential methods. The former uses subsamples such as $t = 1, \cdots, t_b$ for $t_b = t_0, \cdots, T$ where t_0 is a starting point for testing and T is the sample size, and the latter sequentially changes the date of the hypothetical break point using the full sample with some dummies. Zivot and Andrews (1992) and Banerjee, Lumsdaine and Stock (1992) reverse the conclusions of Perron (1989) by adopting endogenous structural change.

A natural extension of the literature on unit root tests in the presence of structural change involves allowing for multiple break points instead of a single point. To this end, Lumsdaine and Papell (1997) allowed for one and two possible endogenous break points in the unit root test adopting Zivot and Andrews' (1992) method, and they obtained results between Perron (1989) and Zivot and Andrews (1992). There are some research results supporting trend stationarity with multiple break points such as Ben-David, Lumsdaine and Papell (2003), Garcia and Perron (1996), Kanas (1998), Papell (1998), and among others.

The research above mentioned is interesting in the case of the null hypothesis of a unit root versus the alternative of a stationary process with break points. However, Leybourne, Mills and Newbold (1998) considered the DF tests in the presence of a change point under the unit root null hypothesis. Their results show that if a break occurs early in the sample of the unit root process, the DF test has severe size distortion which leads to the erroneous decision that an $I(1)$ process is $I(0)$. In this paper, we also consider the case of the presence of structural changes under the unit root null.

1.3 Summary of chapters

Chapter 2

In chapter 2 we provide theoretical background used through the paper. These results include (1) the issues of double-index and independent transformation, (2) the issues of growth rate of m and n, and (3) the issues of joint versus sequential limits in asymptotics. In the transformation discussion we show that our transformation leads to a different asymptotic behavior from the standard unit root test. In the growth rate issue, we address how to select m and n in the process of transformation and study the convergence rate of the ratio of m to n. Finally, as we study the limit issue we can easily show the asymptotic behavior of the estimators in chapter 3.

Chapter 3

In chapter 3 we focus on estimation and inference on unit root tests. For these purposes we consider three commonly used models and their transformed counterparts. The estimation part mainly relies on methodology in panel unit root tests and hence handling bias adjustment due to interval- specific effects is a main issue. As well, we address the asymptotic behavior of estimators which is totally different from those of the standard unit root test in univariate time series data. We derive a limiting normal distribution instead of functionals of standard Brownian motion. We calculate the critical values for test statistics and estimate testing power by simulation studies. We compare those with the results of the standard DF unit root tests.

Chapter 4

This chapter is to extend the results derived in chapter 3 to contaminated time series data where the contaminated data could include structural change, outlier, and missing data. As a main result, we show that if the number of contaminated intervals is finite (usually very small numbers), then this condition does not affect the asymptotic results in our modified unit root tests. We concentrate on the structural change models considering level shift, trend shift, and slope change. We also provide simulation results for this topic.

Chapter 5

In this chapter we apply our method to various real data such as US real GDP (small sample example) and US/Canada exchange rate (large sample) . For the US GDP data, we compare the result to three benchmarking tests, ADF, Zivot and Andrew, and Perron's test in Murray and Nelson (2000). When performing empirical analyses, various practical issues would occur, *e.g,* the choice of m and n, the outlier interval problem, the lag selection issue, and the choice of deterministic parameters (global vs. interval-specific).

Chapter 6

Finally, we summerize our findings through the paper and suggest future work. The findings include data transformation, asymptotic results, finite sample properties, and performance under structural changes. Future studies focus on the extension and improvement of our methodology.

Chapter 2

Data Transformation and Asymptotics

In this chapter we address some theoretical background that is used through the paper. First, we introduce a two-step data transformation, double-indexing and last value subtraction. This transformation leads to different asymptotic behavior from the standard unit root test. Second, we study the issues of growth rate of the index limits m and n where the total sample size $T = mn$. In this section, we study how to select m and n in the process of transformation and study the convergence rate of the ratio of m to n. Finally, we review the concepts of joint, sequential, and diagonal limit in asymptotics. As we study the limit issue we can easily show the asymptotic behavior of the estimators in Chapter 3 and 4.

2.1 Multi-index transformation of random variables

Following Fuller (1996, p.3), we describe what a time series is. Let $(\Omega, \mathscr{A}, P)$ be a probability space, and let T be an index set with $T \subseteq \mathbb{Z}$. A real valued *time series* (or *stochastic process*) is a real valued function $Y(t, \omega)$ defined on $T \times \Omega$ such that for each fixed t, $Y(t, \omega)$ is a random variable on $(\Omega, \mathscr{A}, P)$. The function $Y(t, \omega)$ is often written $Y_t(\omega)$ or Y_t, and a time series is considered as a collection $\{Y_t : t \in T\}$ of random variables.

Now consider a double-index transformation of random variables. Let S be an index set in $\mathbb{Z} \times \mathbb{Z}$ such that $S = \{(i, j) : i \in M; j \in N; \text{ and } M, N \subseteq \mathbb{Z}\}$. If, for an arbitrary function f, $f : T \mapsto S$ is one-to-one and onto, then $\{Y_{i,j} : (i, j) \in S\}$ is probabilistically equivalent to $\{Y_t : t \in T\}$. In practice, we are interested in the observed time series with index set $T = \{t : \ t = 0, 1, 2, 3, \cdots\}$. Hence, we restrict attention to the index set S such that $S = \{(i, j) : \ i \in M \text{ and } j \in N\}$ where $M = \{i : \ i = 1, 2, 3, \cdots\}$ and $N = \{j : \ j = 0, 1, 2, 3, \cdots\}$.

For a time series with a unit root, the double-index transformation itself is not sufficient to obtain asymptotic results different from those of the standard unit root test. It is simply a re-indexing that provides the first step in the development of such a result. The solution is what we will call the *double array* (or *triangular* array) transformation.

Definition 2.1 *(Double array): A double array is a collection of random variables on* $\mathbb{R}$, $\{Y_{nj} : 1 \leq j \leq k_n, n \in \mathbb{N}\}$ *with* $k_n \to \infty$ *such that* $Y_{n1}, \cdots, Y_{nk_n}$ *are independent for each* $n \in \mathbb{N}$. *If* $k_n = n$, *it is called a triangular array.*

In this definition, the key point is the independence of Y for all n. For the unit root process, let $\{Y_t : \ 1 \leq t \leq T\}$ be a sequence of dependent random variables, and let $\{y_{i,j} : \ 1 \leq i \leq m; 1 \leq j \leq n\}$ be a sequence of random variables which are independent across i, but not over j. If we transform Y_t to $y_{i,j}$ in a certain way, then their probabilistic properties are different from each other. Therefore, our goal is to obtain a new time series $\{y_{ij}\}$ from the original series $\{Y_t\}$. This kind of double-index transformation has been addressed by Kunst (1997) and Phillips *et al.*(2001) although their asymptotic results are different from each other due to the different transformation methods.

Double-indexing of random variables

Consider a dependent time series $\{Y_t : \ 0 \leq t \leq T\}$ in which $Y_0 = 0$. Applying a one-to-one and onto double-index transformation $Y_t \mapsto Y_{ij}$, we obtain the double-indexed process $\{Y_{ij} : \ i = 1, \cdots, m; \ j = 1, \cdots, n; T = mn\}$ given by

$$\begin{array}{c} Y_{11}, \ \ Y_{12}, \ \cdots, \ \ Y_{1n}, \\ Y_{21}, \ \ Y_{22}, \ \cdots, \ \ Y_{2n}, \\ \vdots \\ Y_{m1}, \ \ Y_{m2}, \ \cdots, \ \ Y_{mn}, \end{array}$$

where Y_{ij} are obviously dependent across i since the original process $\{Y_t\}$ is a dependent process. Practically, $T = mn$ does not always hold as if T is a prime number. In such a case, we suboptimally choose m and n such that $mn = [T]_n$ where $[T]_n$ is the largest multiple of n that is less than T, and the number of observations in the last interval is $T - [T]_n$. If $T - [T]_n$ is greater than the number of model parameters,

then we also use that interval. Otherwise, we neglect that interval.

Independent transformation

Independence of transformed variables across i is possible under the situation where a univariate time series is a random walk. In this process of transformation, we can consider two cases. First, we can subtract the last value of the $(i-1)^{th}$ interval, $Y_{i-1,n}$ from each value in the i^{th} interval for $i = 2, \cdots, m$:

$$\begin{gathered}
0, \quad Y_{11}, \quad Y_{12}, \quad \cdots, \quad Y_{1n}, \\
Y_{1,n} - Y_{1,n}(=0), \; Y_{21} - Y_{1n}, \; Y_{22} - Y_{1n}, \; \cdots, \; Y_{2n} - Y_{1n}, \\
\vdots \\
Y_{m-1,n} - Y_{m-1,n}(=0), \; Y_{m1} - Y_{m-1,n}, \; Y_{m2} - Y_{m-1,n}, \; \cdots, \; Y_{mn} - Y_{m-1,n},
\end{gathered}$$

where for $i = 1$, we subtract 0 from each observation. Letting $y_{ij} = Y_{ij} - Y_{i-1,n}$, we equivalently have the transformed sequence

$$\begin{gathered}
0, \; y_{11}, \; y_{12}, \; \cdots, \; y_{1n}, \\
0, \; y_{21}, \; y_{22}, \; \cdots, \; y_{2n}, \\
\vdots \\
0, \; y_{m1}, \; y_{m2} \; \cdots, \; y_{mn}.
\end{gathered}$$

Second, we can subtract the initial value, $Y_{i,1}$ of each interval:

$$\begin{gathered}
Y_{11} - Y_{11}(=0), \; Y_{12} - Y_{11}, \; \cdots, \; Y_{1n} - Y_{11}, \\
Y_{21} - Y_{21}(=0), \; Y_{22} - Y_{21}, \; \cdots, \; Y_{2n} - Y_{21}, \\
\vdots \\
Y_{m1} - Y_{m1}(=0), \; Y_{m2} - Y_{m1}, \; \cdots, \; Y_{mn} - Y_{m1}.
\end{gathered}$$

Although both methods give independent rows, the latter loses more observations than the former in the process of transformation, and hence, we consider the former transformation hereafter. Phillips, Moon and Xiao (2001) has a similar transformation to our transformation, but their approach does not guarantee independence across blocks since they initialize for each interval with the last observation of the previous interval.

For illustrating the process of data transformation, consider a simple AR(1) model

$$Y_t \;=\; \rho Y_{t-1} \;+\; \epsilon_t, \qquad t = 1, 2, \cdots, T \tag{2.1}$$

where $\epsilon_t \sim iid(0,\ \sigma^2)$ and $Y_0 = 0$. Under the null hypothesis that $\rho = 1$, it becomes

$$Y_t \;=\; Y_{t-1} \;+\; \epsilon_t \;=\; \sum_{s=1}^{t} \epsilon_s, \qquad t = 1, 2, \cdots, T. \tag{2.2}$$

Under the double-index transformation, for $i = 1, \cdots, m; j = 1, \cdots, n$, we obtain

$$Y_{ij} \;=\; Y_{ij-1} \;+\; \epsilon_{ij} = Y_{i-1,n} + \sum_{s=1}^{j} \epsilon_{is}, \tag{2.3}$$

where $Y_0 = 0$ is not used. Finally under the double array transformation,

$$Y_{ij} \;-\; Y_{i-1,n} \;=\; Y_{ij-1} \;-\; Y_{i-1,n} \;+\; \epsilon_{ij} \tag{2.4}$$

where $Y_{i-1,n} = 0$ for $i = 1$. For simplicity, letting $y_{ii} = Y_{ij} - Y_{i-1,n}$, Equation (2.4) is equivalently

$$y_{ij} \;=\; y_{ij-1} \;+\; \epsilon_{ij} = \sum_{s=1}^{j} \epsilon_{is} \qquad \text{for all } i. \tag{2.5}$$

The last expression obviously indicates that y_{ij} are independent across i if ϵ_{ij} are $i.i.d.$ for all i and j.

Table 2.1: Independent Transformation

Last value transformation	Initial value transformation
$(0), \epsilon_{11},\ \epsilon_{11}+\epsilon_{12},\ \cdots,\ \sum_{j=1}^{n}\epsilon_{1j}$	$0,\ \epsilon_{12},\ \epsilon_{12}+\epsilon_{13},\ \cdots,\ \sum_{j=2}^{n}\epsilon_{1j}$
$(0), \epsilon_{21},\ \epsilon_{21}+\epsilon_{22},\ \cdots,\ \sum_{j=1}^{n}\epsilon_{2j}$	$0,\ \epsilon_{22},\ \epsilon_{22}+\epsilon_{23},\ \cdots,\ \sum_{j=2}^{n}\epsilon_{2j}$
$\vdots$	$\vdots$
$(0), \epsilon_{m1},\ \epsilon_{m1}+\epsilon_{m2},\ \cdots,\ \sum_{j=1}^{n}\epsilon_{mj}$	$0,\ \epsilon_{m2},\ \epsilon_{m2}+\epsilon_{m3},\ \cdots,\ \sum_{j=2}^{n}\epsilon_{mj}$

The transformed model is quite similar to the seasonal unit root model in Dickey, Hasza, and Fuller (1984) where the seasonality is determined by mod m as well as to m panel data model mentioned in Chapter 1. Their seasonal model can be partitioned into m component series such that

$$y_{mk+i} = \rho y_{m(k-1)+i} + \epsilon_{mk+i}; \quad i = 0, 1, \cdots, m-1,$$

and each component runs from $k = 1$ to n where the total number of observations is $T = mn$. This partition into m component series provides a useful link between the double array transformed and seasonal models. However, there is a main difference between two models. That is, the transformation model allows m to arbitrarily increase while the seasonal unit root model has some specific number for m, e.g., $m = 2, 4,$ and 12, which correspond to semi-annual, quarterly, and monthly data. This difference causes different asymptotic behavior for each model. Kunst(1997) extended Dickey *et al.* (1984) and Hylleberg *et al.* (1990) to the more general case in the sense that he considered an arbitrary n. The asymptotic result of Kunst (1997) follows the χ^2 distributions, which are different from Dickey *et al.*

2.2 Growth rate of m and n

When we transform a univariate time series with a unit root to a double-indexed independent process in the previous section, we divide the T observations into m sets of size n. For convenience, let m be the number of intervals which corresponds to the number of individuals in panel data, and let n be the number of observations in each interval which corresponds to the number of observations over time in panel data. In this section, we present some asymptotic behavior of the double-indexed independent process, which depends on m and n. First we note that asymptotic analysis is only useful to the extent that it approximates the finite sample result. None of the assumptions listed above is inherently more justified than another.

A linear double-indexd process for asymptotic results depending on m and n typically has the following form:

$$X_{m,n} = \frac{1}{k_m}\frac{1}{h_n}\sum_{i=1}^{m}\sum_{j=1}^{n} f(y_{i,j}), \tag{2.6}$$

where f is some measurable function on $(\Omega, \mathscr{A}, P)$, and $f(y_{i,j})$ is sometimes independent across i, $\{y_{i,j}\}$ is a time series process, and h_n and k_m are a function of n and m, respectively.

The limiting results of this process may be studied as m and n go simultaneously to infinity, or as $m \to \infty$ follows $n \to \infty$, or vice versa. For this idea we review the following limits as in Phillips and Moon (1999) that well studied the concepts in the double-indexed data framework.

(i) *Sequential Limit:* A sequential limit approach first makes one index, say m

fixed and let the other,say n, go to infinity. This step gives an intermediate limit which still depends on m. Then allowing m to go to infinity, we obtain the final sequential limit result. This sequential limit is often denoted by $(n, m \to \infty)_{\text{seq}}$. Although the reverse sequence is possible, we adopt this type of sequential limit since it gives the nicest results. For Equation (2.6), the sequential limit is:

$$X_{m,n} \overset{n\to\infty}{\longrightarrow} X_m \overset{m\to\infty}{\longrightarrow} X$$

where asymptotic results could be obtained either convergence in probability or in distribution.

(ii) *Diagonal Path Limit:* A diagonal path limit approach allows the two indexes, m and n, to go to infinity along a specific diagonal path in the two dimensional array. This path can be determined by a monotonically increasing functional relationship between m and n, e.g., $m = m(n) = n^{\frac{1}{2}}$. By this approach, $X_{m,n}$ reduces to the single indexed process $X_{m(n),n}$ which implies that the asymptotic theory would be simplified. Quah (1994) used this approach in finding the asymptotic results of panel unit root test statistics. Of course, the assumed path $(m(n), n \to \infty)$ may or may not provide an appropriate approximation for a given index set (m, n).

(iii) *Joint Limit:* A joint limit approach allows both indices, m and n, to go to infinity simultaneously without specific restrictions on the growth rate as in a diagonal path. However, this approach still requires some control over the relative expansion of the two indexes in order to obtain asymptotic results. For Equation (2.6), the joint limit is:

$$X_{mn} \overset{n,m\to\infty}{\longrightarrow} X.$$

Generally, the joint limit approach is more robust than either sequential or diagonal path approaches, but makes it difficult to derive the limiting results and requires stronger conditions such as existence of higher moments for achieving uniform convergence. Comparing to the joint limit, the sequential limit requires an additional assumption that the intermediate limit, say X_m, should be well-defined random variables or constants. In addition to these properties, it is not generally true that a sequential limit is equal to a joint limit. Phillips and Moon (1999) provided the conditions under which those two limits are asymptotically the same. Nevertheless, most panel unit root tests use the sequential limit approach to derive their asymptotic results sometimes without this being explicitly mentioned. We adopt the sequential limit approach as in most panel unit root tests since the seqential limit approach allows us to obtain the asymptotic results relatively easily. With the sequential limit, for the unit root statistics the Functional CLT is first applied as $n \to \infty$ and then the Lindeberg-Feller CLT is applied $m \to \infty$ to obtain the final asymptotic results. In chapter 3 we apply these results to the specific unit root models.

As $T \to \infty$, we can consider three cases for m and n:

- Case 1: $m \to \infty$ and $n \to \infty$
- Case 2: $m \to \infty$ and n is fixed
- Case 3: m is fixed and $n \to \infty$

In the previous section we defined the double array transformation which has the property of independence across i. Hence, if we obtain the first two moments for

each i, then we can apply the Lindeberg-Feller CLT for case 1 and 2. In other words, as long as $m \to \infty$, we achieve asymptotic normality via the Lindeberg-Feller CLT although in case 2 the asymptotic result depends on the number of observations in each interval, n as in Harris and Tzavalis (1999). As mentioned in the previous section, the seasonal unit root model is a special case of case 3 so this model does not achieve the asymptotic normality. Dickey, Hasza, and Fuller (1984) showed that for $m = 12$, the limiting distributions are nonnormal functionals of standard Brownian motions, and Kunst (1997) approach still has a χ^2 distribution for an arbitrary fixed m. Thus we do not pursue this case further.

If we can choose m and n in a certain way, then m and n are functions of the total sample size T, such as $m = m(T)$ and $n = n(T)$. For Equation (2.1) we have two different rates of convergence which depend on the value of ρ. That is, if $|\rho| < 1$ then we obtain a $\sqrt{T}$-consistent estimator of ρ, and if $\rho = 1$, then we obtain a T-consistent (superconsistent) estimator. However, for Equation (2.5) obtained by the independent transformation, as in the panel data model, we obtain a $\sqrt{mn}$-consistent estimator and an $n\sqrt{m}$-consistent estimator of ρ for the stationary and nonstationary processes, respectively.

Without loss of generality, we define the m and n as a function of T using $\delta \in [0,\ 1)$:

$$\begin{aligned} m &= m(T) = T^{\delta}; \\ n &= n(T) = T^{1-\delta}; \\ mn &= T. \end{aligned} \tag{2.7}$$

For the stationary process, the convergence rate of the least squares estimator of

ρ, $\sqrt{mn} = \sqrt{T}$, is free of δ. However, for the case of $\rho = 1$, the convergence rate is $n\sqrt{m} = T^{1-\delta/2}$ which is continuous between $(0.5, 1]$. If $\delta = 0$ then we have the conventional unit root case, and for $\delta \in (0,\ 1)$ we achieve the asymptotic normality with the rate of convergence, $n\sqrt{m}$. Of course in practice, m will be a rounded version of $m = T^{1-\delta}$.

Phillips, Moon and Xiao (2001) obtain a similar result to our convergence rate $T^{1-\delta/2}$ for the local-to-unity case, and they still obtain the conventional Dickey-Fuller distribution for the unit root case. Our last-value subtraction method leads to asymptotic normality, even in the unit root process regardless of whether or not the triangular array is independent across rows. The difference between the two approaches depends on the double-indexing method and the initial value setup. That is, we first subtract the last value of the previous interval, we use the value of zero as the initial value while they use the last value of the previous interval (or block). Hence, in the presence of a unit root our process restarts from 0 in every interval, but Phillips *et al.* process preserves the previous shock without change.

In Chapter 3 and 4 we address the unit root tests in Case I: $m \to \infty$ and $n \to \infty$, which is studied in most panel unit root tests. Although we allow both m and n to increase, there is another issue to obtain the asymptotic normality related to the growth rate of the ratio m to n.

The following summary provides the growth rate of the ratio in the panel unit root test for various papers that use different assumptions.

- Quah (1994): $\frac{\sqrt{m}}{n} \to 0$, $m = m(n)$
- Levin and Lin (1992): $\frac{\sqrt{m}}{n} \to 0$
- Levin and Lin (1993, 2002): $\frac{m}{n} \to 0$
- Im, Pesaran, Shin (2003): $\frac{m}{n} \to k, k$ constant
- Phillips and Moon (1999, 2000): $\frac{n}{m} \to \infty$

When we develop the asymptotics of test statistics in chapter 3, we mainly adopt the Levin and Lin (1992) approach. However, the growth rate is highly dependent on the model specification.

Chapter 3

Estimation

In this chapter we consider three typical models for unit root tests in a univariate time series. We obtain estimators and test statistics for each model based on our transformed data in Chapter 2 and study their asymptotic behavior. We consider three models which are usually studied in unit root tests as follows:

$$\text{Model 1: } Y_t \;=\; \rho Y_{t-1} \;+\; \epsilon_t \tag{3.1}$$

$$\text{Model 2: } Y_t - \alpha \;=\; \rho(Y_{t-1} - \alpha) \;+\; \epsilon_t, \qquad \text{for } t = 0, 1, \cdots, T \tag{3.2}$$

$$\text{Model 3: } Y_t - \alpha - \beta t \;=\; \rho(Y_{t-1} - \alpha - \beta(t-1)) \;+\; \epsilon_t. \tag{3.3}$$

Or equivalently,

$$\text{Model 1: } Y_t \;=\; \rho Y_{t-1} \;+\; \epsilon_t \tag{3.4}$$

$$\text{Model 2: } Y_t \;=\; \mu_1 \;+\rho Y_{t-1} \;+\; \epsilon_t, \qquad \text{for } t = 0, 1, \cdots, T \tag{3.5}$$

$$\text{Model 3: } Y_t \;=\; \mu_2 \;+\; \delta t \;+\rho Y_{t-1} \;+\; \epsilon_t, \tag{3.6}$$

where $y_0 = 0$, $\mu_1 = \alpha(1-\rho)$, $\mu_2 = \alpha(1-\rho) + \beta\rho$, and $\delta = \beta(1-\rho)$. We formally define the error process $\{\epsilon_t\}_{t=0}^{\infty}$ as in Fuller (1996, pp. 408-9).

Assumption 3.1 *$\{\epsilon_t\}$ is a sequence of random variables with $E(\epsilon_t|\mathscr{A}_{t-1}) = 0$, $E(\epsilon_t^2|\mathscr{A}_{t-1}) = \sigma^2$, and $E\{|\epsilon_t|^{2+\delta}|\mathscr{A}_{t-1}\} < M < \infty$ for some $\delta > 0$ and for all t, where $\mathscr{A}_{t-1}$ is the sigma-algebra generated by $\{\epsilon_j : j \leq t-1\}$.*

Based on three regression models, we consider four scenarios for regression in the context of panel data models. As is true with most unit root tests, we think of data being generated by a random walk, possibly with a constant or linear trend added. One then runs a regression of Y_t on Y_{t-1} and possibly 1 and t, thus making the Y_{t-1} coefficient invariant to the actual values of the constant or trend. The following classifications describe the regressions to be performed. We use the transformed data $y_{i,j}$ so these "models" are instructions on how to run the regressions rather than descriptions of data generating mechanism. For example, if $Y_t = \mu + \sum_{s=1}^{t} \epsilon_s$, then the transformation $y_{i,j}$ automatically annihilates μ and produces m independent random walks and yet we anticipate that a user might run the regression of $y_{i,j}$ on 1 and $y_{i,j-1}$ which we indicate as the regression model $y_{i,j} = \mu + \rho y_{i,j-1} + \epsilon_{i,j}$.

Class 1. Global Parameter Regression Models

$$M1: \quad y_{ij} = \rho y_{ij-1} + \epsilon_{ij} \tag{3.7}$$

$$M2 - G: y_{ij} = \mu + \rho y_{ij-1} + \epsilon_{ij} \tag{3.8}$$

$$M3 - G: y_{ij} = \alpha + \beta \cdot j + \rho y_{ij-1} + \epsilon_{ij} \tag{3.9}$$

Class 1-1. Demeaned Global Parameter Regression Models

$$M1: \qquad y_{ij} = \rho y_{ij-1} + \epsilon_{ij} \tag{3.10}$$

$$M2 - GD: y_{ij} = \rho y_{ij-1} + \epsilon_{ij} \tag{3.11}$$

$$M3 - GD: y_{ij} = \beta \cdot j + \rho y_{ij-1} + \epsilon_{ij} \tag{3.12}$$

Class 2. Interval-Specific Parameter Regression Models

$$M1: \qquad y_{ij} = \rho y_{ij-1} + \epsilon_{ij} \tag{3.13}$$

$$M2 - I: y_{ij} = \mu_i + \rho y_{ij-1} + \epsilon_{ij} \tag{3.14}$$

$$M3 - I: y_{ij} = \alpha_i + \beta_i \cdot j + \rho y_{ij-1} + \epsilon_{ij} \tag{3.15}$$

Class 2-1. Demeaned Interval-Specific Parameter Regression Models

$$M1: \qquad y_{ij} = \rho y_{ij-1} + \epsilon_{ij} \tag{3.16}$$

$$M2 - ID: y_{ij} = \rho y_{ij-1} + \epsilon_{ij} \tag{3.17}$$

$$M3 - ID: y_{ij} = \beta_i \cdot j + \rho y_{ij-1} + \epsilon_{ij} \tag{3.18}$$

Assumption 3.2 *The doubly indexed sequence of random variables* $\{\epsilon_{ij}\}$ *has the same properties as* $\{\epsilon_t\}$ *in Assumption* 3.1. *Thus,* $E(\epsilon_{ij}) = 0$ *and* $E(\epsilon_{ij}^2) = \sigma^2$ *for all* i *and* j.

In the following sections, for the above regression models, we study coefficient estimates and their asymptotic behavior. Although under the null hypothesis that $\rho = 1$, the asymptotic result for each model is the same as that of the corresponding model in Levin and Lin (1992), it is meaningful work to present the results here since

(1) we introduce the sequential limit assumption that makes the process of obtaining asymptotic results clear; (2) we correct some theorems and proofs; and (3) we provide proofs in detail.

3.1 Model I: Simple AR (1) model

3.1.1 Model setup

The first model is a simple AR (1) model without intercept or time trend. Although this model is empirically unrealistic, it provides us with insight to understand our methodology.

Original model: $Y_t = \rho Y_{t-1} + \epsilon_t,$

Double-index transformed model:

$$Y_{ij} = \rho Y_{i,j-1} + \epsilon_{ij} \tag{3.19}$$

Independent transformed model under the null hypothesis $\rho = 1$:

$$Y_{ij} - Y_{i-1,n} = \rho(Y_{i,j-1} - Y_{i-1,n}) + \epsilon_{ij} \tag{3.20}$$

$$\Leftrightarrow \quad y_{ij} = \rho y_{ij-1} + \epsilon_{ij}, \tag{3.21}$$

where

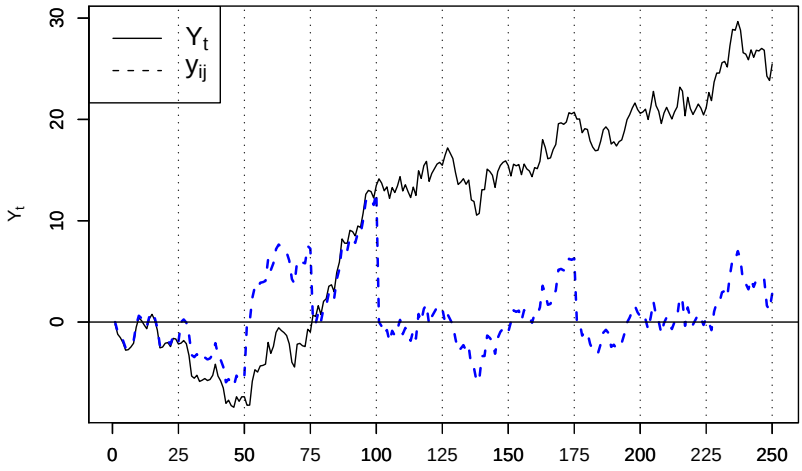
$$y_{ij} = \begin{cases} Y_{1j} & \text{if} \quad i = 1, \\ Y_{ij} - Y_{i-1,n} & \text{else if } i = 2, 3, \cdots, m. \end{cases} \tag{3.22}$$

We also assume that $Y_0 = 0$ and $y_{i0} = 0$ for all i in the sense that $y_{i0} = Y_{i-1,n} - Y_{i-1,n}$. Equation 3.21 is exactly the same as the simple AR(1) dynamic panel model independent across i, which is found in the standard panel data textbooks, *e.g.* Baltagi (2005) and Hsiao (2002). Note that the last observation, $Y_{i-1,n}$, from the previous interval is subtracted from each observation in interval i in the independent transformed model. Also note that this is not algebraically equivalent to the original model except under the null hypothesis $H_0 : \rho = 1$. Under that hypothesis, observations in any given interval are independent of those in any other interval. Table 3.1 further illustrates the transformation process.

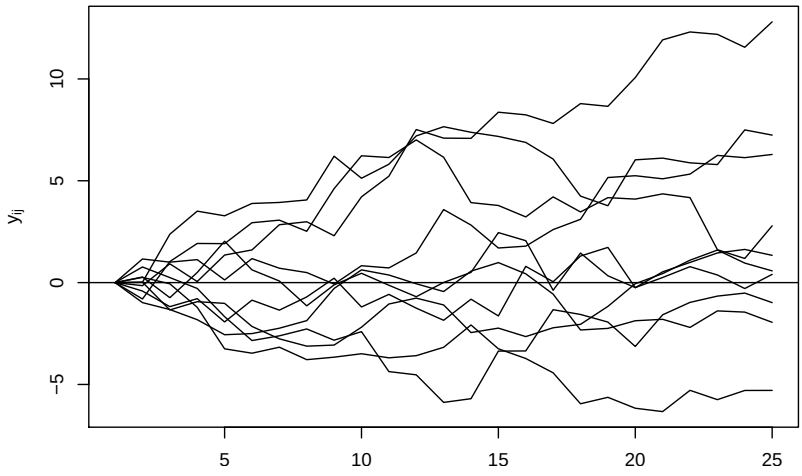
In Figure 3.1, we illustrate the process of data transformation for a simple AR (1) model under the null hypothesis that $\rho = 1$. In the upper plot, the solid line denotes the original series with $T = 250$, and the dotted line is the independent transformation with $m = 10$ and $n = 25$ where the error process $\{\epsilon\}$ is generated from $N(0,\ 1)$. The lower plot draws the independent transformed data in each interval over $j = 0, 1, 2, \cdots, 25$. This plot is the same plot as the panel unit root data with 10 individuals over 25 time points. As can be seen, the plot spreads out as n increases, which display a property of a unit root process, that is, the increase of variance.

Table 3.1: Data Transformation Process

Original Data	Double-index Transformation	Independent Transformation
$Y_0, Y_1, Y_2, \cdots, Y_T$	$Y_{11}, Y_{12}, \cdots, Y_{1n}$	$0, Y_{11}, Y_{12}, \cdots, Y_{1n}$
	$\vdots$	$\vdots$
	$Y_{m1}, Y_{m2}, \cdots, Y_{mn}$	$0, Y_{m1} - Y_{m-1,n}, \cdots, Y_{mn} - Y_{m-1,n}$



(a) Original



(b) Transformation

Figure 3.1: Transformation of Simple AR (1) Model

3.1.2 Estimation and asymptotic results

The transformed model is the same form as in the panel data model so that the Pooled OLS (POLS) estimator for ρ is given by

$$\hat{\rho} \;=\; \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij}y_{ij-1}}{\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij-1}^2} \tag{3.23}$$

As mentioned in Chapter 2, we apply the sequential limits to obtain the limiting distribution of $\hat{\rho}$. Under the null hypothesis that $\rho = 1$, $y_{ij} - y_{i,j-1} = \epsilon_{ij}$, so we obtain the $n\sqrt{m}$-consistent ρ-statistic as in the panel unit root

$$n\sqrt{m}\cdot(\hat{\rho} \;-\; 1) \;= \frac{(\sqrt{m}n)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij-1}\epsilon_{ij}}{m^{-1}n^{-2}\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij-1}^2} \tag{3.24}$$

The t-statistic is given by

$$t_\rho = \frac{\hat{\rho}-1}{\sqrt{\hat{\sigma}^2/\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij-1}^2}} = \frac{\sigma^{-2}\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij-1}\epsilon_{ij}}{\frac{\hat{\sigma}}{\sigma}\sqrt{\sigma^{-2}\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij-1}^2}} \tag{3.25}$$

These results imply that the POLS estimator adopting the sequential limit is "super-consistent" for n, and is asymptotically normally distributed as $m \to \infty$ in contrast to the results of standard univariate unit root tests.

Lemma 3.1 *Under the null hypothesis $\rho = 1$, for given Model M1,*

$$\hat{\sigma}^2 = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij} - \hat{\rho}y_{ij-1})^2 \xrightarrow{p} \sigma^2. \tag{3.26}$$

Proof of Lemma 3.1

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij}-\hat{\rho}y_{ij-1})^2 \\ &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(\epsilon_{ij}-(\hat{\rho}-\rho)y_{ij-1})^2 \\ &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\epsilon_{ij}^2+(\hat{\rho}-\rho)^2\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}y_{ij-1}^2-2(\hat{\rho}-\rho)\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}y_{ij-1}\epsilon_{ij}\end{aligned}$$

since $\epsilon_{ij} = y_{i,j} - \rho y_{ij-1}$,

$$\begin{aligned}&= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\epsilon_{ij}^2-(\hat{\rho}-\rho)\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}y_{ij-1}\epsilon_{ij} \\ &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\epsilon_{ij}^2-(\hat{\rho}-\rho)\frac{1}{\sqrt{m}}\left(\frac{1}{\sqrt{mn}}\sum_{i=1}^{m}\sum_{j=1}^{n}y_{ij-1}\epsilon_{ij}\right)\end{aligned}$$

since $\hat{\rho}-\rho = O_p(m^{-1/2}n^{-1})$ and $\frac{1}{\sqrt{mn}}\sum_{i=1}^{m}\sum_{j=1}^{n}y_{ij-1}\epsilon_{ij} = O_p(1)$,

$$= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\epsilon_{ij}^2+O_p(\frac{1}{mn}) \xrightarrow{p} \sigma^2.$$

Theorem 3.1 *(Levin and Lin, 1992)Under the null hypothesis that $\rho = 1$ and Asumption 3.1, as $(n,m)_{\text{seq}} \to \infty$, the following hold:*

$$n\sqrt{m}\cdot(\hat{\rho}\ -\ 1)\ \ \Rightarrow\ \ N\big(0,\ 2\big) \tag{3.27}$$

$$t_\rho\ \ \Rightarrow\ \ N\big(0,\ 1\big) \tag{3.28}$$

Proof of Theorem 3.1

$$n\sqrt{m}\cdot(\hat{\rho}\ -\ 1)\ =\ \frac{\frac{1}{n\sqrt{m}}\sum_{i=1}^{m}\sum_{j=1}^{n}y_{ij-1}\epsilon_{ij}}{\frac{1}{n^2m}\sum_{i=1}^{m}\sum_{j=1}^{n}y_{ij-1}^2} = \frac{\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left(\frac{1}{n}\sum_{j=1}^{n}y_{ij-1}\epsilon_{ij}\right)}{\frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{n^2}\sum_{j=1}^{n}y_{ij-1}^2\right)}$$

By applying sequential limits, the n limit is

$$\overset{n\to\infty}{\Longrightarrow}\ \frac{\frac{1}{\sqrt{m}}\frac{1}{2}\sum_{i=1}^{m}(W_i^2(1)-1)}{\frac{1}{m}\sum_{i=1}^{m}\int_0^1 W_i^2(r)dr},$$

where $\Longrightarrow$ denotes weak convergence and $W_i(r), i = 1, 2, \cdots, m$, are independent standard Wiener processes. See Chan and Wei (1987, 1988) and Hall and Heyde (1980) for the derivation of the functional of Wiener process. Since $\int_0^1 W_i^2(r)dr$ are well defined random variables by Fuller (1996) in the sense that they have the finite first two moments and the W_i are independent of each other,

$$\frac{1}{m}\sum_{i=1}^{m}\int_0^1 W_i^2(r)dr \xrightarrow{p} \mathrm{E}\left(\int_0^1 W^2(r)dr\right) = \frac{1}{2} \quad \text{as } m \to \infty$$

$$\frac{1}{2\sqrt{m}}\sum_{i=1}^{m}\left(W_i^2(1) - 1\right) \xrightarrow{d} N\left(0,\ \frac{1}{2}\right) \quad \text{as } m \to \infty$$

Therefore, by the Slutsky theorem, this sequential limit approach gives

$$n\sqrt{m}\big(\hat{\rho}\ -\ 1\big) \ \Longrightarrow\ N\big(0,\ 2\big)$$

The t_ρ-statistic is given by

$$t_\rho = \frac{\hat{\rho} - 1}{\sqrt{\hat{\sigma}^2 / \sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij-1}^2}} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij-1}\epsilon_{ij}}{\hat{\sigma}\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij-1}^2}}$$

$$= \frac{\frac{1}{n\sqrt{m}}\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij-1}\epsilon_{ij}/\sigma^2}{\frac{\hat{\sigma}}{\sigma}\sqrt{\frac{1}{mn^2}\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij-1}^2/\sigma^2}} \xrightarrow{d} N\big(0,\ 1\big)$$

In the last equality, we used Lemma 3.1 and the above asymptotic results.

Alternatively, we can use exact moments for interval i which appear in Dickey (1976). Especially, the exact moments obtained are very useful to study the asymptotic behavior when the size of interval, n, is fixed. In that case, the asymptotics are functionals of n as mentioned in chapter 2. Define

$$N_{in} = \frac{1}{n\sigma^2}\sum_{j=1}^{n} y_{ij-1}\epsilon_{ij}, \qquad \text{and} \qquad D_{in} = \frac{1}{n^2\sigma^2}\sum_{j=1}^{n} y_{ij-1}^2.$$

Table 3.2: Moments of an Simple AR(1) Model (Dickey, 1976)

$\mathrm{E}(N_i) = 0$	$\mathrm{E}(N_{in}) = 0$
$\mathrm{E}(N_i^2) = \frac{1}{2}n(n-1)\sigma^4$	$\mathrm{E}(N_{in}^2) = \frac{1}{2} + o(\frac{1}{n})$
$\mathrm{Var}(N_i) = \frac{1}{2}n(n-1)\sigma^4$	$\mathrm{Var}(N_{in}) = \frac{1}{2} + o(\frac{1}{n})$
$\mathrm{E}(D_i) = \frac{1}{2}n(n-1)\sigma^2$	$\mathrm{E}(D_{in}) = \frac{1}{2} + o(\frac{1}{n})$
$\mathrm{E}(D_i^2) = \frac{1}{12}n(n-1)(7n^2-7n+4)\sigma^4$	$\mathrm{E}(D_{in}^2) = \frac{7}{12} + o(\frac{1}{n})$
$\mathrm{Var}(D_i) = \frac{1}{3}n(n-1)(n^2-n+1)\sigma^4$	$\mathrm{Var}(D_{in}) = \frac{1}{3} + o(\frac{1}{n})$
$\mathrm{Cov}(N_i, D_i) = \frac{1}{3}(n-1)n(n-2)\sigma^4$	$\mathrm{Cov}(N_{in}, D_{in}) = o(\frac{1}{n})$

The moments for N_i and D_i are listed in the above table which is also useful for the case that $m \to \infty$ and n is fixed.

Using these results, we can obtain the asymptotic results:

$$\begin{aligned} n\sqrt{m}\cdot(\hat{\rho} \ - \ 1) = & \ \frac{(\sqrt{m}n)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij-1}\epsilon_{ij}}{m^{-1}n^{-2}\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij-1}^2} \\ = & \ \frac{m^{-1/2}\sum_{i=1}^{m} N_{in}}{m^{-1}\sum_{i=1}^{m} D_{in}} = \frac{N_m}{D_m} \ \xrightarrow{d} \ N(0,2) \quad \text{as } m \to \infty \end{aligned}$$

and

$$\begin{aligned} t_\rho = & \ \frac{\hat{\rho}-1}{\sqrt{\hat{\sigma}^2/\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij-1}^2}} = \frac{m^{-1/2}\sum_{i=1}^{m} N_{in}}{\left(\frac{\hat{\sigma}}{\sigma}\right)\sqrt{m^{-1}\sum_{i=1}^{m} D_{in}}} \\ & \xrightarrow{d} N(0,1). \end{aligned}$$

Since we know that

$$\frac{1}{\sqrt{m}}\sum_{i=1}^{m} N_{in} \xrightarrow{d} N\left(0, \frac{1}{2}\right), \quad \frac{1}{m}\sum_{i=1}^{m} D_{in} \xrightarrow{p} \frac{1}{2}, \quad \text{and} \quad \hat{\sigma}^2 \xrightarrow{p} \sigma^2,$$

by the Slutsky theorem we have the results as required.

3.1.3 Behavior under the alternative

Now we study the statistical behavior of the POLS estimator under the alternative $|\rho| < 1$. To this end, we first check whether or not the estimator of ρ for Model 1 (3.21) is consistent.

$$\hat{\rho} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{i,j-1} - Y_{i-1,n})(Y_{ij} - Y_{i-1,n})}{\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{i,j-1} - Y_{i-1,n})^2} \qquad \text{with } Y_{0,n} = 0$$

$$= \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij}Y_{ij-1} - Y_{ij}Y_{i-1,n} - Y_{ij-1}Y_{i-1,n} + Y_{i-1,n}^2)}{\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij-1}^2 - 2Y_{ij-1}Y_{i-1,n} + Y_{i-1,n}^2)}$$

$$= \frac{\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij}Y_{ij-1} - Y_{ij}Y_{i-1,n} - Y_{ij-1}Y_{i-1,n} + Y_{i-1,n}^2)}{\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij-1}^2 - 2Y_{ij-1}Y_{i-1,n} + Y_{i-1,n}^2)}$$

$$\xrightarrow{p} \frac{(\rho+1)\gamma_0}{2\gamma_0} = \frac{1}{2}(\rho+1)$$

where

$$\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} Y_{ij}Y_{ij-1} \xrightarrow{p} \gamma_1 = \rho\gamma_0$$
$$\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} Y_{ij}Y_{i-1,n} \xrightarrow{p} \gamma_j = \rho\gamma_{j-1}.$$

Under the null hypothesis $\rho = 1$, the POLS estimator of ρ is consistent, that is, $\hat{\rho} \xrightarrow{p} 1$. On the other hand, under the alternative this estimator is inconsistent and the fact that it converges to the average of ρ and the null hypothesis value 1 suggests a loss of power. Breitung and Meyer (1994) also showed this property, but they did not show the limiting distribution of the estimator.

Next we study the limiting distribution of the POLS estimator under the alternative. With simple algebra, we know that $\widehat{\rho}-1 \xrightarrow{p} \frac{1}{2}(\rho-1)$. Let us define $\rho^* = \frac{1}{2}(\rho-1)$ and $\widehat{\rho^*} = \widehat{\rho} - 1$. We consider the following regression

$$y_{ij} - y_{ij-1} = (\rho - 1)y_{ij-1} + \epsilon_{ij}.$$

Using $y_{ij} = Y_{ij} - Y_{i-1,n}$, it can be rewritten as

$$Y_{ij} - Y_{ij-1} = (\rho - 1)(Y_{ij-1} - Y_{i-1,n}) + \epsilon_{ij}.$$

Then the POLS estimator of ρ^* is given by

$$\widehat{\rho^*} = \frac{\sum_{i=1}^m \sum_{j=1}^n (Y_{ij} - Y_{ij-1})(Y_{ij-1} - Y_{i-1,n})}{\sum_{i=1}^m \sum_{j=1}^n (Y_{ij-1} - Y_{i-1,n})^2}.$$

In order to obtain asymtotic normality, we subtract $\rho^* = \frac{1}{2}(\rho-1)$ from the both sides

$$\widehat{\rho^*}-\rho^* = \frac{\sum_{i=1}^m \sum_{j=1}^n \{Y_{ij} - Y_{ij-1} - \rho^*(Y_{ij-1} - Y_{i-1,n})\}(Y_{ij-1} - Y_{i-1,n})}{\sum_{i=1}^m \sum_{j=1}^n (Y_{ij-1} - Y_{i-1,n})^2}$$

$$= \frac{\sum_{i=1}^m \sum_{j=1}^n \{\epsilon_{ij} + (\rho-1)Y_{ij-1} - \rho^*(Y_{ij-1} - Y_{i-1,n})\}(Y_{ij-1} - Y_{i-1,n})}{\sum_{i=1}^m \sum_{j=1}^n (Y_{ij-1} - Y_{i-1,n})^2}$$

$$= \frac{\{\epsilon_{ij} + \rho^*(Y_{ij-1} + Y_{i-1,n})\}(Y_{ij-1} - Y_{i-1,n})}{\sum_{i=1}^m \sum_{j=1}^n (Y_{ij-1} - Y_{i-1,n})^2}$$

$$= \frac{\sum_{i=1}^m \sum_{j=1}^n \{\epsilon_{ij} + \rho^*(Y_{ij-1} + Y_{i-1,n})\}(Y_{ij-1} - Y_{i-1,n})}{\sum_{i=1}^m \sum_{j=1}^n (Y_{ij-1} - Y_{i-1,n})^2}$$

As in the panel data framework, multiplying $\sqrt{mn}$ on the both sides yields

$$\sqrt{mn}\big(\widehat{\rho^*}-\rho^*\big) = \frac{\frac{1}{\sqrt{mn}}\big[\sum_{i=1}^m \sum_{j=1}^n \{\epsilon_{ij} + \rho^*(Y_{ij-1} + Y_{i-1,n})\}(Y_{ij-1} - Y_{i-1,n})\big]}{\frac{1}{mn}\sum_{i=1}^m \sum_{j=1}^n (Y_{ij-1} - Y_{i-1,n})^2}.$$

Since we know the asymptotic behavior of the denominator, if we know the asymptotic behavior of the numerator, then we obtain the limiting distribution of the pooled OLS estimator of ρ under the alternative. The numerator, say N_{mn}, can be rewritten as

$$\begin{aligned} N_{mn} &= \frac{1}{\sqrt{mn}} \sum_{i=1}^{m} \sum_{j=1}^{n} \epsilon_{ij}(Y_{ij-1} - Y_{i-1,n}) + \rho^* \frac{1}{\sqrt{mn}} \sum_{i=1}^{m} \sum_{j=1}^{n} (Y_{ij-1}^2 - Y_{i-1,n}^2) \\ &\stackrel{\text{def}}{=} A \ + \ \rho^* B. \end{aligned}$$

In Assumption 3.2, we assume that $\mathrm{E}\{|\epsilon_{ij}|^{2+\delta}\} \leq M < \infty$ for some $\delta > 0$. In order to obtain the result in part A, especially, we assume that $\mathrm{E}\{\epsilon_{ij}^4\} = \eta\sigma^4 < \infty$ as in Theorem 6.2.1 of Fuller (1996, p. 314). Also, to prove part B, we assume that $\mathrm{E}\{\epsilon_{ij}^6\} = \xi\sigma^6 < \infty$ as in Theorem 6.2.3 of Fuller(1996, p.317) where if ϵ_{ij} are normally distributed, then $\eta = 3$. That is, to obtain the probability limit of the estimator of the autocovariance function for a stationary series, we need the finite fourth moment, and we need the finite sixth moment for the probability limit of the variance-covariance estimator of autocovariance function.

For part A, since ϵ_{ij} is independent of $(Y_{ij-1} - Y_{i-1,n})$ $\forall i, j$, $\mathrm{E}(A) = 0$ and

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^{n} \mathrm{E}\big[\epsilon_{ij}^2 (Y_{ij-1} - Y_{i-1,n})^2\big] &= \frac{1}{n} \sum_{j=1}^{n} \mathrm{E}\Big[\epsilon_{ij}^2 (Y_{i,j-1}^2 + Y_{i-1,n}^2) - 2\epsilon_{ij}^2 Y_{i,j-1} Y_{i-1,n}\Big] \\ &= \frac{1}{n} \sum_{j=1}^{n} 2(\gamma_0 + \gamma_{j-1})\sigma^2 \qquad (\because \sum_h |\gamma_h| < \infty) \\ &= 2\gamma_0\sigma^2 + O(n^{-1}). \end{aligned}$$

Hence, we have the following result

$$\frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \big[\epsilon_{ij}^2 (Y_{ij-1} - Y_{i-1,n})^2\big] \xrightarrow{p} 2\gamma_0\sigma^2 = \frac{2}{1-\rho^2}\sigma^4.$$

Thus we obtain the asymptotic distribution for A as follows:

$$A \xrightarrow{d} N\left(0,\ \frac{2}{1-\rho^2}\sigma^4\right).$$

For part B, by WLLN we obtain the asymptotic results for the first two moments:

$$\begin{aligned}\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij-1}^2 - Y_{i-1,n}^2) &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}Y_{i,j-1}^2 - \frac{1}{m}\sum_{i=1}^{m}Y_{i-1,n}^2 \\ &\xrightarrow{p} \frac{1}{n}\sum_{j=1}^{n}\mathrm{E}Y_{i,j-1}^2 - \mathrm{E}Y_{i-1,n}^2 = 0.\end{aligned}$$

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n}\mathrm{E}(Y_{ij-1}^2 - Y_{i-1,n}^2)^2 &= \frac{1}{n}\sum_{j=1}^{n}[\mathrm{E}(Y_{ij-1}^4) + \mathrm{E}(Y_{i-1,n}^4) - 2\mathrm{E}(Y_{ij-1}^2 Y_{i-1,n}^2)] \\ &= \frac{1}{n}\sum_{j=1}^{n}[2\mathrm{E}(Y_{ij-1}^4) - 2\mathrm{E}(Y_{ij-1}^2 Y_{i-1,n}^2)] \\ &= \frac{1}{n}\sum_{j=1}^{n}\left[2\cdot 3\{\mathrm{E}(Y_{ij-1}^2)\}^2 - 2\{\mathrm{E}(Y_{ij-1}^2)\mathrm{E}(Y_{i-1,n}^2) + 2(\mathrm{E}(Y_{ij-1}Y_{i-1,n}))^2\}\right] \\ &= \frac{1}{n}\sum_{j=1}^{n}(4\gamma_0^2 - 4\gamma_j^2) \\ &= 4\gamma_0^2 + O(n^{-1}) \ \text{ uniformly in } i.\end{aligned}$$

Hence,

$$\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij-1}^2 - Y_{i-1,n}^2)^2 \xrightarrow{p} 4\gamma_0^2.$$

For B we have the asymptotic distribution

$$B \xrightarrow{d} N\left(0,\ \frac{4}{(1-\rho^2)^2}\sigma^4\right).$$

Now we should check the correlation between A and B.

$$\frac{1}{n}\sum_{j=1}^{n}\mathrm{E}[\epsilon_{ij}(Y_{ij-1} - Y_{i-1,n})(Y_{ij-1}^2 - Y_{i-1,n}^2)] = 0.$$

$$\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}[\epsilon_{ij}(Y_{ij-1} - Y_{i-1,n})(Y_{ij-1}^2 - Y_{i-1,n}^2)] \xrightarrow{p} 0.$$

This result implies that A and B are asymptotically uncorrelated.

$$\mathrm{E}(AB) \;=\; 0.$$

The result that A and B are asymptotically individually normal does not directly imply the bivariate asymptotic normality of the pair of (A, B). In this case, we can prove the asymptotic normality via the Cramer-Wold device. Let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)' \neq \mathbf{0}$ and $\mathbf{X}_n = (A, B)'$. Then $\boldsymbol{\lambda}\mathbf{X}_n \xrightarrow{d} \boldsymbol{\lambda}\mathbf{X}$ where

$$\mathbf{X} \sim N\bigg(0,\; \frac{2}{1-\rho^2}\sigma^4 + \frac{4}{(1-\rho^2)^2}\sigma^4\bigg).$$

Using the preceding results, it follows rather easily that

$$\boldsymbol{\lambda}\mathbf{X}_n = \frac{1}{\sqrt{mn}} \sum_{i=1}^{m}\sum_{j=1}^{n} \Big[\lambda_1 \epsilon_{i,j}(Y_{i,j-1} - Y_{i-1,n}) + \lambda_2(Y_{i,j-1}^2 - Y_{i-1,n}^2)\Big]$$

converges to a normal random variable with mean 0 and variance $\lambda_1^2\mathrm{Var}(A)+\lambda_2^2\mathrm{Var}(B)$, thus establishing the joint asymptotic normality of A and B and in particular for $\lambda_1 = 1$ and $\lambda_2 = \rho^*$ it establishes the asymptotic distribution of N_{mn}.

Therefore,

$$\begin{aligned} N_{mn} \;&\xrightarrow{d}\; N\bigg(0,\; \frac{2}{1-\rho^2}\sigma^4 + \frac{1}{4}(\rho-1)^2 4\gamma_0^2\bigg) \\ &= N\bigg(0,\; \frac{2}{1-\rho^2}\sigma^4 + \frac{1}{(1+\rho)^2}\sigma^4\bigg) \\ &= N\bigg(0,\; \frac{3+\rho}{(1-\rho^2)(1+\rho)}\sigma^4\bigg) \end{aligned}$$

Finally, by the Slutsky theorem we obtain the limiting distribution of $\widehat{\rho^*} - \rho^*$

$$\sqrt{mn}\big(\widehat{\rho^*} - \rho^*\big) \;\xrightarrow{d}\; N\bigg(0,\; \frac{(3+\rho)(1-\rho)}{4}\bigg) \tag{3.29}$$

For this regression model, the mean squared error is given by

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}[Y_{ij} - Y_{ij-1} - (\hat{\rho}-1)(Y_{ij-1} - Y_{i-1,n})]^2 \\
&= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}[\epsilon_{ij} + (\rho-1)Y_{ij-1} - (\hat{\rho}-1)(Y_{ij-1} - Y_{i-1,n})]^2 \\
&= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}[\epsilon_{ij} + (\rho-\hat{\rho})Y_{ij-1} + (\hat{\rho}-1)Y_{i-1,n}]^2 \\
&= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left[\epsilon_{ij} + \left(\frac{\rho+1}{2} - \hat{\rho}\right)(Y_{ij-1} - Y_{i-1,n}) + \left(\frac{\rho-1}{2}\right)(Y_{ij-1} + Y_{i-1,n})\right]^2 \\
&= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\epsilon_{ij}^2 + \left(\hat{\rho} - \frac{\rho+1}{2}\right)^2\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij-1} - Y_{i-1,n})^2 \\
&\quad + \left(\frac{\rho-1}{2}\right)^2\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij-1} + Y_{i-1,n})^2 \\
&\quad - 2\left(\frac{\rho-1}{2}\right)\left(\hat{\rho} - \frac{\rho+1}{2}\right)\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij-1}^2 - Y_{i-1,n}^2) + \text{other cross product terms} \\
&\xrightarrow{p} \sigma^2 + 2\left(\frac{1-\rho}{2}\right)^2\gamma_0 = \frac{3+\rho}{2(1+\rho)}\sigma^2
\end{aligned}
$$

Now we can obtain the asymptotic distribution of the t-statistic

$$
\begin{aligned}
t &= \frac{\widehat{\rho^*} - \rho^*}{\sqrt{\hat{\sigma^2}\cdot\left(\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{i,j-1} - Y_{i-1,n})^2\right)^{-1}}} \\
&= \frac{\sqrt{mn}(\widehat{\rho^*} - \rho^*)}{\sqrt{\hat{\sigma^2}\cdot\left(\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{i,j-1} - Y_{i-1,n})^2\right)^{-1}}} \\
&\xrightarrow{d} N(0,\ V) \qquad \text{by the Slutsky theorem}
\end{aligned}
$$

where

$$V = \frac{2(1+\rho)}{3+\rho}\frac{2}{1-\rho^2}\frac{(3+\rho)(1-\rho)}{4} = 1.$$

Therefore, we have the following results:

$$\sqrt{mn}(\widehat{\rho^*} - \rho^*) \xrightarrow{d} N\left(0,\ \frac{(3+\rho)(1-\rho)}{4}\right) \quad \text{for } \rho^* = \frac{1}{2}(\rho - 1)$$

$$t \xrightarrow{d} N(0,\ 1).$$

3.2 Model 2: Intercept AR (1) model

In this section we consider an AR (1) model with intercept under the alternative hypothesis given by

$$Y_t \ = \ \alpha(1-\rho) \ + \ \rho Y_{t-1} \ + \ \epsilon_t \tag{3.30}$$

where $t = 1, 2, \cdots, T$ and ϵ_t follows Asummption 3.1. Using the double-index transformation we can rewrite Equation (3.30):

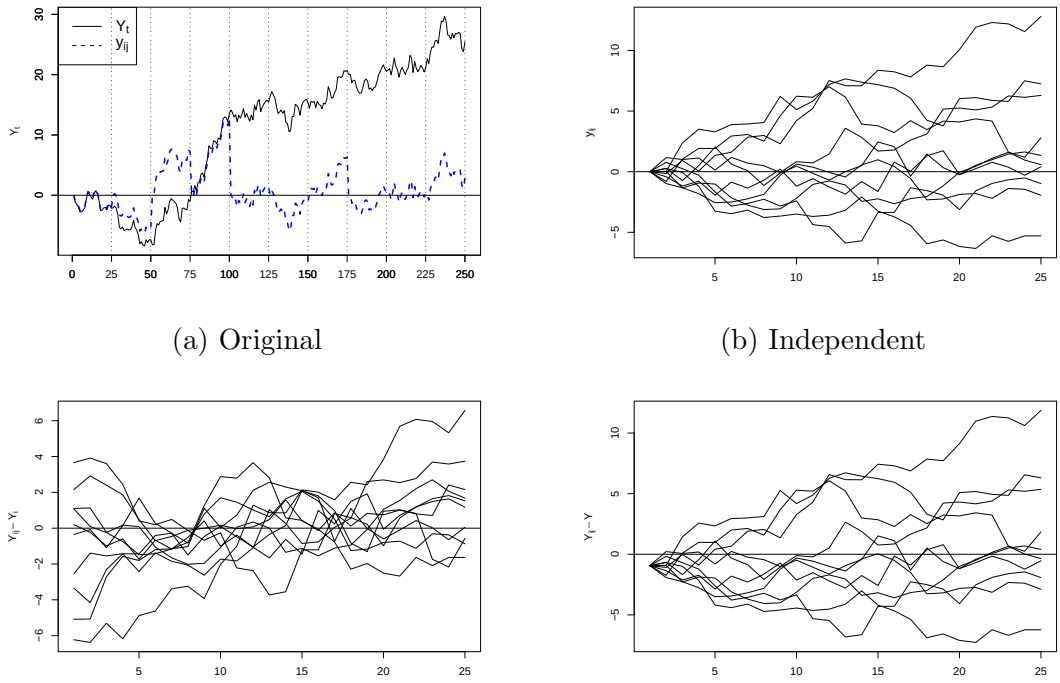
$$Y_{i,j} \ = \ \alpha(1-\rho) \ + \ \rho Y_{i,j-1} \ + \ \epsilon_{ij}, \tag{3.31}$$

where $t = (i-1)n + j, i = 1, \cdots, m,\ j = 1, \cdots, n$, and $mn = T$. Now, under the null hypothesis that $\rho = 1$ the independent transformation for Equation (3.31) yields

$$y_{i,j} \ = \ y_{i,j-1} \ + \ \epsilon_{i,j} \tag{3.32}$$

where $y_{i,j} = Y_{i,j} - Y_{i-1,n}$. The last value subtraction for the independent transformation removes the constant term in both null and alternative hypotheses. We call

Equation (3.32) model M2-GD or M2-ID which are exactly the same as model M1. Thus, for Equation (3.32) we can use the methodology for the simple AR (1). This method is adopted by Breitung and Meyer (1994) for panel unit root tests so they could avoid the bias in the estimate of ρ due to the incidental parameter.



(a) Original

(b) Independent

(c) Interval Mean

(d) Global Mean

Figure 3.2: Transformation of Intercept AR (1) Model

We draw the transformation process for the intercept AR (1) regression model in Figure 3.2. The total number of observations (T) is 250, the number of intervals (m) is 10, and the number of observations in each interval (n) is 25. Because the simple AR(1) model and the intercept AR(1) model are the same under the null hypothesis, we have the same plots for the original and independent transformed data in (a) and (b). Plot (c) illustrates the subtraction of interval means from the independent

transformed data in each interval, and plot (d) represents the subtraction of the global mean from the transformed data. In the following subsection, we address the issues related to those means in detail. Comparing to the independent transformation, the subtraction of interval means results in a modification of the plot, but the global mean case makes just a shift-down without modification of the shape. This shift decreases as m gets large.

3.2.1 Global mean model: M2-G

In this section we consider an AR (1) model with global mean under the alternative hypothesis for the transformed data. Levin and Lin (1992) used the term "aggregate mean" in panel unit root tests instead of global mean. The regression model after the independent transformation is given by

$$y_{i,j} = \mu + \rho y_{i,j-1} + \epsilon_{i,j} \tag{3.33}$$

where $y_{i,j} = Y_{i,j} - Y_{i-1,n}$, $\mu = \alpha(1-\rho)$, and $\epsilon_{i,j}$ follows Assumption 3.2. We call Equation (3.33) model M2-G. As mentioned above, although the true model after transformation does not involve a constant term, we introduce the term in the regression model because an intercept is often supplied by default in regression software. Since the estimate of the constant term converges in probability to zero, this constant does not affect the limiting distribution of the estimate of ρ. The POLS estimator of

ρ for Equation (3.33) is given by

$$\hat{\rho}_\mu = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{i,j} - \bar{y}_0)(y_{i,j-1} - \bar{y}_{-1})}{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{i,j-1} - \bar{y}_{-1})^2} \tag{3.34}$$

$$= \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{i,j} - \bar{Y}_0)(Y_{i,j-1} - \bar{Y}_{-1})}{\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{i,j-1} - \bar{Y}_{-1})^2} \tag{3.35}$$

where $\bar{y}_s = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} y_{i,j+s} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{i,j+s} - Y_{i-1,n})$ for $s = 0, -1$. At this point, we distinguish $\bar{y}$ from $\overline{Y} = T^{-1}\sum_{t=1}^{T} Y_t$ which is calculated in the original data set. Equation (3.34) is the same as the estimator of ρ for the double-index transformed model in Equation (3.30) since $y_{i,j} - \bar{y_0} = Y_{i,j} - \bar{Y}$. That means that the POLS estimate of ρ is consistent for both hypotheses unlike the estimate in M1. The advantage of using model M2-G is that the estimate of ρ is consistent under both the null and alternative hypotheses. Also, the M2-G model attenuates the effect of a structural change in level (or coefficient of intercept)on the test statistics because after transformation only one interval is contaminated and the others remain with truly zero mean. Thus, as m increases the influence of the contaminated interval disappears. We will thoroughly discuss the topic of structual changes in Chapter 4. From Equation (3.34), under the null hypothesis that $\rho = 1$ we obtain

$$\hat{\rho}_\mu \ - \ 1 = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{i,j-1} - \bar{y}_{-1})\epsilon_{i,j}}{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{i,j-1} - \bar{y}_{-1})^2} \tag{3.36}$$

Lemma 3.2 *(Levin and Lin, 1992) Under Assumption* 3.2 *and* $H_0 : \rho = 1$*, for model M2-G,*

$$n^{-1/2}\bar{y} \xrightarrow{p} 0 \quad \textit{and} \quad (3m/n)^{1/2}\bar{y} \xrightarrow{d} N(0,\ 1). \tag{3.37}$$

Proof of Lemma 3.2

$$\frac{1}{n^{1/2}}\bar{y} = \frac{1}{n^{1/2}}\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} y_{i,j}$$
$$= \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{n^{3/2}}\sum_{j=1}^{n} y_{i,j}\right) \overset{n\to\infty}{\Longrightarrow} \frac{1}{m}\sum_{i=1}^{m}\int_0^1 W_i(r)dr \overset{m\to\infty}{\longrightarrow} \mathrm{E}\int_0^1 W(r)dr = 0.$$

$$\left(\frac{3m}{n}\right)^{\frac{1}{2}}\bar{y} = \sqrt{\frac{3}{m}}\sum_{i=1}^{m}\left(\frac{1}{n^{3/2}}\sum_{j=1}^{n} y_{i,j}\right)$$
$$\overset{n\to\infty}{\Longrightarrow} \sqrt{\frac{3}{m}}\sum_{i=1}^{m}\int_0^1 W_i(r)dr \overset{m\to\infty}{\longrightarrow} N(0,\ 1) \qquad \text{by the Lindeberg-Feller CLT}$$

Theorem 3.2 *(Levin and Lin, 1992) Under Assumption* 3.2 *and* $H_0 : \rho = 1$*, for model M2-G, as* $(n, m)_{\text{seq}} \to \infty$

$$n\sqrt{m}(\hat{\rho}_\mu - 1) \xrightarrow{d} N(0,\ 2) \tag{3.38}$$

$$t_\mu \xrightarrow{d} N(0,\ 1) \tag{3.39}$$

Proof of Theorem 3.2

$$n\sqrt{m}(\hat{\rho}_\mu - 1) = \frac{\frac{1}{n\sqrt{m}}(\sum_{i=1}^{m}\sum_{j=1}^{n} y_{i,j-1}\epsilon_{i,j}\bar{y}_{-1})\sum_{i=1}^{m}\sum_{j=1}^{n}\epsilon_{i,j}}{\frac{1}{mn^2}(\sum_{i=1}^{m}\sum_{j=1}^{n} y_{i,j-1}^1 - mn\bar{y}_{-1}^2)}$$

$$\frac{\frac{1}{n\sqrt{m}}\sum_{i=1}^{m}\sum_{j=1}^{n} y_{i,j-1}\epsilon_{i,j} - \frac{1}{\sqrt{n}}\bar{y}_{-1}\cdot\frac{1}{\sqrt{mn}}\sum_{i=1}^{m}\sum_{j=1}^{n}\epsilon_{i,j}}{\frac{1}{mn^2}\sum_{i=1}^{m}\sum_{j=1}^{n} y_{i,j-1}^2 - \frac{1}{n}\bar{y}_{-1}^2}$$

$$= \frac{\frac{1}{n\sqrt{m}}\sum_{i=1}^{m}\sum_{j=1}^{n} y_{i,j-1}\epsilon_{i,j} + o_p(1)}{\frac{1}{mn^2}\sum_{i=1}^{m}\sum_{j=1}^{n} y_{i,j-1}^2 + o_p(1)}.$$

using the result in Model 1 and by the Slutsky theorem, we obtain

$$n\sqrt{m}(\hat{\rho}_\mu - 1) \xrightarrow{d} N(0,\ 2).$$

Similarly, we can prove the asymptotic result for t_μ-statistic. Q.E.D.

3.2.2 Interval-specific mean model

Alternatively, when we transform the intercept model to a double-index model, we can consider interval-specific means instead of the global mean, which is the same as individual-specific fixed effects in panel data. This interval-specific mean model is fitting the interval mean for each interval, so we would expect that the M2-I model reflects the circumstance of each interval more accurately. However, in the presense of structural change in level we still have misfitting in the contaminated interval. The more serious problem in model M2-I is that the interval-specific mean leads to the biased estimator of ρ since the interval-specific mean depends only on the number of observations in each interval n, but not on the number of intervals m. In this case, the double-indexed transformation is given by

$$Y_{ij} = \mu_i + \rho Y_{ij-1} + \epsilon_{ij} \ ; \qquad i = 1, \cdots, m; \ j = 1, \cdots, n; \ mn = T \tag{3.40}$$

where $\mu_i = (1-\rho)\alpha_i$. For obtaining independence of each interval under the null hypothesis $\rho = 1$, we subtract the last value of the $(i-1)^{th}$ interval from each observation in the i^{th} interval for $i = 2, 3, \cdots, m$.

$$Y_{ij} - Y_{i-1,n} = \mu_i + \rho(Y_{ij-1} - Y_{i-1,n}) + \epsilon_{ij} \tag{3.41}$$

or equivalently,

$$y_{ij} = \mu_i + \rho y_{ij-1} + \epsilon_{ij}, \tag{3.42}$$

where $y_{ij} = Y_{ij} - Y_{i-1,n}$.

In the context of estimating the panel data model with individual-specific fixed

effects, the POLS estimator of ρ is given by

$$\begin{aligned}\hat{\rho}_{\mu_i} &= \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i,0})(y_{ij-1}-\bar{y}_{i,-1})}{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij-1}-\bar{y}_{i,-1})^2} \\ &= \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij}-\bar{Y}_{i,0})(Y_{ij-1}-\bar{Y}_{i,-1})}{\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij-1}-\bar{Y}_{i,-1})^2} \\ &= \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i,-1}-\bar{\epsilon}_{i,0})(y_{ij-1}-\bar{y}_{i,-1})}{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij-1}-\bar{y}_{i,-1})^2},\end{aligned} \tag{3.43}$$

where $\bar{Y}_{i,s} = n^{-1}\sum_{j=1}^{n} Y_{i,j+s}$, and similarly $\bar{\epsilon}_{i,s} = n^{-1}\sum_{j=1}^{n}\epsilon_{i,j+s}$ for $s = 0, -1$. Note that $y_{ij} - \bar{y}_{i,0} = Y_{ij} - \bar{Y}_{i,0}$ in the above equation. In the panel data regression, this estimator is called the least-squares dummy-variable (LSDV) estimator or covariance (CV) estimator. In vector notation, for $i = 1, \cdots, m$, Equation (3.42) can be written as:

$$\mathbf{y}_i = \mu_i\boldsymbol{\iota} + \rho\mathbf{y}_{i,-1} + \boldsymbol{\epsilon}_i.$$

To delete out the interval-specific mean effect, premultiplying by an $n \times n$ idempotent matrix, $\mathbf{Q} = \mathbf{I}_n - n^{-1}\boldsymbol{\iota}\boldsymbol{\iota}'$, yields

$$\mathbf{y}_i = \rho\mathbf{Q}\mathbf{y}_{i,-1} + \mathbf{Q}\boldsymbol{\epsilon}_i$$

and

$$\hat{\rho}_{CV} = \left[\sum_{i=1}^{m}\mathbf{y}'_{i,-1}\mathbf{Q}\ \mathbf{y}_{i,-1}\right]^{-1}\left[\sum_{i=1}^{m}\mathbf{y}'_{i,-1}\mathbf{Q}\ \mathbf{y}_i\right]$$

which is the same as the result in Equation (3.43). Under the null hypothesis that

$\rho = 1$, the estimator can be written as:

$$\begin{aligned}\hat{\rho}_{\mu_i} &= \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij-1} + \epsilon_{ij} - \bar{y}_{i,-1} - \bar{\epsilon}_{i,0})(y_{ij-1} - \bar{y}_{i,-1})}{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij-1} - \bar{y}_{i,-1})^2} \\ &= 1 \ + \ \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij-1} - \bar{y}_{i,-1})(\epsilon_{ij} - \bar{\epsilon}_{i,0})}{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij-1} - \bar{y}_{i,-1})^2}.\end{aligned} \tag{3.44}$$

Since $\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij-1} - \bar{y}_{i,-1})\bar{\epsilon}_{i,0} = \sum_{i=1}^{m}\bar{\epsilon}_{i,0}\sum_{j=1}^{n}(y_{ij-1} - \bar{y}_{i,-1}) = 0$, we obtain

$$\hat{\rho}_{\mu_i} = 1 \ + \ \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij-1} - \bar{y}_{i,-1})\epsilon_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij-1} - \bar{y}_{i,-1})^2}. \tag{3.45}$$

That is,

$$\hat{\rho}_{\mu_i} \ - \ 1 = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij-1} - \bar{y}_{i,-1})\epsilon_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij-1} - \bar{y}_{i,-1})^2} \qquad \text{under } H_0 \tag{3.46}$$

Unfortunately, this estimator is downward biased, and goes to the negative infinity with rate $m^{-1/2}$, as in the panel data model with individual-specific fixed effect as proved in Hsiao (2002, pp. 71-72). Hence we need to make a bias correction for this model. Using the delta method, we can approximate the degree of biasedness. Hence, we need to know the moment information of the estimator.

Theorem 3.3 *(Levin and Lin, 1992): Under Assumption* 3.2 *and* $H_0 : \rho = 1$*, for model M2-I, as* $(n, m)_{\text{seq}} \to \infty$*,*

$$\begin{aligned}
&a) \ \ \sqrt{m}n(\hat{\rho}_{\mu_i} - 1) - \frac{\sqrt{m}EN_n}{ED_n} \ \xrightarrow{d} \ N(0, 10.2) \\
&b) \ \ \textit{If } \frac{\sqrt{m}}{n} \to 0, \ \textit{then } \sqrt{m}n(\hat{\rho}_{\mu_i} - 1) + 3\sqrt{m} \ \xrightarrow{d} \ N(0, 10.2) \\
&c) \ \ \sqrt{1.25}\left\{t_\mu - \frac{\sqrt{m}EN_n}{\sqrt{ED_n}}\right\} \ \xrightarrow{d} \ N(0, 1) \\
&d) \ \ \textit{If } \frac{\sqrt{m}}{n} \to 0, \ \textit{then } \sqrt{1.25}\ t_\mu + \sqrt{1.875m} \ \xrightarrow{d} \ N(0, 1).
\end{aligned}$$

Table 3.3: Moments of the Intercept AR(1) Model (Dickey, 1976)

$\mathrm{E}\{N_i\} = -\frac{(n-2)}{2}\sigma^2$	$\mathrm{E}\{N_{in}\} = -\frac{1}{2} + o(\frac{1}{n})$
$\mathrm{E}\{N_i^2\} = \frac{n(n-2)}{3}\sigma^4$	$\mathrm{E}\{N_{in}^2\} = \frac{1}{3} + o(\frac{1}{n})$
$\mathrm{Var}\{N_i\} = \frac{(n-2)(n+6)}{12}\sigma^4$	$\mathrm{Var}\{N_{in}\} = \frac{1}{12} + o(\frac{1}{n})$
$\mathrm{E}\{D_i\} = \frac{n(n-2)}{6}\sigma^2$	$\mathrm{E}\{D_{in}\} = \frac{1}{6} + o(\frac{1}{n})$
$\mathrm{E}\{D_i^2\} = \frac{n(n-2)(n^2-n+2)}{20}\sigma^4$	$\mathrm{E}\{D_{in}^2\} = \frac{1}{20} + o(\frac{1}{n})$
$\mathrm{Var}\{D_i\} = \frac{n(n-2)(2n^2-4n+9)}{90}\sigma^4$	$\mathrm{Var}\{D_{in}\} = \frac{1}{45} + o(\frac{1}{n})$
$\mathrm{Cov}(N_i, D_i) = -\frac{n(n-2)}{6}\sigma^4$	$\mathrm{Cov}(N_{in}, D_{in}) = o(\frac{1}{n})$

In order prove Theorem 3.3 first define:

$$N_i = \sum_{j=1}^{n}(y_{ij-1} - \bar{y}_{i,-1})e_{ij} \quad \text{and} \qquad D_i = \sum_{j=1}^{n}(y_{ij-1} - \bar{y}_{i,-1})^2$$

$$N_{in} = \frac{1}{n\sigma^2}\sum_{j=1}^{n}(y_{ij-1} - \bar{y}_{i,-1})e_{ij} \quad \text{and} \qquad D_{in} = \frac{1}{n^2\sigma^2}\sum_{j=1}^{n}(y_{ij-1} - \bar{y}_{i,-1})^2$$

Proof of Theorem 3.3

For part (a)

$$n\sqrt{m}(\hat{\rho}_{\mu_i} - 1) \;=\; \sqrt{m}\,\frac{\sum_{i=1}^{m}\left\{\frac{1}{n}\sum_{j=1}^{n}(Y_{ij-1} - \bar{Y}_{i,-1})(\epsilon_{ij} - \bar{\epsilon}_{i,0})\right\}}{\sum_{i=1}^{m}\left\{\frac{1}{n^2}\sum_{j=1}^{n}(Y_{ij-1} - \bar{Y}_{i,-1})^2\right\}} \tag{3.47}$$

$$= \sqrt{m}\,\frac{\sum_{i=1}^{m} N_{in}}{\sum_{i=1}^{m} D_{in}} \tag{3.48}$$

Applying the sequential limit that $(n,m)_{seq} \to \infty$, as $n \to \infty$ we obtain that

$$\begin{aligned} N_{in} &= \frac{1}{n\sigma^2}\sum_{j=1}^{n} Y_{ij-1}e_{ij} - \frac{1}{n^{3/2}\sigma}\sum_{j=1}^{n} Y_{i\cdot}\frac{1}{n^{1/2}\sigma}\sum_{j=1}^{n} e_{ij} \\ &\overset{n\to\infty}{\Longrightarrow} \int_0^1 W_i(r)dW_i(r) - \int_0^1 W_i(r)dr\int_0^1 dW_i(r) = \int_0^1 \Psi_i(r)dW_i(r), \end{aligned}$$

and similarly

$$D_{in} \overset{n\to\infty}{\Longrightarrow} \int_0^1 \Psi_i^2(r)dr$$

where $\Psi_i(r) = W_i(r) - \int_0^1 W_i(r)dr$ is the demeaned standard Brownian motion. Therefore, we obtain the intermediate asymptotic result:

$$n\sqrt{m}(\hat{\rho}_{\mu_i} - 1) \overset{n\to\infty}{\Longrightarrow} \sqrt{m}\frac{\sum_{i=1}^{m}\int \Psi_i(r)dr}{\sum_{i=1}^{m}\int \Psi_i^2(r)dr} \overset{\text{def}}{=} \sqrt{m}\frac{N_m}{D_m} \tag{3.49}$$

where $N_m = \sum_{i=1}^{m} N_{in}$ and $D_m = \sum_{i=1}^{m} D_{in}$. The next step is to obtain the final asymptotic result as $m \to \infty$. As can be seen in Table 3.2.2, the numerator N_m has nonzero expected value, $-1/2$ so multiplying N_m by $\sqrt{m}$ diverges to negative infinity as $m \to \infty$. Delta method is a typcial method for removing this effect. Using the delta method and moments in Table 3.2.2, we obtain that

$$\mathrm{E}(\hat{\rho}_{\mu_i} - 1) \approx \frac{\mathrm{E}(N_i)}{\mathrm{E}(D_i)} = -\frac{3}{n}, \tag{3.50}$$

$$\begin{aligned} &\mathrm{Var}(\hat{\rho}_{\mu_i} - 1) \\ &\quad\approx \begin{pmatrix} \frac{1}{\mathrm{E}(D_i)} & -\frac{\mathrm{E}(N_i)}{\mathrm{E}^2(D_i)} \end{pmatrix} \begin{pmatrix} \frac{1}{mn^2}\mathrm{Var}(N_i) & \frac{1}{mn^2}C(N_i,D_i) \\ \frac{1}{mn^2}C(N_i,D_i) & \frac{1}{mn^2}\mathrm{Var}(D_i) \end{pmatrix} \begin{pmatrix} \frac{1}{\mathrm{E}(D_i)} \\ -\frac{\mathrm{E}(N_i)}{\mathrm{E}^2(D_i)} \end{pmatrix} \\ &\quad= \begin{pmatrix} 6 & 18 \end{pmatrix} \begin{pmatrix} \frac{1}{12} & 0 \\ 0 & \frac{1}{45} \end{pmatrix} \begin{pmatrix} 6 \\ 18 \end{pmatrix} = 10.2 \end{aligned}$$

Therefore, we know the bias pattern and variance of $\hat{\rho}_\mu$, which is given by

$$\hat{\rho}_{\mu_i} \;\approx\; N\!\left(1-\frac{3}{n},\; \frac{10.2}{mn^2}\right) \tag{3.51}$$

and hence, as $m \to \infty$ via the result in (3.49)

$$\sqrt{m}n(\rho_{\mu_i}-1) - \sqrt{m}\frac{\mathrm{E}(N_i n)}{\mathrm{E}(D_i n)} \;\xrightarrow{d}\; N(0,\; 10.2). \tag{3.52}$$

Since $\mathrm{E}(N_{in}) = -1/2 + o_p(n^{-1})$ and $\mathrm{E}(D_{in}) = 1/6 + o_p(n^{-1})$, if $\sqrt{m}/n \to 0$, then we obtain the result in part (b). For the t-statistic parts, we have similar results. Under the assumption that $\sqrt{m}/n \to 0$,

$$\mathrm{E}(t_\mu) \approx \frac{\mathrm{E}(N_i)}{\sqrt{\mathrm{E}(D_i)}} = -\frac{\sqrt{6}}{2n}, \tag{3.53}$$

$$\begin{aligned}
&\mathrm{Var}(t_\mu) \\
&\approx \begin{pmatrix} \frac{1}{\sqrt{\mathrm{E}(D_i)}} & -\frac{1}{2}\frac{\mathrm{E}(N_i)}{\mathrm{E}(D_i)^{3/2}} \end{pmatrix} \begin{pmatrix} \frac{1}{mn^2}V(N_i) & \frac{1}{mn^2}C(N_i,D_i) \\ \frac{1}{mn^2}C(N_i,D_i) & \frac{1}{mn^2}V(D_i) \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{\mathrm{E}(D_i)}} \\ -\frac{1}{2}\frac{\mathrm{E}(N_i)}{\mathrm{E}(D_i)^{3/2}} \end{pmatrix} \\
&= \begin{pmatrix} \sqrt{6} & \frac{1}{4}6^{3/2} \end{pmatrix} \begin{pmatrix} \frac{1}{12} & 0 \\ 0 & \frac{1}{45} \end{pmatrix} \begin{pmatrix} \sqrt{6} \\ \frac{1}{4}6^{3/2} \end{pmatrix} = \frac{4}{5}
\end{aligned}$$

For Model M2-I, the t-statistics has the following distribution:

$$t_{\mu_i} \;\approx\; N\!\left(-\frac{\sqrt{6}}{2},\; \frac{4}{5}\right) \tag{3.54}$$

and, the results in part (c) and (d) follow.

Alternatively, from equation 3.49, we can directly calculate the moments of N_m and

D_m. Since $\int_0^1 dW_i(r) = W_i(1)$, we have the following equalities

$$\int_0^1 \Psi_i(r)dW_i(r) = \frac{1}{2}(W_i^2(1) - 1) - W_i(1)\int_0^1 W_i(r)dr$$
$$\int_0^1 \Psi_i^2(r)dr = \int_0^1 W_i^2(r)dr - \left(\int_0^1 W_i(r)dr\right)^2.$$

$$\begin{aligned}
&\mathrm{E}\left[\frac{1}{2}(W_i^2(1) - 1) - W_i(1)\int_0^1 W_i(r)dr\right] \\
&= \frac{1}{2}\big\{\mathrm{E}(W_i^2(1)) - 1\big\} - \mathrm{E}\int_0^1 W_i(1)W_i(r)dr \\
&= -\mathrm{E}\int_0^1 W_i(1)W_i(r)dr \qquad (W_i^2(1) \sim \chi^2(1), \mathrm{E}(W_i^2(1)) = 1) \\
&= -\int_0^1 \mathrm{Cov}\{W_i(r), W_i(1)\}dr = -\int_0^1 \mathrm{Cov}\{W_i(r), [W_i(1) - W_i(r)] + W_i(r)\}dr \\
&= -\int_0^1 [\mathrm{Cov}\{W_i(r), W_i(1) - W_i(r)\} + \mathrm{Var}\{W_i(r)\}]dr \qquad (\text{ independent increments}) \\
&= -\int_0^1 \mathrm{Var}\{W_i(r)\}dr = -\int_0^1 r dr = -\frac{1}{2}
\end{aligned}$$

$$\begin{aligned}
&\mathrm{Var}\left[\frac{1}{2}(W_i^2(1) - 1) - W_i(1)\int_0^1 W_i(r)dr\right] \\
&= \frac{1}{4}\mathrm{Var}\{W_i^2(1)\} + \mathrm{Var}\left[\int_0^1 W_i(1)W_i(r)dr\right] - \mathrm{Cov}\left[W_i^2(1), \int_0^1 W_i(1)W_i(r)dr\right] \\
&\stackrel{def}{=} A + B + C.
\end{aligned}$$

Now we calculate A, B, and C separately.

$$\begin{aligned}
A &= \frac{1}{2} \\
B &= \mathrm{E}\int_0^1\int_0^1 W_i^2(1)W_i(r)W_i(s)drds - \left[\mathrm{E}\int_0^1 W_i(1)W_i(r)dr\right]^2 \\
&= \frac{5}{6} - \frac{1}{4} = \frac{7}{12}.
\end{aligned}$$

$$
\begin{aligned}
C &= -\mathrm{E}\int_0^1 W_i^3(1)W_i(r)dr - \mathrm{E}\{W_i^2(1)\} + \mathrm{E}\{W_i^2(1)\}\mathrm{E}\Big\{\int_0^1 W_i(1)W_i(r)dr\Big\} \\
&= -\mathrm{E}\int_0^1 \{W_i(1) - W_i(r) + W_i(r)\}^3 W_i(r)dr + \frac{1}{2} \\
&= -\int_0^1 \{\mathrm{E}(W_i(1) - W_i(r))^3 W_i(r) + 3\mathrm{E}(W_i(1) - W_i(r))^2 W_i^2(r) \\
&\quad + 3\mathrm{E}(W_i(1) - W_i(r))W_i^3(r) + \mathrm{E}W_i^4(r)\}dr + \frac{1}{2} \qquad \text{independent increments} \\
&= -\int_0^1 \{3\mathrm{E}(W_i(1) - W_i(r))^2 W_i^2(r) + \mathrm{E}W_i^4(r)\}dr + \frac{1}{2} \\
&= -\int_0^1 [3\{\mathrm{E}\{(W_i(1) - W_i(r))^2\}\mathrm{E}\{W_i^2(r)\} + \mathrm{E}W_i^4(r)]dr + \frac{1}{2} \\
&= -\int_0^1 \{3(1-r)r\} + 3r^2]dr + \frac{1}{2} = -\int_0^1 3rdr + \frac{1}{2} = -\frac{3}{2} + \frac{1}{2} = -1.
\end{aligned}
$$

Therefore,

$$
\mathrm{Var}\left[\frac{1}{2}(W_i^2(1) - 1) - W_i(1)\int_0^1 W_i(r)dr\right] = A + B + C = \frac{1}{2} + \frac{7}{12} - 1 = \frac{1}{12}.
$$

$$
\begin{aligned}
&\mathrm{E}\Big[\int_0^1 W_i^2(r)dr - \big(\int_0^1 W_i(r)dr\big)^2\Big] \\
&= \int_0^1 \mathrm{E}W_i^2(r)dr - \int_0^1\int_0^1 W_i(r)W_i(s)drds = \int_0^1 rdr - \int_0^1\int_0^1 \min\{r,s\}drds \\
&= \frac{1}{2} - \int_0^1 \left[\int_0^r sds + \int_r^1 rds\right]dr = \frac{1}{2} - \int_0^1 \left[\frac{1}{2}s^2\Big|_{s=0}^{r} + rs\Big|_{s=r}^{1}\right]dr \\
&= \frac{1}{2} - \frac{1}{3} = \frac{1}{6} \\
&\mathrm{Var}\Big[\int_0^1 W_i^2(r)dr - \big(\int_0^1 W_i(r)dr\big)^2\Big] \\
&= \mathrm{E}\big[\int\int\int W_i^2(r)W_i^2(s)drds - 2\int\int\int W_i^2(r)w_i(s)W_i(t)drdsdt \\
&\quad + \int\int\int\int W_i(r)W_i(sW_i(t)W_i(u)drdsdtdu)\big] - \frac{1}{36} \\
&= \frac{7}{12} + \frac{13}{15} + \frac{1}{3} - \frac{1}{36}.
\end{aligned}
$$

This result indicates that the small sample bias of $\widehat{\rho}_{\mu_i}$ depends only on the number of observations in each interval, and it vanishes as n gets large. For the fixed-effects in the dynamic panel model on the bias of $\widehat{\rho}_{\mu_i}$, refer to Anderson and Hsiao (1982), and Nickell (1981). The increase in the number of intervals reduces the sample variance but not the bias of $\widehat{\rho}_{\mu_i}$. However, since the bias does not depend on the interval-specific fixed effects, we can study the test statistics based on the sample sizes m and n.

The presence of interval-specific (individual-specific in panel data) effects causes the unit root test statistics to have non-central asymptotic normal distributions. The non-centrality depends only on the number of observations in each interval (time dimension), but not on the number of the intervals (cross-sectional dimension). Levin and Lin (1992) showed that small values of $\sqrt{m}/n$ work well for finite samples. That is, $\sqrt{m}/n \to 0$ is a relevant property in the sense that the bias of $\hat{\rho}_{\mu_i}$ is relatively small.

3.3 Model III: Linear trend AR (1) model

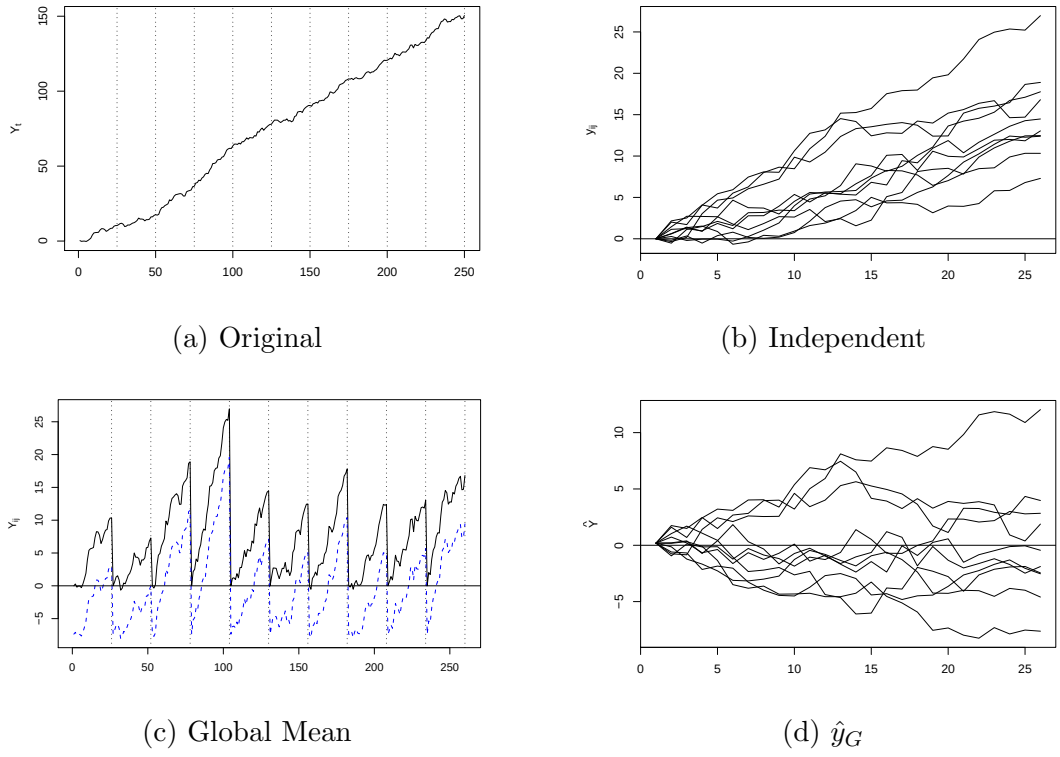
Consider the following model

$$Y_t - \alpha - \beta t = \rho(Y_{t-1} - \alpha - \beta(t-1)) + \epsilon_t, \qquad t = 1, \cdots, T \tag{3.55}$$

By double-index transformation, Equation (3.55) can be written as

$$Y_{i,j} - \alpha - \beta\{(i-1)n + j\} = \rho[Y_{i,j-1} - \alpha - \beta\{(i-1)n + j - 1\}] + \epsilon_{i,j} \tag{3.56}$$

where $t = (i-1)n + j; i = 1, \cdots, m;\ j = 1, \cdots, n$; and $mn = T$.



(a) Original

(b) Independent

(c) Global Mean

(d) $\hat{y}_G$

Figure 3.3: Transformation of Linear Trend AR (1) Model:M3-G

3.3.1 Global parameter model: M3-G and M3-GD

The global parameter model for the linear trend AR (1) model is given by

$$y_{i,j} = \mu + \delta j + \rho y_{i,j-1} + \epsilon_{i,j} \tag{3.57}$$

where $y_{i,j} = Y_{i,j} - Y_{i-1,n}, \mu = \alpha(1-\rho) + \beta\rho, \delta = \beta(1-\rho)$ and $\epsilon_{i,j}$ satisfies Assumption 3.2. Unlike Model M1, the last value transformation does not lead to inconsitency of the POLS estimate of ρ since the following properties hold:

- $y_{ij} - \overline{y} = Y_{ij} - \overline{Y}$
- $\hat{d} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij}-\overline{y})(j-\overline{j})}{\sum_{i=1}^{m}\sum_{j=1}^{n}(j-\overline{j})^2} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij}-\overline{Y})(j-\overline{j})}{\sum_{i=1}^{m}\sum_{j=1}^{n}(j-\overline{j})^2}$

- $\hat{y}_{ij} = y_{ij} - \overline{y} - \hat{d}(j - \overline{j}) = Y_{ij} - \overline{Y} - \hat{d}(j - \overline{j}) \;=\; \hat{Y}_{ij}$
- $\hat{\rho}_\tau = \frac{\sum_{i=1}^m \sum_{j=1}^n \hat{y}_{ij}\hat{y}_{i,j-1}}{\sum_{i=1}^m \sum_{j=1}^n \hat{y}_{i,j-1}^2} = \frac{\sum_{i=1}^m \sum_{j=1}^n \hat{Y}_{ij}\hat{Y}_{ij-1}}{\sum_{i=1}^m \sum_{j=1}^n \hat{Y}_{ij-1}^2}$

The two-step OLS estimators of M3-G can be expressed as follows:

Step 1. Regress $y_{i,j}$ on 1 and j to obtain $\hat{y}_{i,j}$:

$$\hat{y}_{i,j} = y_{i,j} - \bar{y} - \hat{d}(j - \bar{j})$$

where

$$\hat{\mu} = \bar{j} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} j = \frac{n+1}{2} \quad \text{and} \quad \hat{d} = \frac{\sum_{i=1}^m \sum_{j=1}^n (y_{i,j} - \bar{y})(j - \bar{j})}{\sum_{i=1}^m \sum_{j=1}^n (j - \bar{j})^2} \tag{3.58}$$

Step 2. Regress $\hat{y}_{i,j}$ on $\hat{y}_{i,j-1}$ to estimate ρ:

$$\hat{\rho}_\tau = \frac{\sum_{i=1}^m \sum_{j=1}^n \hat{y}_{i,j}\hat{y}_{i,j-1}}{\sum_{i=1}^m \sum_{j=1}^n \hat{y}_{i,j-1}^2}. \tag{3.59}$$

Lemma 3.3 *(Levin and Lin, 1992) Given Model M3-G and Assumption 3.2, under the null hypothesis that $\rho = 1$, as $(n, m)_{\text{seq}} \to \infty$*

$$\frac{1}{\sigma}\sqrt{mn}(\hat{d} - \beta) \xrightarrow{d} N\left(0,\ \frac{6}{5}\right) \quad \textit{and} \quad \sqrt{n}(\hat{d} - \beta) \xrightarrow{p} 0. \tag{3.60}$$

Proof of Lemma 3.3

$$
\begin{aligned}
(i)\frac{1}{\sigma}\sqrt{mn}(\hat{d}-\beta) &= \frac{m^{-1/2}n^{-5/2}\sigma^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}\{y_{i,j}-\bar{y}-\beta(j-\bar{j})\}(j-\bar{j})}{mn^{-3}\sum_{i=1}^{m}\sum_{j=1}^{n}(j-\bar{j})^2} \\
&= \frac{m^{-1/2}\sum_{i=1}^{m}\left(n^{-5/2}\sum_{j=1}^{n}\{y_{i,j}-\bar{y}-\beta(j-\bar{j})\}(j-\bar{j})\right)}{m^{-1}\sum_{i=1}^{m}\left(n^{-3}\sum_{j=1}^{n}(j-\bar{j})^2\right)} \\
&= \frac{m^{-1/2}\sum_{i=1}^{m}\left(n^{-5/2}\sum_{j=1}^{n}(w_{i,j}-\bar{w})(j-\bar{j})\right)}{m^{-1}\sum_{i=1}^{m}\left(n^{-3}\sum_{j=1}^{n}(j-\bar{j})^2\right)}, \qquad w_{i,j}=\sum_{s=1}^{j}\epsilon_{i,s} \\
&\stackrel{\text{def}}{=} \frac{m^{-1/2}\sigma^{-1}\sum_{i=1}^{m}N_{in}}{m^{-1}\sum_{i=1}^{m}D_{in}}.
\end{aligned}
$$

$$
\begin{aligned}
N_{in} &= n^{-5/2}\sum_{j=1}^{n}(j-\bar{j})w_{i,j} - n^{-5/2}\bar{w}\sum_{j=1}^{n}(j-\bar{j}) \stackrel{\text{def}}{=} A+B, \\
A &= n^{-5/2}\sum_{j=1}^{n}jw_{i,j} - \frac{1}{2}n^{-3/2}\sum_{j=1}^{n}w_{i,j} + o_p(n^{-1}) \\
&\Rightarrow \sigma\left(\int_0^1 rW_i(r)dr - \frac{1}{2}\int_0^1 W_i(r)dr\right) \\
B &= n^{-1/2}\bar{w}(n^{-2}\sum_{j=1}^{n}(j-\bar{j})) = 0 \qquad (\because \sum_{j=1}^{n}(j-\bar{j})=0) \\
D_{in} &= n^{-3}\big(\sum_{j=1}^{n}j^2 - n^{-1}\bar{j}^2\big) = \frac{1}{3}-\frac{1}{4}+o(n^{-1}) \to \frac{1}{12}.
\end{aligned}
$$

Hence, as $n\to\infty$ we obtain an intermediate asymptotic result for Equation (3.61)

$$
\begin{aligned}
\frac{m^{-1/2}\sum_{i=1}^{m}N_{in}}{m^{-1}\sum_{j=1}^{n}D_{in}} &\stackrel{n\to\infty}{\Longrightarrow} \frac{m^{-1/2}\sum_{i=1}^{m}\left(\int_0^1 rW_i(r)dr - \frac{1}{2}\int_0^1 W_i(r)dr\right)}{m^{-1}\sum_{i=1}^{m}\frac{1}{12}} \\
&= 12\cdot m^{-1/2}\sum_{i=1}^{m}\left(\int_0^1 rW_i(r)dr - \frac{1}{2}\int_0^1 W_i(r)dr\right) \\
&\stackrel{\text{def}}{=} 12\cdot m^{-1/2}\sum_{i=1}^{m}N_i
\end{aligned}
$$

Now

$$\mathrm{E}N_i = \mathrm{E}\bigg(\int_0^1 rW_i(r)dr - \frac{1}{2}\int_0^1 W_i(r)dr \bigg) = \mathrm{E}\int_0^1 rW_i(r)dr - \frac{1}{2}\mathrm{E}\int_0^1 W_i(r)dr = 0$$

$$\begin{aligned}
\mathrm{Var}N_i &= \mathrm{Var}\bigg(\int_0^1 rW_i(r)dr - \frac{1}{2}\int_0^1 W_i(r)dr \bigg) \\
&= \mathrm{E}\bigg(\int_0^1 rW_i(r)dr \bigg)^2 - \mathrm{E}\bigg(\int_0^1 rW_i(r)dr \cdot \int_0^1 W_i(r)dr \bigg) + \frac{1}{4}\mathrm{E}\bigg(\int_0^1 W_i(r)dr \bigg)^2 \\
&= \frac{2}{15} - \frac{5}{24} + \frac{1}{12} = \frac{1}{120}
\end{aligned}$$

where the calculation of a stochastic integral with Brownian motion, for example, is as follows:

$$\begin{aligned}
\mathrm{E}\bigg(\int_0^1 rW_i(r)dr \bigg)^2 &= \mathrm{E}\int_0^1 \int_0^1 rsW_i(r)W_i(s)drds \\
&= \int_0^1 \int_0^1 rs\mathrm{E}W_i(r)W_i(s)drds = \int_0^1 \int_0^1 rs\min\{r,s\}drds \\
&= \int_0^1 \bigg[\int_0^r rs^2 ds + \int_r^1 r^2 sds \bigg] dr = \int_0^1 \bigg(\frac{1}{2}r^2 - \frac{1}{6}r^4 \bigg) dr \\
&= \frac{1}{6}r^3 - \frac{1}{30}r^5 \bigg|_0^1 = \frac{2}{15}.
\end{aligned}$$

Therefore, by the Slutsky theorem we obtain the result

$$\frac{1}{\sigma}\sqrt{mn}(\hat{d} - \beta) \xrightarrow{d} N\bigg(0,\ \frac{6}{5}\bigg).$$

$$\begin{aligned}
(ii)\quad \sqrt{n}(\hat{d} - \beta) &= \frac{m^{-1}\sum_{i=1}^m n^{-5/2}\sum_{j=1}^n \{y_{i,j} - \bar{y} - \beta(j - \bar{j})\}(j - \bar{j})}{m^{-1}\sum_{i=1}^m n^{-3}\sum_{j=1}^n (j - \bar{j})^2} \\
&\overset{n\to\infty}{\Longrightarrow} 12\sigma \cdot m^{-1}N_i \xrightarrow{p} 0,\ \text{as}\ \ m \to \infty.
\end{aligned}$$

Theorem 3.4 *(Levin and Lin, 1992)Given Model M3-G and Assumption 3.2, under the null hypothesis that* $\rho = 1$*, as* $(n, m)_{\text{seq}} \to \infty$

$$n\sqrt{m}(\hat{\rho}_\tau - 1) \xrightarrow{d} N(0,\ 2) \tag{3.61}$$

$$t_\tau \xrightarrow{d} N(0,\ 1) \tag{3.62}$$

Proof of Theorem 3.4

$$\begin{aligned} n\sqrt{m}(\hat{\rho}_\tau - 1) &= \frac{m^{-1/2}n^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}(\hat{y}_{i,j} - \hat{y}_{i,j-1})\hat{y}_{i,j-1}}{m^{-1}n^{-2}\sum_{i=1}^{m}\sum_{j=1}^{n}\hat{y}_{i,j-1}^2} \\ &= \frac{m^{-1/2}n^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}(\epsilon_{i,j} - (\hat{d} - \beta))\hat{y}_{i,j-1}}{m^{-1}n^{-2}\sum_{i=1}^{m}\sum_{j=1}^{n}\hat{y}_{i,j-1}^2} \\ &= \frac{m^{-1/2}n^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}\epsilon_{i,j}\hat{y}_{i,j-1} - n^{-1/2}(\hat{d} - \beta))(mn)^{-1/2}\sum_{i=1}^{m}\sum_{j=1}^{n}\hat{y}_{i,j-1}}{m^{-1}n^{-2}\sum_{i=1}^{m}\sum_{j=1}^{n}\hat{y}_{i,j-1}^2} \end{aligned}$$

by Lemma 3.3

$$= \frac{m^{-1/2}n^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n}\epsilon_{i,j}\hat{y}_{i,j-1} - o_p(1)\cdot O_p(1)}{m^{-1}n^{-2}\sum_{i=1}^{m}\sum_{j=1}^{n}\hat{y}_{i,j-1}^2}$$

Hence, we have the same asymptotic result as in the simple AR(1) model as $(n, m)_{\text{seq}} \to \infty$.

3.3.2 Interval-specific parameter model: M3-I and M3-ID

The model can be transformed to the interval-specific mean and trend model

$$Y_{ij} - \alpha_i - \beta_i j = \rho(Y_{ij-1} - \alpha_i - \beta_i(j-1)) + \epsilon_{ij}, \tag{3.63}$$

$$Y_{ij} = \alpha_i(1-\rho) + \beta_i\rho + \beta_i(1-\rho)j + \rho Y_{ij} + \epsilon_{ij} \tag{3.64}$$

$$Y_{ij} = \mu_i + \delta_i j + \rho Y_{ij} + \epsilon_{ij}, \tag{3.65}$$

where $i = 1, \cdots, m;\ j = 1, \cdots, n;\ mn = T, \mu_i = \alpha_i(1-\rho) + \beta_i\rho,$ and $\delta_i = \beta_i(1-\rho)$.

However, we know that each interval is not independent of the others because the observations of each interval depend on those of the previous interval. As in the previous section, we will modify the model by subtracting the last value of the previous interval. So Model M3-I is given as:

$$Y_{ij} - Y_{i-1,n} = \mu_i + \delta_i j + \rho(Y_{i,j-1} - Y_{i-1,n}) + \epsilon_{ij}, \tag{3.66}$$

or equivalently,

$$y_{ij} = \mu_i + \delta_i j + \rho y_{i,j-1} + \epsilon_{ij}. \tag{3.67}$$

Note that the original and modified models are equivalent only when $\rho = 1$, for then we are simply subtracting $Y_{i-1,n}$ from both sides of Equation (3.64).

For this model, the estimator of ρ can be obtained in two stages. We first detrend using the interval-specific mean and trend for each interval, i.e., $\bar{y}_i$ and $\hat{d}_i$, and then we obtain the POLS estimate of ρ in Stage 2.

Stage 1: $\hat{y}_{ij} = y_{ij} - \bar{y}_i - \hat{d}_i(j - \bar{j}), \qquad i = 1, \cdots, m;\ j = 1, \cdots, n$

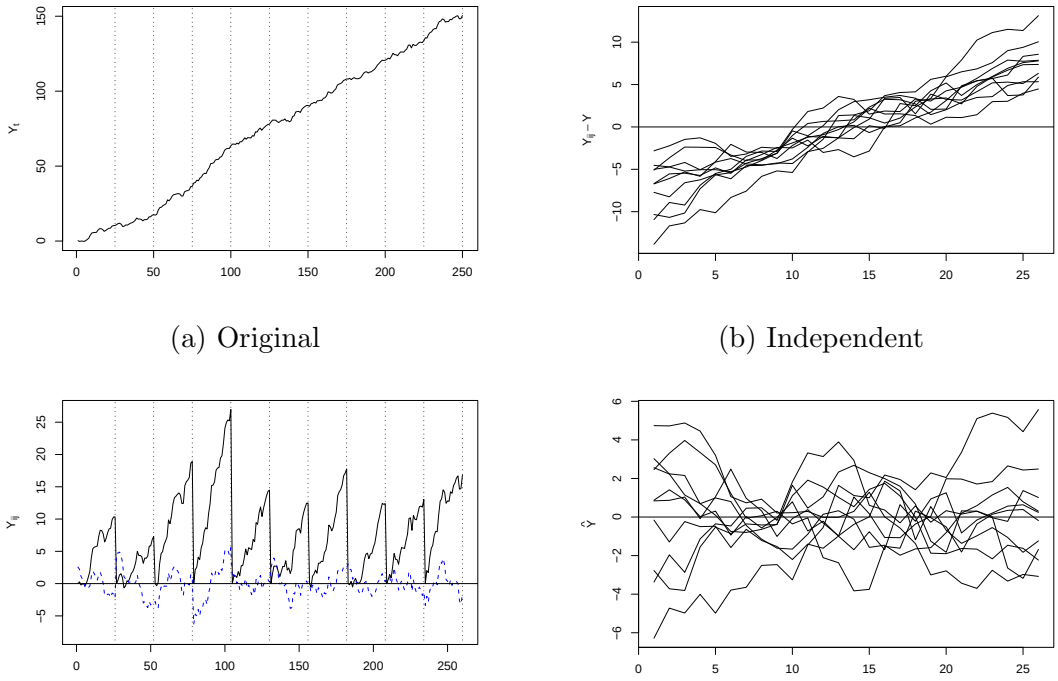
Stage 2: $\hat{y}_{ij} = \rho\hat{y}_{ij-1} + e_{ij}$

From stage 2, we have the pooled least squares estimator of ρ given by

$$\hat{\rho}_{\tau_i} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\hat{y}_{ij}\hat{y}_{ij-1}}{\sum_{i=1}^{m}\sum_{j=1}^{n}\hat{y}_{ij-1}^2},$$

under $H_0 : \rho = 1$,

$$= \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(\hat{y}_{ij-1} + e_{ij})\hat{y}_{ij-1}}{\sum_{i=1}^{m}\sum_{j=1}^{n}\hat{y}_{ij-1}^2} = 1 + \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\hat{y}_{ij-1}e_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n}\hat{y}_{ij-1}^2}$$



(a) Original

(b) Independent

(c) Interval Mean

(d) $\hat{y}_I$

Figure 3.4: Transformation of Linear Trend AR (1) Model: M3-I

which is also biased because of interval-specific effects. Again, by the delta method, we obtain

$$\hat{\rho}_{\tau_i} \approx N\left(1 - \frac{7.5}{n+1}, \frac{25.85}{mn^2}\right)$$

The following properties guarantee that the POLS estimate of ρ is consistent both under the hypotheses.

- $y_{ij} - \overline{y}_i = Y_{ij} - \overline{Y}_i$
- $\hat{d}_i = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij}-\overline{y}_i)(j-\bar{j})}{\sum_{i=1}^{m}\sum_{j=1}^{n}(j-\bar{j})^2} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij}-\overline{Y}_i)(j-\bar{j})}{\sum_{i=1}^{m}\sum_{j=1}^{n}(j-\bar{j})^2}$
- $\hat{y}_{ij} = y_{ij} - \overline{y}_i - \hat{d}_i(j-\bar{j}) = Y_{ij} - \overline{Y}_i - \hat{d}_i(j-\bar{j}) = \hat{Y}_{ij}$

Table 3.4: Moments of the Linear Trend AR (1) Model (Dickey, 1976)

$\mathrm{E}\{N_i\} = -\frac{(n-3)}{2}\sigma^2$	$\mathrm{E}\{N_{in}\} = -\frac{1}{2} + o(\frac{1}{n})$
$\mathrm{E}\{{N_i}^2\} = \frac{(n-3)(8n-7)}{30}\sigma^4$	$\mathrm{E}\{N_{in}^2\} = \frac{4}{15} + o(\frac{1}{n})$
$\mathrm{Var}\{N_i\} = \frac{(n-3)(n+31)}{60}\sigma^4$	$\mathrm{Var}\{N_{in}\} = \frac{1}{60} + o(\frac{1}{n})$
$\mathrm{E}\{D_i\} = \frac{(n+1)(n-3)}{15}\sigma^2$	$\mathrm{E}\{D_{in}\} = \frac{1}{15} + o(\frac{1}{n})$
$\mathrm{E}\{{D_i}^2\} = \frac{(n+1)(n-3)(13n^2-26n+36)}{2100}\sigma^4$	$\mathrm{E}\{{D_{in}}^2\} = \frac{1}{2100} + o(\frac{1}{n})$
$\mathrm{Var}\{D_i\} = \frac{(n+1)(n-3)(11n^2-22n+192)}{6300}\sigma^4$	$\mathrm{Var}\{D_{in}\} = \frac{11}{6300} + o(\frac{1}{n})$
$\mathrm{Cov}(N_i, D_i) = -\frac{(n+1)(n-3)}{15}\sigma^4$	$\mathrm{Cov}(N_{in}, D_{in}) = o(\frac{1}{n})$

- $\hat{\rho}_{\tau_i} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\hat{y}_{ij}\hat{y}_{i,j-1}}{\sum_{i=1}^{m}\sum_{j=1}^{n}\hat{y}_{i,j-1}^2} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\hat{Y}_{ij}\hat{Y}_{ij-1}}{\sum_{i=1}^{m}\sum_{j=1}^{n}\hat{Y}_{ij-1}^2}$

Define

$$N_i = \sum_{j=1}^{n}\hat{y}_{ij-1}e_{ij} \quad \text{and} \quad D_i = \sum_{j=1}^{n}\hat{y}_{ij-1}^2,$$

or

$$N_{in} = \frac{1}{n\sigma^2}\sum_{j=1}^{n}\hat{y}_{ij-1}e_{ij} \quad \text{and} \quad D_{in} = \frac{1}{n^2\sigma^2}\sum_{j=1}^{n}\hat{y}_{ij-1}^2,$$

Recall the moments. Using the delta method, we obtain that

$$\mathrm{E}(\hat{\rho}_{\tau_i} - 1) \approx \frac{\mathrm{E}(N_i)}{\mathrm{E}(D_i)} \approx -\frac{7.5}{n+1}, \tag{3.68}$$

and

$$
\begin{aligned}
&\mathrm{Var}(\hat{\rho}_{\tau_i} - 1) \approx \\
&\left(\begin{array}{cc} \frac{1}{\mathrm{E}(D_i)} & -\frac{\mathrm{E}(N_i)}{\mathrm{E}^2(D_i)} \end{array} \right) \left(\begin{array}{cc} \frac{1}{mn^2}\mathrm{Var}(N_i) & \frac{1}{mn^2}\mathrm{Cov}(N_i, D_i) \\ \frac{1}{mn^2}\mathrm{Cov}(N_i, D_i) & \frac{1}{mn^2}\mathrm{Var}(D_i) \end{array} \right) \left(\begin{array}{c} \frac{1}{\mathrm{E}(D_i)} \\ -\frac{\mathrm{E}(N_i)}{\mathrm{E}^2(D_i)} \end{array} \right) \\
&\approx \left(\begin{array}{cc} \frac{1}{\mathrm{E}(D_{in})} & -\frac{\mathrm{E}(N_{in})}{\mathrm{E}^2(D_{in})} \end{array} \right) \left(\begin{array}{cc} \frac{1}{m}\mathrm{Var}(N_{in}) & \frac{1}{m}\mathrm{Cov}(N_{in}, D_{in}) \\ \frac{1}{m}\mathrm{Cov}(N_{in}, D_{in}) & \frac{1}{m}\mathrm{Var}(D_{in}) \end{array} \right) \left(\begin{array}{c} \frac{1}{\mathrm{E}(D_{in})} \\ -\frac{\mathrm{E}(N_{in})}{\mathrm{E}^2(D_{in})} \end{array} \right) \\
&= \left(\begin{array}{cc} 15 & \frac{15^2}{2} \end{array} \right) \left(\begin{array}{cc} \frac{1}{60mn^2} & 0 \\ 0 & \frac{11}{6300mn^2} \end{array} \right) \left(\begin{array}{c} 15 \\ \frac{15^2}{2} \end{array} \right) = \frac{2895}{112mn^2}
\end{aligned} \tag{3.69}
$$

For the t-statistic, we have similar results. Using the delta method, we obtain that

$$
\mathrm{E}(t_{\tau_i}) \approx \frac{\mathrm{E}(N_i)}{\sqrt{\mathrm{E}(D_i)}} = -\frac{\sqrt{15}}{2n} \tag{3.70}
$$

$$
\begin{aligned}
&\mathrm{Var}(t_\mu) \approx \\
&\left(\begin{array}{cc} \frac{1}{\sqrt{\mathrm{E}(D_i)}} & -\frac{1}{2}\frac{\mathrm{E}(N_i)}{\mathrm{E}(D_i)^{3/2}} \end{array} \right) \left(\begin{array}{cc} \frac{1}{mn^2}V(N_i) & \frac{1}{mn^2}C(N_i, D_i) \\ \frac{1}{mn^2}C(N_i, D_i) & \frac{1}{mn^2}V(D_i) \end{array} \right) \left(\begin{array}{c} \frac{1}{\sqrt{\mathrm{E}(D_i)}} \\ -\frac{1}{2}\frac{\mathrm{E}(N_i)}{\mathrm{E}(D_i)^{3/2}} \end{array} \right) \\
&\approx \left(\begin{array}{cc} \frac{1}{\sqrt{\mathrm{E}(D_{in})}} & -\frac{1}{2}\frac{\mathrm{E}(N_{in})}{\mathrm{E}(D_{in})^{3/2}} \end{array} \right) \left(\begin{array}{cc} \frac{1}{m}V(N_{in}) & \frac{1}{m}C(N_{in}, D_{in}) \\ \frac{1}{m}C(N_{in}, D_{in}) & \frac{1}{m}V(D_{in}) \end{array} \right) \left(\begin{array}{c} \frac{1}{\sqrt{\mathrm{E}(D_{in})}} \\ -\frac{1}{2}\frac{\mathrm{E}(N_{in})}{\mathrm{E}(D_{in})^{3/2}} \end{array} \right) \\
&= \left(\begin{array}{cc} \sqrt{15} & \frac{1}{4}15^{3/2} \end{array} \right) \left(\begin{array}{cc} \frac{1}{60} & 0 \\ 0 & \frac{11}{6300} \end{array} \right) \left(\begin{array}{c} \sqrt{15} \\ \frac{1}{4}15^{3/2} \end{array} \right) = \frac{277}{448}
\end{aligned}
$$

Theorem 3.5 (Levin and Lin, 1992): *Given Model M3-I and Assumption 3.2, under the null hypothesis that* $\rho = 1$, *as* $(n, m)_{\text{seq}} \to \infty$

$$
\begin{aligned}
&a) \quad \sqrt{m}n(\hat{\rho}_{\tau_i} - 1) - \frac{\sqrt{m}EN_n}{ED_n} \xrightarrow{d} N\left(0, \frac{2895}{112}\right) \\
&b) \quad \textit{If } \frac{\sqrt{m}}{n} \to 0, \textit{ then } \sqrt{m}n(\hat{\rho}_{\tau_i} - 1) + 7.5\sqrt{m} \xrightarrow{d} N\left(0, \frac{2895}{112}\right) \\
&c) \quad \sqrt{\frac{448}{277}}\left\{t_\tau - \frac{\sqrt{m}EN_n}{\sqrt{ED_n}}\right\} \xrightarrow{d} N(0,1) \\
&d) \quad \textit{If } \frac{\sqrt{m}}{n} \to 0, \textit{ then } \sqrt{\frac{448}{277}}\left(t_\tau + \sqrt{3.75m}\right) \xrightarrow{d} N(0,1).
\end{aligned}
$$

The proof of Theorem 3.5 is similar to that of Theorem 3.3, and we correct the variance in (a) and (b).

3.4 Simulation Results

In this section we perform simulation work to understand the finite sample behavior of our transformed models so that we mainly focus on calculating the critical value and empirical power for each model. To this end, we consider regression models, M1, M2-G, M2-I, M3-G, M3-GD, M3-I, and M3-ID developed in the previous sections.

The total sample sizes are chosen as $T = 100, 250, 500$ and 2500 which are also used in Fuller (1996). So we easily check how our approach is different from the conventional Dickey-Fuller test. We generate the data set with $Y_0 = 0$ and $\epsilon_t \overset{iid}{\sim} N(0,\ 1)$.

In the independent data transformation, we are required to divide the original univariate series into m intervals with size n, say $T = mn$. For our study, we choose m and T as follows:

Table 3.5: Simulation scheme: Selection of m and n from T

T	100	250	500	2500
m	1	1	1	1
	5	5	5	5
	10	10	10	10
	20	25	25	25
		50	50	50
			100	100
				250
				500

For each model, the number of replications is at least 5000, and the standard error of simulation is less than 0.014 in most cases.

First, under the null hypothesis that $\rho = 1$ we check whether or not the normality of the t-statistic holds (approximately) in the independent data transformation. As we have shown in the previous sections, the simple random walk model (M1) and global parameter models (M2-G, M3-G, and M3-GD) asymptotically have the standard normal distribution under the null hypothesis while the interval-specific parameter models such as M2-I, M3-I, and M3-ID have a nonstandard normal distribution due to the biased estimator of ρ. Through Figure 7.2 - 7.6, we see that the simulation results nicely support the normality of our approach even in the small sample ($T = 100$) case. Each figure has a smoothed plot of the simulated t-statistics and the plot of the standard normal distribution where we drew the mean-adjusted t-statistics for M2-G, M2-I, M3-G, and M3-I because the parameters in the deterministic term lead to noncentrality. For example, in Model M2-G (Figure 7.3) we adjusted by $1.52, 0.55, 0.35$, and 0.22 for $m = 1, 5, 10$, and 20, respectively. For fixed T, the adjustment value for noncentrality becomes smaller in the global parameter model

but larger in the interval-specific parameter models as m becomes large. These phenomena are also explained by the biased estimator of ρ and σ^2 in the interval-specific parameter models while the global parameter models asymptotically have no bias. Especially, the magnitude of the bias in the interval-specific models depends only on the size of each interval since the bias of $\widehat{\rho}_{\mu_i}$ and $\widehat{\rho}_{\tau_i}$ is $O_p(n^{-1})$, which implies that for fixed T, increasing m (decreasing n) results in increasing bias and noncentrality.

After the mean adjustment, the patterns of the global parameter models are quite close to M1 and the standard normal distribution excepting for $m = 1$, which suggests that the testing power of M2-G and M3-G has a similar pattern to M1. However, the interval-specific parameter models have different patterns from each other, and the smoothed plot of t-statistics gets farther from the standard normal distribution as the number of incidental parameters increase.

Second, we investigate the simulated critical values of each model under the null hypothesis. Through Table $7.1 - 7.5$ we provide the $1\%, 5\%$, and 10% quantiles for ρ- and/or t-statistics which are defined in the previous sections of Chapter 3. For M1, M2-G, M3-G, and M3-GD, ρ-statistics are $N(0,\ 2)$ and t-statistics are $N(0,\ 1)$ asymptotically. On the other hand, M2-I, M3-I, and M3-ID are asymptotically non-standard normal. Although the simulated critical values in the smaller finite samples do not achieve their asymptotic values, the simulation results show that the quantiles quickly approach those values as m and n increase. As can be seen in the tables, for fixed T, as m gets large the quantiles are close to the limiting quantiles. We can also check that as n gets large the quantile values approach the limiting values for fixed m. However, if $\sqrt{m}/n$ is not close to zero, then the bias-corrected t-statistic

overcorrects and the critical values are located to the right of those of the limiting distribution. For example, for the combination of $(m, n) = (500, 5)$, the bias-correted 10% simulated critical value is 0.14, which is quite far from -1.28.

For fixed T, the quantiles of both $\rho-$ and $t-$statistics drastically vary as m changes from 1 to 5. For example, in Table 7.1 the 5% quantile of the ρ-statistic changes from -13.39 to -5.49 when m increases from 1 to 5. This is because the size of each interval, n, decreases relatively more in the change of m from 1 to 5 than it does in the other cases.

Next, we illustrate the rejection probability (or empirical power) of tests for the basic models without structural changes through Table $7.6-7.8$ and Figure $7.7-7.10$. In the tables we provide the empirical power for the values of alternative hypothesis, $\rho = 0.8, 0.9$, and 0.95 and in the plots we draw the power curve for the various values of ρ such as $0.8, 0.85, 0.90, 0.925, 0.95, 0.975, 0.98$, and 0.99 for each m.

The main result from the simulation is that for fixed T increasing m seriously reduces the power of the test, which verifies that the power of such a test highly depends on the number of observations in each interval. In Table 7.6, for given $T = 250$ and $\rho = 0.9$, we can see the severe reduction of the power from 99% to 11%, at size 0.01, as m increases. However, as the total sample size, T, increases the power of the tests approach that of the original Dickey-Fuller test and ultimately to 1 since the size of each interval, n, becomes large for fixed m. That is, the tests are consistent.

Another important feature is that the interval-specific parameter models, M2-I

and M3-I, have a more serious decrease in power than the global parameter models. We report the empirical power of these models in Table 7.9 and 7.10 for $\rho = 0.8, 0.9,$ and 0.95.

In section 3.1.3, we developed the properties of $\hat{\rho}$ under the alternative hypothesis. As we have shown, because of the last value subtraction $\hat{\rho}$ is in general an inconsistent estimator of ρ. It converges to $\frac{1}{2}(1+\rho)$ and thus is consistent only under the null hypothesis. In our framework, the case of $m = 1$ produces the conventional Dickey-Fuller test in which the estimate of ρ is consistent for all ρ. Hence, it is reasonable that we compare the power at a value ρ in cases of $m = 5, 10$ and 20 to the power at $\frac{1}{2}(1+\rho)$ in the case of $m = 1$. From investigation in Table 7.6, for $m = 5$ the power of the transformed model is quite close to that of the Dickey-Fuller test.

3.5 Extension to an AR (p) process

Up to now, we focused on the AR (1) model, which allows the independent transformation. In this section, we extend our methodology to a general AR (p) process that makes sence since time series data in the real world have various types of ARMA structure. We first consider the model

$$Y_t = d_t + \rho Y_{t-1} + Z_t, \qquad t = \cdots, -1, 0, 1 \cdots, \tag{3.71}$$

$$Z_t = \alpha_1 Z_{t-1} + \alpha_2 Z_{t-2} + \cdots + \alpha_p Z_{t-p} + \epsilon_t \tag{3.72}$$

where d_t is the deterministic part, Z_t is a stationary series with $\mathrm{E}(Z_t) = 0$ and $\sum_{-\infty}^{\infty} \gamma_z(h) < \infty$, and $\epsilon_t \overset{iid}{\sim} (0, \sigma^2)$ follows Assumption 3.1. For simplicity, we assume that $d_t = 0$. Under the null hypothesis that $\rho = 1$, $Y_t = \sum_{k=1}^{t} Z_k$ and $\Delta Y_t =$

$Y_t - Y_{t-1} = Z_t$. Note that for the simple AR(1) process with a unit root, $Y_t = \sum_{k=1}^{t} \epsilon_k$ and $\Delta Y_t = \epsilon_t$.

We can rewrite (3.71) as

$$\begin{aligned} \Delta Y_t &= (\rho - 1)Y_{t-1} + Z_t \\ &= (\rho - 1)Y_{t-1} + \alpha_1 Z_{t-1} + \alpha_2 Z_{t-2} + \cdots + \alpha_p Z_{t-p} + \epsilon_t \\ &= (\rho - 1)Y_{t-1} + \alpha_1 \Delta Y_{t-1} + \alpha_2 \Delta Y_{t-2} + \cdots + \alpha_p \Delta Y_{t-p} + \epsilon_t \end{aligned} \tag{3.73}$$

As in the AR(1) model, under the null hypothesis we make the double-indexing and last value subtraction, sequentially.

Double-indexing: $Y_t \mapsto Y_{i,j}$ with $t = (i-1)n + j$:

$$\begin{array}{cccc} Y_{1,1} & Y_{1,2} & \cdots & Y_{1,n} \\ Y_{2,1} & Y_{2,2} & \cdots & Y_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{m,1} & Y_{m,2} & \cdots & Y_{m,n} \end{array} \overset{\rho=1}{\Longleftrightarrow} \begin{array}{cccc} Z_1 & Z_1 + Z_2 & \cdots & \sum_{k=1}^{n} Z_k \\ \sum_{k=1}^{n+1} Z_k & \sum_{k=1}^{n+2} Z_k & \cdots & \sum_{k=1}^{2n} Z_k \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^{(m-1)n+1} Z_k & \sum_{k=1}^{(m-1)n+2} Z_k & \cdots & \sum_{k=1}^{mn} Z_k \end{array} \tag{3.74}$$

Last value subtraction: $Y_{i,j} - Y_{i-1,n} \overset{\text{say}}{=} y_{i,j}$:

$$\begin{array}{cccc} y_{1,1} & y_{1,2} & \cdots & y_{1,n} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m,1} & y_{m,2} & \cdots & y_{m,n} \end{array} \Longleftrightarrow \begin{array}{cccc} Z_{1,1} & Z_{1,1} + Z_{1,2} & \cdots & \sum_{k=1}^{n} Z_{1,k} \\ Z_{2,1} & Z_{2,1} + Z_{2,2} & \cdots & \sum_{k=1}^{n} Z_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{m,1} & Z_{m,1} + Z_{m,2} & \cdots & \sum_{k=1}^{n} Z_{m,k} \end{array} \tag{3.75}$$

where

$$Z_{i,j} = \begin{cases} \sum_{k=1}^{j} Z_k & \text{for } \ i = 1 \\ \sum_{k=1}^{(i-1)n+j} Z_k - \sum_{k=1}^{(i-1)n} Z_k & \text{for } \ i = 2, \cdots, m \end{cases}.$$

Unlike the AR(1) process, a general ARMA(p,q) series has no independent double-indexing transformation under the null hypothesis that $\rho = 1$. However, we can see that the covariance between any two intervals after the last value subtraction decays

more quickly than that of within an interval and that of the simply double indexing process, which has the same probabilistic properties as the Y_t. Now we check the covariance between two intervals where we change the indices in order to check the effect of n.

$$\text{Interval } i: \ Z_{in+1}, \ Z_{in+1} + Z_{in+2}, \ \cdots, Z_{in+1} + \cdots + Z_{(i+1)n}$$

$$\text{Interval } i': \ Z_{i'n+1}, \ Z_{i'n+1} + Z_{i'n+2}, \ \cdots, Z_{i'n+1} + \cdots + Z_{(i'+1)n}$$

For some $0 < M < \infty$ and $0 < \lambda < 1$,

$$\begin{aligned}
&|\text{Cov}(Z_{in+1} + \cdots + Z_{in+k}, Z_{i'n+1} + \cdots + Z_{i'n+k'})| \\
&\qquad = |\sum_{t=1}^{k}\sum_{s=1}^{k'} \gamma_z\{(i-i')n + t - s\}| \\
&\qquad \leq \sum_{t=1}^{k}\sum_{s=1}^{k'} |\gamma_z\{(i-i')n + t - s\}| \\
&\qquad \leq \sum_{t=1}^{n}\sum_{s=1}^{n} |\gamma_z\{(i-i')n + t - s\}| \\
&\qquad \leq \sum_{t=1}^{n}\sum_{s=1}^{n} |\gamma_z\{(i-i'-1)n + 1\}| \\
&\qquad = n^2 |\gamma_z\{(i-i'-1)n + 1\}| \\
&\qquad \leq M\lambda^{|(i-i'-1)n+1|}
\end{aligned}$$

where the last inequality holds since $|\gamma(h)| \leq M\lambda^{|h|}$. If $i = i'$, then the result is the simple covariance within the interval. We see that the covariance within interval dominates that between intervals and both are absolutely summable, which implies that we can apply the CLT for dependent variable to this process.

In order to estimate the coefficients in the AR (p) model given in (3.75), we follow the methods in Said and Dickey (1984, 1985). Now define:

1. $\alpha_0 = \rho - 1$

2. $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \cdots, \alpha_p)'$

3. $X_{i,j} = (Z_{i,j-1}, \cdots, Z_{i,j-p}) = (\Delta y_{i,j-1}, \cdots, \Delta y_{i,j-p})$

4. $U_{i,j} = (y_{i,j-1}; X_{i,j})$

Then the regression model in (3.75) can be rewritten as

$$y_{i,j} = U_{i,j}\boldsymbol{\alpha} + \epsilon_{i,j}. \tag{3.76}$$

1. $\mathbf{y}_{i,0} = (y_{i,p+1}, \cdots, y_{i,n})$ and $\mathbf{y}_{i,-1} = (y_{i,p}, \cdots, y_{i,n-1})$

2. $\mathbf{X}_i = (\mathbf{X}'_{i,p+1}, \cdots, \mathbf{X}'_{i,n})'$

3. $\mathbf{U}_i = (\mathbf{U}'_{i,p+1}, \cdots, \mathbf{U}'_{i,n})' = (\mathbf{y}_{i,-1};\ \mathbf{X}_i)$

4. $\boldsymbol{\epsilon}_i = (\epsilon_{i,p+1}, \cdots, \epsilon_{i,n})'$

Using these vector notations, for $i = 1, \cdots, m$, the regression model is equivalently given as

$$\mathbf{y}_{i,0} = \mathbf{U}_i\boldsymbol{\alpha} + \boldsymbol{\epsilon}_i. \tag{3.77}$$

The POLS estimate of $\boldsymbol{\alpha}$ is given by

$$\widehat{\boldsymbol{\alpha}} = \left[\sum_{i=1}^{m} \mathbf{U}'_i\mathbf{U}_i\right]^{-1} \sum_{i=1}^{m} \mathbf{U}'_i\mathbf{y}_{i,0}$$

and

$$\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = \left[\sum_{i=1}^{m} \mathbf{U}'_i\mathbf{U}_i\right]^{-1} \sum_{i=1}^{m} \mathbf{U}'_i\boldsymbol{\epsilon}_i.$$

Define

$$D_{mn} = \sum_{i=1}^{m} \mathbf{U}_i'\mathbf{U}_i = \sum_{i=1}^{m} \begin{bmatrix} \mathbf{y}_{i,-1}'\mathbf{y}_{i,-1} & \mathbf{y}_{i,-1}'\mathbf{X}_i \\ \mathbf{X}_i'\mathbf{y}_{i,-1} & \mathbf{X}_i'\mathbf{X}_i \end{bmatrix} \tag{3.78}$$

$$= \sum_{i=1}^{m}\sum_{j=p+1}^{n} U_{i,j}U_{i,j}' = \sum_{i=1}^{m} \begin{bmatrix} \sum_{j=p+1} y_{i,j-1}^2 & \sum_{j=p+1} y_{i,j-1}X_{i,j} \\ \sum_{j=p+1} y_{i,j-1}X_{i,j}' & \sum_{j=p+1} X_{i,j}X_{i,j}' \end{bmatrix},$$

$$N_{mn} = \sum_{i=1}^{m} \mathbf{U}_i'\boldsymbol{\epsilon}_i = \sum_{i=1}^{m} \begin{bmatrix} \mathbf{y}_{i,-1}'\boldsymbol{\epsilon}_i \\ \mathbf{X}_i'\boldsymbol{\epsilon}_i \end{bmatrix} \tag{3.79}$$

$$= \sum_{i=1}^{m}\sum_{j=p+1}^{n} U_{i,j}\epsilon_{i,j} = \sum_{i=1}^{m} \begin{bmatrix} \sum_{j=p+1}^{n} y_{i,j-1}\epsilon_{i,j} \\ \sum_{j=p+1}^{n} X_{i,j}'\epsilon_{i,j} \end{bmatrix},$$

and

$$R_{mn} = \text{diag}\{m^{-1/2}(n-p)^{-1}, \underbrace{(m(n-p))^{-1/2}, \cdots, (m(n-p))^{-1/2}}_{\text{p terms}}\} \tag{3.80}$$

where R_{mn} can be decomposed into two normalizing matrices R_m and R_n such as $R_{mn} = R_m R_n$ where $R_m = \text{diag}\{m^{-1/2}, \cdots, m^{-1/2}\} = m^{-1/2}I_{p+1}$ and $R_n = \text{diag}\{R_{n1}, R_{n2}\} \overset{\text{say}}{=} \text{diag}\{(n-p)^{-1}, (n-p)^{-1/2}I_p\}$. This decomposition is very useful when we apply the sequential limit to achieve asymptotic results since we can easily identify the effects of m and n. The next step is to develop the limiting distribution of the coefficient estimates. Using the above definitions, we have the following form:

$$\begin{aligned} R_{mn}^{-1}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) &= (R_{mn}D_{mn}R_{mn})^{-1}R_{mn}N_{mn} \\ &= [R_m(R_nD_{mn}R_n)R_m]^{-1}R_m(R_nN_{mn}) \\ &= [R_m^{-1}(R_nD_{mn}R_n)^{-1}R_m^{-1}]R_m(R_nN_{mn}). \end{aligned} \tag{3.81}$$

Applying the sequential limit to Equation (3.81), we first try to find the intermediate asymptotic results as $n \to \infty$. To this end, we consider two parts, $(R_nD_{mn}R_n)$ and R_nN_{mn}, separately.

$$
\begin{aligned}
R_n D_{mn} R_n \quad = \quad & \\
& \sum_{i=1}^{m} \begin{bmatrix}
\frac{1}{(n-p)^2}\sum_{j=p+1}^{n} y_{i,j-1}^2 & \frac{1}{(n-p)^{3/2}}\sum_{j=p+1}^{n} y_{i,j-1}Z_{i,j-1} \\
\frac{1}{(n-p)^{3/2}}\sum_{j=p+1}^{n} y_{i,j-1}Z_{i,j-1} & \frac{1}{(n-p)}\sum_{j=p+1}^{n} Z_{i,j-1}^2 \\
\vdots & \vdots \\
\frac{1}{(n-p)^{3/2}}\sum_{j=p+1}^{n} y_{i,j-1}Z_{i,j-p} & \frac{1}{(n-p)}\sum_{j=p+1}^{n} Z_{i,j-1}Z_{i,j-p}
\end{bmatrix} \\
& \qquad \begin{bmatrix}
\cdots & \frac{1}{(n-p)^{3/2}}\sum_{j=p+1}^{n} y_{i,j-1}Z_{i,j-p} \\
\cdots & \frac{1}{(n-p)}\sum_{j=p+1}^{n} Z_{i,j-1}Z_{i,j-p} \\
\ddots & \vdots \\
\cdots & \frac{1}{(n-p)}\sum_{j=p+1}^{n} Z_{i,j-p}^2
\end{bmatrix} \overset{\text{def}}{=} \sum_{i=1}^{m} \begin{bmatrix} A_{in} & B_{in} \\ B_{in}' & C_{in} \end{bmatrix}
\end{aligned}
\tag{3.82}
$$

where B_{in} is the $1 \times p$ vector and C_{in} is the $p \times p$ matrix. Using that $y_{i,j} = \sum_{s=1}^{j} Z_{i,s}$ and $\Delta y_{i,j} = Z_{i,j}$, we see that for $j = p+1, \cdots, n$,

$$
\begin{aligned}
\big(1 - \textstyle\sum_{k=1}^{p} \alpha_k\big) y_{i,j} \quad & = \epsilon_{i,p+1} + \epsilon_{i,p+2} + \cdots + \epsilon_{i,j} \\
& \quad + \textstyle\sum_{l=1}^{p} \big(1 - \sum_{k=1}^{p-l} \alpha_k\big) Z_{i,l} + \sum_{l=1}^{p} \big(\sum_{k=l}^{p} \alpha_k\big) Z_{i,j-l} \\
& = \omega_{i,j} + O_p(1)
\end{aligned}
\tag{3.83}
$$

where $\omega_{i,j} = \sum_{s=p+1}^{j} \epsilon_{i,s}$ are independent across i. Hence,

$$
\begin{aligned}
A_{in} &= \frac{1}{(n-p)^2} \sum_{j=p+1}^{n} y_{i,j-1}^2 \\
&= \frac{1}{(1 - \sum_{k=1}^{p} \alpha_k)^2 (n-p)^2} \sum_{j=p+1}^{n} \omega_{i,j}^2 + o_p(n^{-1}) \\
&\Longrightarrow \frac{\sigma^2}{(1 - \sum_{k=1}^{p} \alpha_k)^2} \int W_i^2(r) dr \overset{\text{def}}{=} A_i,
\end{aligned}
\tag{3.84}
$$

where W_i are independent across i. Let B_l be an element of B for $l = 1, 2, \cdots, p$.

Then

$$B_l = \frac{1}{(n-p)^{3/2}} y_{i,j-1} Z_{i,j-l}$$
$$= \frac{1}{(n-p)^{1/2}} \left(\frac{1}{(n-p)} \sum_{j=p+1}^{n} y_{i,j-1} Z_{i,j-l} \right) \qquad \xrightarrow{p} 0, \qquad \text{for } l = 1, \cdots, p,$$

where the last property holds since for $l = 1, \cdots, p$,

$$\frac{1}{n-p} \sum_{j=p+1}^{n} y_{i,j-1} Z_{i,j-l} \overset{n\to\infty}{\Longrightarrow} \frac{1}{2} \{ \frac{\sigma^2}{(1 - \sum_{k=1}^{p} \alpha_k)^2} W_i^2(1) - \gamma_z(0) \} + \gamma_z(0) + \cdots + \gamma_z(l-1)$$
$$= O_p(1).$$

Thus, we obtain that

$$B_{in} \xrightarrow{p} \mathbf{0}_{1\times p}. \tag{3.85}$$

Since $Z_{i,j}$ is a stationary series, as $n \to \infty$ we see the following result

$$C_{in} \xrightarrow{p} \Gamma = \begin{bmatrix} \gamma_z(0) & \gamma_z(1) & \cdots & \gamma_z(p-1) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_z(p-1) & \gamma_z(p-2) & \cdots & \gamma_z(0) \end{bmatrix} \tag{3.86}$$

In order to achieve the asymptotic result for a general ARMA series in our framework, the following properties are important:

$$A_{in} \overset{n\to\infty}{\Longrightarrow} f(W_i) \text{ independent across } i \text{ and } \; B_{in} \overset{n\to\infty}{\longrightarrow} \mathbf{0}$$

where A_{in} and B_{in} are matrices if the model has the deterministic part such as intercept and/or trend, and $f(W_i)$ denote a functional of the standard Wiener process. We can achieve the same asymptotic results as the ADF for the intercept and/or linear trend models with modifying $A_{i,n}$ and B_{in} in Equation (3.82).

For details, refer to Theorem 10.1.2 in Fuller (1996, p.556) or Section 17.7 in Hamilton (1994). Combining the results in (3.84), (3.85), and (3.86), we obtain the intermediate asymptotic result for $R_n D_{mn} R_n$ as $n \to \infty$:

$$\begin{aligned} R_{mn} D_{mn} R_{mn} &= R_m (R_n D_{mn} R_n) R_m \\ &\overset{n\to\infty}{\longrightarrow} R_m D_m R_m, \end{aligned} \tag{3.87}$$

where

$$D_m = \sum_{i=1}^{m} \begin{bmatrix} A_i & \mathbf{0} \\ \mathbf{0}' & \Gamma \end{bmatrix}.$$

Now we consider the second part in the sequential limit as $n \to \infty$.

$$R_n N_{mn} = \sum_{i=1}^{m} \sum_{j=p+1}^{n} U_{i,j} \epsilon_{i,j} = \sum_{i=1}^{m} \begin{bmatrix} R_{n1} \sum_{j=p+1}^{n} y_{i,j-1} \epsilon_{i,j} \\ R_{n2} \sum_{j=p+1}^{n} X'_{i,j} \epsilon_{i,j} \end{bmatrix}$$

$$\begin{aligned} R_{n1} \sum_{j=p+1}^{n} y_{i,j-1} \epsilon_{i,j} &= \frac{1}{n(1 - \sum_{k=1}^{p} \alpha_k)} \sum_{j=p+1}^{n} \omega_{i,j-1} \epsilon_{i,j} + o_p(1) \\ &\overset{n\to\infty}{\Longrightarrow} \frac{\sigma^2}{1(1 - \sum_{k=1}^{p} \alpha_k)} \{W_i^2(1) - 1\} \overset{\text{def}}{=} \xi_i. \end{aligned} \tag{3.88}$$

$$R_{n2} \sum_{j=p+1}^{n} X'_{i,j} \epsilon_{i,j} \xrightarrow{d} \Phi_i, \qquad \Phi_i \sim N_p(\mathbf{0},\ \sigma^2 \Gamma). \tag{3.89}$$

Hence,

$$R_n N_{mn} \overset{n\to\infty}{\Longrightarrow} N_m, \tag{3.90}$$

where $N_m = (\xi_i,\ \Phi'_i)'$.

Combining (3.87) and (3.90), we have the intermediate asymptotic result:

$$\begin{aligned} R_{mn}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) &= (R_{mn}D_{mn}R_{mn})^{-1}R_{mn}N_{mn} \\ &= [R_m^{-1}(R_nD_{mn}R_n)^{-1}R_m^{-1}]R_m(R_nN_{mn}) \\ &\overset{n\to\infty}{\Longrightarrow} (R_mD_mR_m)^{-1}R_mN_m \\ &= (R_m^2D_m)^{-1}R_mN_m \quad (\because R_m = m^{-1/2}I_{p+1}). \end{aligned} \tag{3.91}$$

Each element in R_m and N_m is identically distributed across i, and a well defined random variable in the sense that it has at least the finite first two moments. In order to obtain the final asymptotic result as $m \to \infty$, we consider two parts, $(R_m^2D_m)^{-1}$ and R_mN_m, separately. For the first part, as $m \to \infty$, by WLLN,

$$R_m^2D_m = \begin{bmatrix} \frac{1}{m}\sum_{i=1}^m A_i & \mathbf{0} \\ \mathbf{0}' & \Gamma \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \mathrm{E}(A_i) & \mathbf{0} \\ \mathbf{0}' & \Gamma \end{bmatrix} \overset{\text{def}}{=} D \tag{3.92}$$

where $\mathrm{E}(A_i) = \frac{1}{2(1-\sum_{k=1}^p \alpha_k)^2}$. Since D_m is a nonsingular matrix, there exists D^{-1}. By the CLT for dependent processes (Davidson, 1994, Ch. 24),

$$R_mN_m = \left(\frac{1}{\sqrt{m}}\sum_{i=1}^m \xi_i, \frac{1}{\sqrt{m}}\sum_{i=1}^m \Phi_i'\right)' \xrightarrow{d} N = (\xi, \Phi')'$$

where using the result for Model M1 in section 3.1, we obtain

$$\frac{1}{\sqrt{m}}\sum_{i=1}^m \xi_i \xrightarrow{d} \xi \sim N\left(0,\ \frac{1}{2(1-\sum_{k=1}^p \alpha_k)^2}\sigma^4\right) \tag{3.93}$$

and

$$\frac{1}{\sqrt{m}}\sum_{i=1}^m \Phi_i \xrightarrow{d} \Phi \sim N\big(\mathbf{0},\ \sigma^2\Gamma\big). \tag{3.94}$$

Therefore, by Slutsky's theorem, as $m \to \infty$, we obtain the final asymptotic result for Equation (3.91):

$$\begin{aligned} (R_m^2D_m)^{-1}R_mN_m \xrightarrow{d} D^{-1}N &= \begin{bmatrix} \frac{1}{2(1-\sum_{k=1}^p \alpha_k)^2}\sigma^2 & \mathbf{0} \\ \mathbf{0}' & \Gamma \end{bmatrix}^{-1} \begin{bmatrix} \xi \\ \Phi \end{bmatrix} \\ &= \begin{bmatrix} \xi^* \\ \Phi^* \end{bmatrix} \end{aligned} \tag{3.95}$$

where

$$\frac{1}{(1-\sum_{k=1}^{p}\alpha_k)}\xi^* \ \sim \ N(0,\ 2) \ \text{ and } \ \Phi^* \ \sim \ N(0,\ \sigma^2\Gamma^{-1}).$$

Using that $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$ and the above asymptotic results, we have the following properties:

$$t_{\boldsymbol{\alpha}} \quad \xrightarrow{d} \quad N(0, \mathbf{I_{p+1}}). \tag{3.96}$$

The result that $n\sqrt{m}(\widehat{\alpha}_0 - \alpha_0) = n\sqrt{m}(\hat{\rho} - 1) \xrightarrow{d} (1-\sum_{k=1}^{p}\alpha_k)^{-1}\xi^* \ \sim \ N(0,\ 2)$ is the same as that of the AR(1) model in Section 3.1. Similarly, we have the same asymptotic results as the simple AR(1) model for the estimate of the unit root coefficient in the AR (p) model with intercept and linear trend regardless of whether or not the deterministic parameters is interval-specific.

Chapter 4

Data Contamination and Unit Root Tests

4.1 Data contamination

Perron (1989) argued that if there is a structural change in the deterministic part of a time series model, then unit root tests tend to report a misleading conclusion that there is a unit root when in fact there is not. His paper triggered a controversy about the effect of structural changes on the unit root tests. In structural change problems, there are two main issues (i) whether the structural change point is known or not; and (ii) whether the structural change is single or multiple.

In the unit root test with a single break, the known (exogenous) structural change wasintroduced by Perron (1989), Lütkepohl, Müller and Saikkonen (2001) among others. The unknown (endogenous) change assumption has been made by Banerjee, Lumsdaine and Stock (1992), Evans (1989), Perron and Rodriguez (1998), Saikkonen

and Lütkepohl (2002), Zivot and Andrews (1992).

For the multiple breaks, see Ben-David, Lumsdaine and Papell (2003), Garcia and Perron (1996), Kanas (1998), Kapetanios (2005), and Lumsdaine and Papell (1997).

Our models have a local nature. We have seen a disadvantage is loss of power when a single model globally describes the data. Our local model should have an advantage under some Perron-type shifts in that only one of our intervals would be affected.

The basic idea of our model is that under the independent double array transformation we construct break-robust statistics which achieve the same asymptotic results as the standard unit root tests in the previous chapter.

There are mainly two types of structural change, namely, model change and parameter change. The parameter change model also can be divided into two parts: coefficient change and disturbance shock (variance) change. Here we focus on the coefficient change model which usually has one of four patterns of structural change:

- Level shift in Intercept Model
- Trend shift in Linear Trend Model
- Slope change in Linear Trend Model
- Trend shift and slope change in Linear Trend Model

Define

$$\delta(t) = \begin{cases} 1 & \text{if } t \geq t_0 \\ 0 & \text{otherwise} \end{cases} \qquad \delta_0(t) = \begin{cases} 1 & \text{if } t = t_0 \\ 0 & \text{otherwise} \end{cases}$$

$$\delta_1(t) = \begin{cases} t - t_0 & \text{if } t \geq t_0 \\ 0 & \text{otherwise} \end{cases}$$

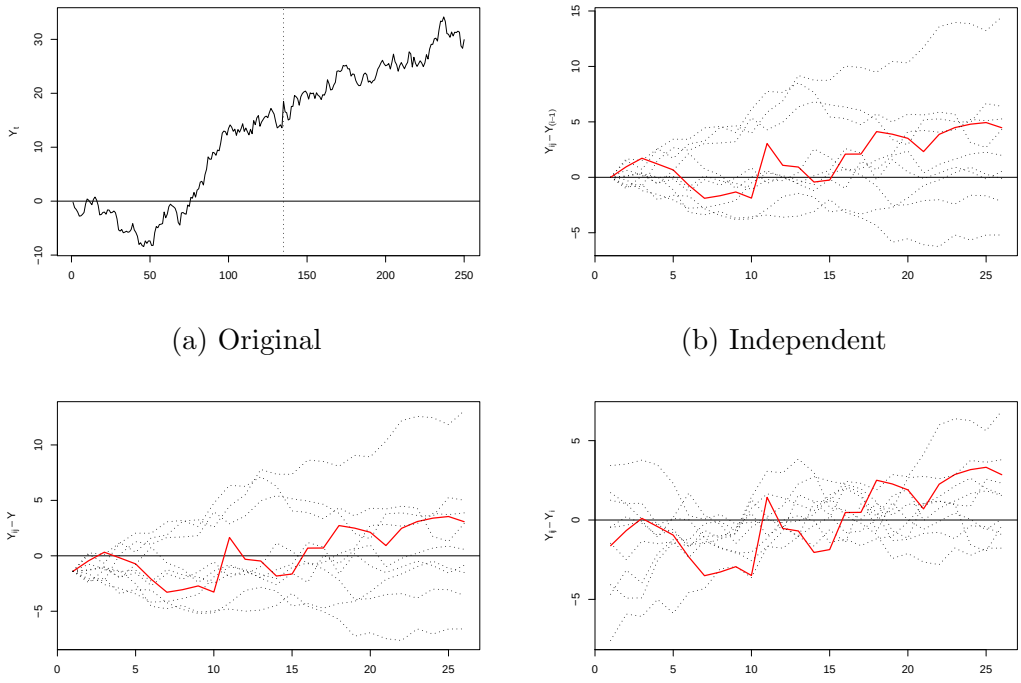
where t_0 is a break point.

Let α_1 denote the break size in the shift model, β_1 denote the magnitude of slope change, and let t_0 denote the break point in the original univariate sample with $t = 1, 2, \cdots, T$.

Case I: Level shift

$$\begin{aligned} t < t_0: &\quad Y_t - \alpha_0 = \rho\big(y_{t-1} - \alpha_0\big) + \epsilon_t \\ t = t_0: &\quad Y_t - \alpha_0 - \alpha_1 = \rho\big(Y_{t-1} - \alpha_0\big) + \epsilon_t \\ t > t_0: &\quad Y_t - \alpha_0 - \alpha_1 = \rho\big(Y_{t-1} - \alpha_0 - \alpha_1\big) + \epsilon_t. \end{aligned} \tag{4.1}$$

In Figure 4.1, we illustrate the data transformation with a structural change in Model 2. Plot (a) shows that the original model (Model 2) with size $\alpha_1 = 5$ at break point $t_0 = 135$ which is denoted by the dotted line. Plot (b) demonstrates the independent transformation in which the solid line represents the interval that was contaminated by a break in level. Plot (c) and (d) represent the subtraction of the global mean $(\bar{y})$ and interval-specific mean $(\bar{y}_i)$ from the independently transformed data, respectively. Comparing with no break case in Figure 3.2, we observe that only the one interval is affected by the structural change, and the others remain unchanged. Therefore, as m increases, the effect of the contaminated interval on the test statistics decreases, which implies that our data transformation is robust to the



(a) Original (b) Independent

(c) Global Mean (d) Interval Mean

Figure 4.1: Transformation of M2-G with Structural Change

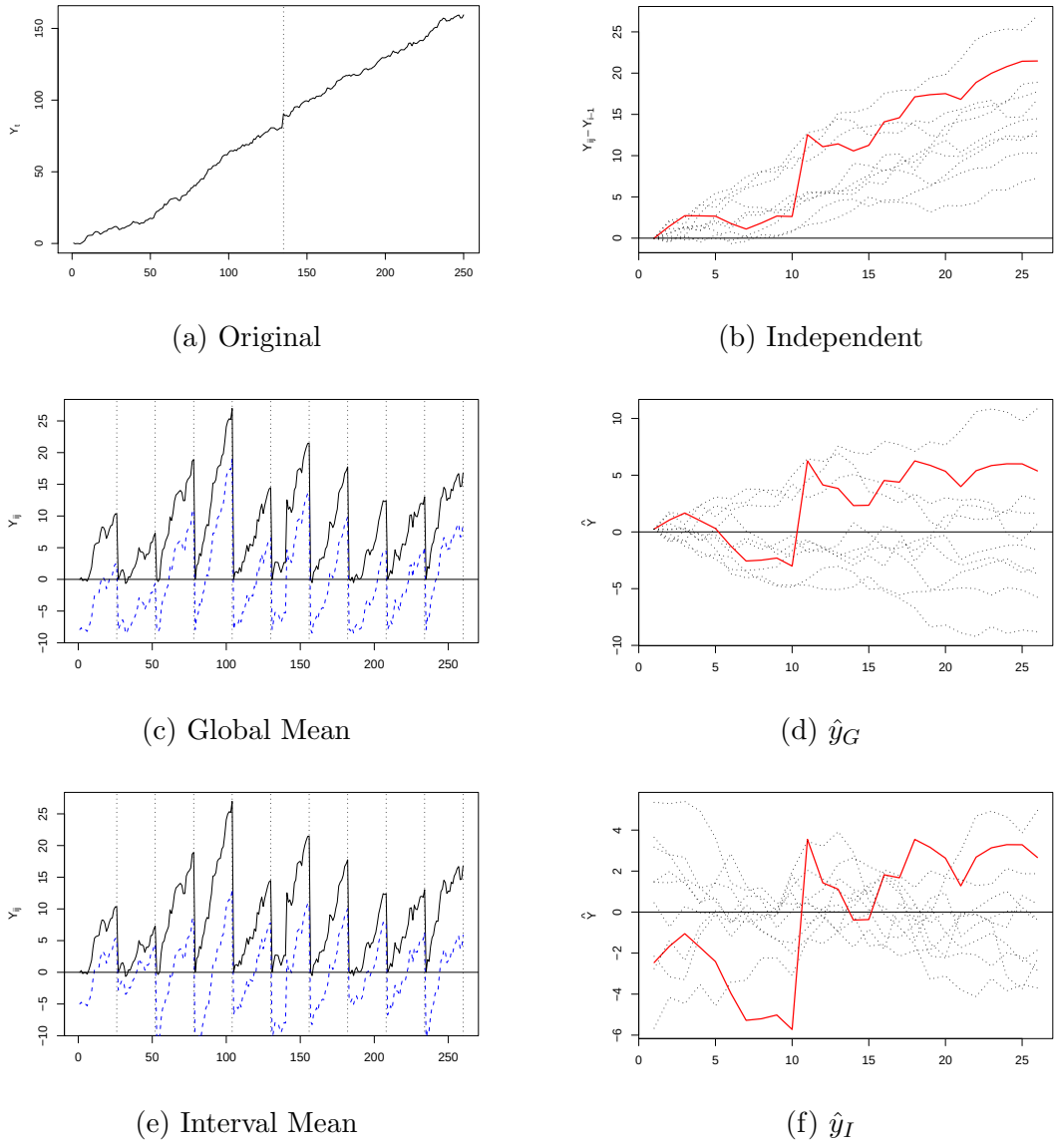
structural change, especially, the break in level type. Our transformation may not have an advantage in the presence of innovation variance breaks as in Cook (2002) and Kim, Leybourne, and Newbold (2002).

Case II: Trend shift

$$
\begin{aligned}
t < t_0: &\quad Y_t - \alpha_0 - \beta_0 t = \rho\big(Y_{t-1} - \alpha_0 - \beta_0(t-1)\big) + \epsilon_t \\
t = t_0: &\quad Y_t - \alpha_0 - \alpha_1 - \beta_0 t = \rho\big(Y_{t-1} - \alpha_0 - \beta_0(t-1)\big) + \epsilon_t \\
t > t_0: &\quad \begin{cases} Y_t = Y_{t-1} + \beta_0 + \epsilon_t, & \text{if } \rho = 1 \\ Y_t = \rho Y_{t-1} + (\alpha_0 + \alpha_1)(1-\rho) + \beta_0 t^* + \epsilon_t, & \text{if } \rho < 1 \end{cases}
\end{aligned}
\tag{4.2}
$$

where $t^* = (t - \rho(t-1))$.

In Figure 4.2 we draw the process of data transformation for the linear trend model with trend shift where we use $T = 250, m = 10, n = 25, \alpha_1 = 10$, and $t_0 = 135$.



(a) Original

(b) Independent

(c) Global Mean

(d) $\hat{y}_G$

(e) Interval Mean

(f) $\hat{y}_I$

Figure 4.2: Transformation of M3 with Trend Shift

Plot (a) represents the original process with a trend break, and Plot (b) illustrates the independent transformation. The trend shift permanently affects the sample period after the shock in the original data. On the other hand, the shift affects only one interval in the independently transformed data. In Plot (c), the solid line represents

the indenpendently transformed data, and the dotted line represents the subtraction of a global mean from the independent data. Plot (d) draws the residuals calculated by $y_{i,j} - \hat{\alpha} - \hat{\beta}(j - \bar{j})$. Plot (e) and (f) are very similar to Plot (c) and (d), but use the interval-specific parameters.

Case III: Slope Change

$$
\begin{aligned}
&t \leqq t_0: \qquad Y_t - \alpha_0 - \beta_0 t = \rho\big(Y_{t-1} - \alpha_0 - \beta_0(t-1)\big) + \epsilon_t \\
&t > t_0: \\
&\begin{cases} Y_t = Y_{t-1} + \beta_0 + \beta_1 + \epsilon_t, & \text{if } \rho = 1 \\ Y_t = \rho Y_{t-1} + \alpha_0(1-\rho) + \beta_0 t^* + \beta_1 t_0^* + \epsilon_t, & \text{if } \rho < 1, \end{cases}
\end{aligned}
$$

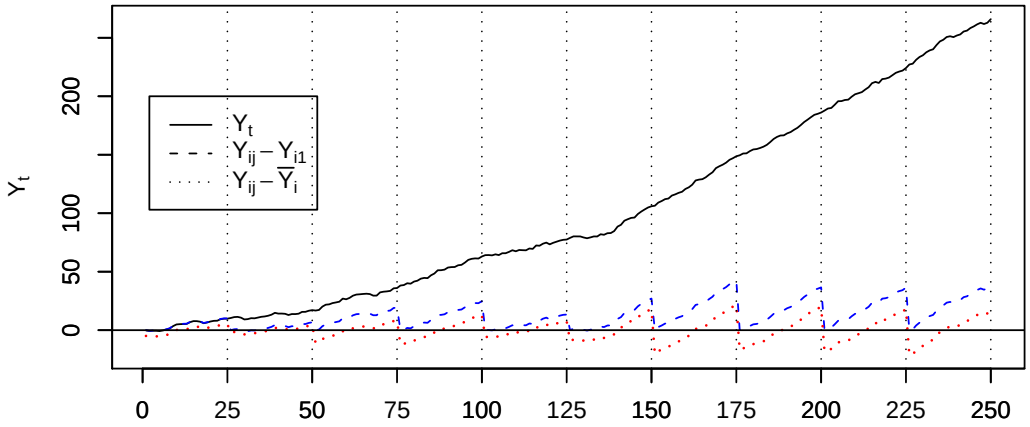
where $t^* = (t - \rho(t-1))$ and $t_0^* = (t - t_0 - \rho(t - t_0 - 1))$ for all $t > t_0$.

Case IV: Trend shift and slope change

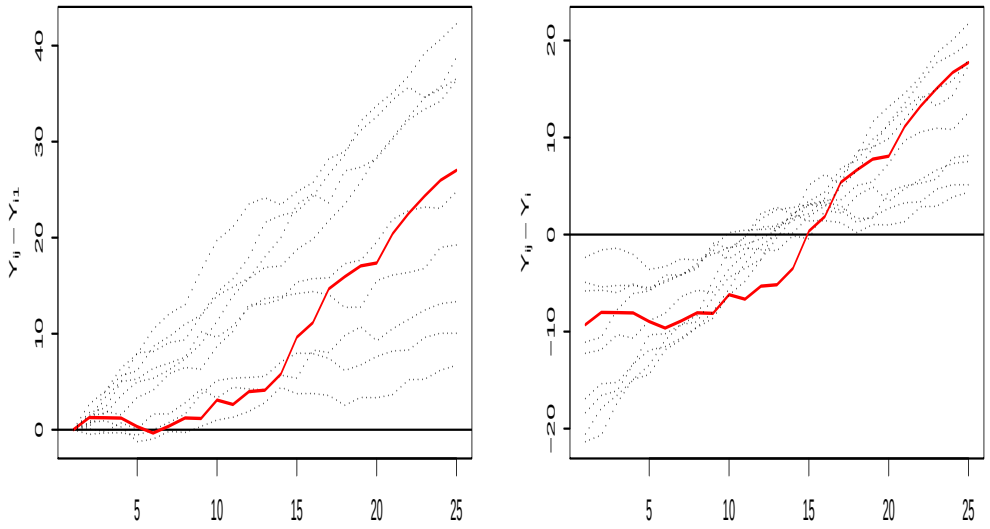
$$
\begin{aligned}
&t < t_0: \qquad Y_t - \alpha_0 - \beta_0 t = \rho\big(Y_{t-1} - \alpha_0 - \beta_0(t-1)\big) + \epsilon_t \\
&t = t_0: \qquad Y_t - \alpha_0 - \alpha_1 - \beta_0 t = \rho\big(Y_{t-1} - \alpha_0 - \beta_0(t-1)\big) + \epsilon_t \\
&t > t_0: \\
&\begin{cases} Y_t = Y_{t-1} + \beta_0 + \beta_1 + \epsilon_t, & \text{if } \rho = 1 \\ Y_t = \rho Y_{t-1} + (\alpha_0 + \alpha_1)(1-\rho) + \beta_0 t^* + \beta_1 t_0^* + \epsilon_t, & \text{if } \rho < 1, \end{cases}
\end{aligned}
$$

where $t^* = (t - \rho(t-1))$ and $t_0^* = (t - t_0 - \rho(t - t_0 - 1))$ for all $t > t_0$.

As in Figure 4.3 abd 4.4, if a single slope change in trend, which is denoted by the solid line, is contained in the model, the global trend model (M3-G) may not work well since the last value transformation data still have the two different slopes. As can be seen in (b) of Figure 4.3, there seem to be two groups with different slopes, and one messy interval. In this case, we expect that the interval-specific model, say M3-I, is appropriate excepting for the intervals containing break points. Model M3-I seems to be useful when trend shift and slope change occur simultaneously as in Figugure



(a) Original

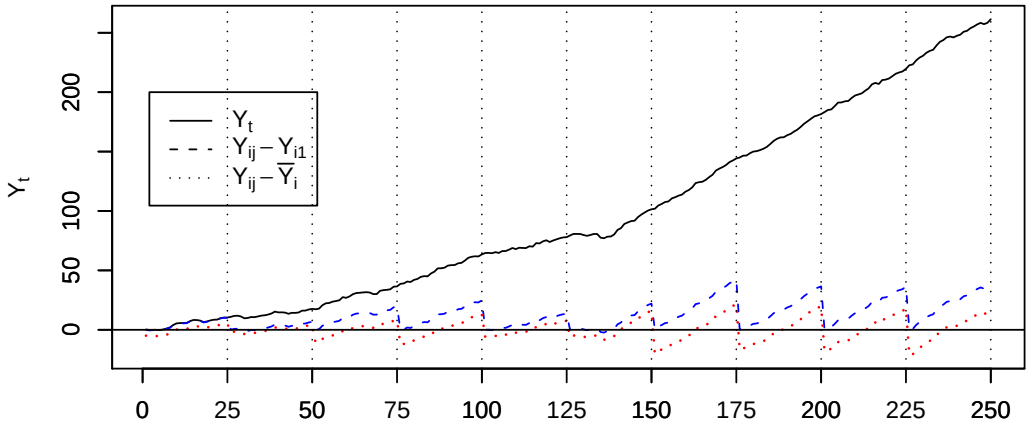


(b) Initial

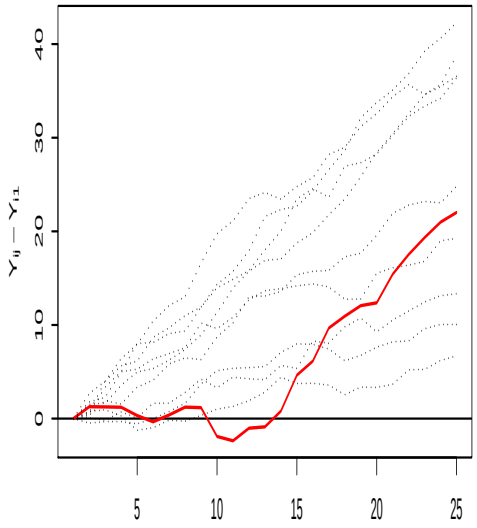
(c) Mean

Figure 4.3: Slope Change in Model 3

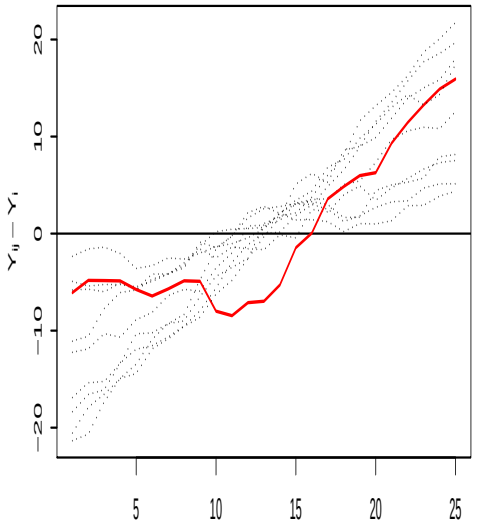
4.4. The following section provides the asymptotic behavior of our approach in the presence of structural breaks.



(a) Original



(b) Initial



(c) Mean

Figure 4.4: Trend Shift and Slope Change in Model 3

4.2 Asymptotic Behavior

Although Perron (1989) considered a single known break point in the unit root test, the recent literature has focused on multiple break points as in Lumsdain and Papell (1997). Hence, we generally assume the existence of multiple break points in the model.

Assumption 4.1 *A time series* $\{y_{ij} : 1 \leq i \leq m; 1 \leq j \leq n;\}$ *has at most* b *structural changes where* $b/m \to 0$ *as* $m \to \infty$. *The magnitude of break in level is* $O(1)$, *and hence, the* b *structural breaks are also* $O(1)$.

As mentioned above, we only consider the structural change in coefficients such as intercept and trend parameters. Rewriting the models which could allow the structural changes yields

$$\text{MG}: \quad y_{ij} - \theta d = \rho(y_{ij-1} - \theta d) + \epsilon_{ij} \tag{4.3}$$

$$\text{MI}: \quad y_{ij} - \theta_i d = \rho(y_{ij-1} - \theta_i d) + \epsilon_{ij} \tag{4.4}$$

where

$$d = \begin{cases} 1 & \text{for } M2 \\ (1,\ j)' & \text{for } M3, \end{cases} \quad \theta = \begin{cases} \alpha & \text{for } M2-G \\ (\alpha, \beta)' & \text{for } M3-G, \end{cases} \quad \theta_i = \begin{cases} \alpha_i & \text{for } M2-I \\ (\alpha_i, \beta_i)' & \text{for } M3-I. \end{cases}$$

Now define that $r_{i,j} = y_{i,j} - \hat{\theta} d$ or $r_{i,j} = y_{i,j} - \hat{\theta}_i d$.

No Break Case: Using the definition of $r_{i,j}$ and the asymptotic results in Chapter 3, we have the following result

$$n\sqrt{m}(\hat{\rho}-1) = \frac{\frac{1}{n\sqrt{m}}\sum_{i=1}^{m}\sum_{j=2}^{n} r_{ij-1}(r_{ij}-r_{ij-1})}{\frac{1}{n^2 m}\sum_{i=1}^{m}\sum_{j=2}^{n} r_{ij-1}^2} \tag{4.5}$$

$$= \frac{\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left(\frac{1}{n}\sum_{j=2}^{n} r_{ij-1}(r_{ij}-r_{ij-1})\right)}{\frac{1}{m}\left(\frac{1}{n^2}\sum_{j=2}^{n} r_{ij-1}^2\right)} \tag{4.6}$$

$$\xrightarrow{d} \frac{N}{D} \tag{4.7}$$

where $N_{mn} \overset{n\to\infty}{\Longrightarrow} N_m \xrightarrow{d} N$ and $D_{mn} \xrightarrow{p} D_m \overset{m\to\infty}{\longrightarrow} D$.

Structural Change Case:

As can be seen in Figure $4.1 - 4.2$, in this case the estimator can be divided into two parts: a break-free part and a contaminated part. Let index i denote the break-free intervals and let i^* denote the contaminated intervals where $i^* = 1, 2, \cdots, b$.

$$n\sqrt{m}(\hat{\rho}-1) = \frac{\frac{1}{n\sqrt{m}}\left\{\sum_{i\neq i^*}^{m}\sum_{j=2}^{n} r_{ij-1}(r_{ij}-r_{ij-1}) + \sum_{i^*=1}^{b}\sum_{j=2}^{n} r^*_{ij-1}(r^*_{ij}-r^*_{ij-1})\right\}}{\frac{1}{n^2 m}\left\{\sum_{i\neq i^*}^{m}\sum_{j=2}^{n} r_{ij-1}^2 + \sum_{i^*=1}^{b}\sum_{j=2}^{n} {r^*_{ij-1}}^2\right\}}$$

$$\overset{def}{=} \frac{N_{mn,0}+N_{mn,b}}{D_{mn,0}+D_{mn,b}}. \tag{4.8}$$

Lemma 4.1 *Under Assumption* 4.1*,*

$$N_{mn,b} = O_p(m^{-1/2}) \quad \text{and} \quad D_{mn,b} = O_p(m^{-1}).$$

Proof of Lemma 4.1

$$\begin{aligned} N_{mn,b} &= \frac{1}{n\sqrt{m}} \sum_{i^*=1}^{b} \sum_{j=2}^{n} r_{i^*,j-1}(r_{i^*,j} - r_{i^*,j-1}) \\ &= \frac{1}{\sqrt{m}} \sum_{i^*=1}^{b} \left(\frac{1}{n} \sum_{j=2}^{n} r_{i^*,j-1}(r_{i^*,j} - r_{i^*,j-1}) \right) \\ &\stackrel{\text{def}}{=} \frac{1}{\sqrt{m}} \sum_{i^*=1}^{b} A_{i^*} \qquad (\because A_{i^*} = O_p(1)) = O_p(m^{-1/2}). \end{aligned}$$

Similarly, we have $D_{mn,b} = O_p(m^{-1})$. Q.E.D.

Theorem 4.1 *Under Assumptions* 3.2 *and* 4.1*, the following statistics have the same asymptotic distribution.*

$$(1) \quad \frac{N_{mn}}{D_{mn}}; \tag{4.9}$$

$$(2) \quad \frac{N_{mn,0}}{D_{mn,0}}; \tag{4.10}$$

$$(3) \quad \frac{N_{mn,0} + N_{mn,b}}{D_{mn,0} + D_{mn,b}}, \tag{4.11}$$

where

$\frac{N_{mn}}{D_{mn}}$: *statistic for no break case*

$\frac{N_{mn,0}+N_{mn,b}}{D_{mn,0}+D_{mn,b}}$: *statistic for break case; and*

$\frac{N_{mn,0}}{D_{mn,0}}$: *statistic for break-free parts of break case.*

Proof of Theorem 4.1 We know that t-statistics for the break-free models are a function of $\hat{\sigma}^2$, N_{mn}, and D_{mn}. Now define $\hat{\sigma}_b^2$ as the estimator of σ^2 for the structural change models. Then the t-statistics for the stuctural change models are a function of $\hat{\sigma}_b^2$, $N_{mn,0}$, $N_{mn,b}$, $D_{mn,0}$, and $D_{mn,b}$. Q.E.D.

Lemma 4.2 *Under Assumption* 3.2 *and* 4.1*,*

$$\hat{\sigma}_b^2 = \hat{\sigma}^2 + O_p(m^{-1}). \tag{4.12}$$

Proof of Lemma 4.2

Let $e_{i,j}^*$ denote the residual of the structural change model.

$$\begin{aligned}
\frac{1}{mn}\sum_{i=1}^{b}\sum_{j=1}^{n} {e_{i,j}^*}^2 &= \frac{1}{mn}\sum_{i=1}^{b}\sum_{j=1}^{n}(e_{i,j} + O_p(1))^2 \\
&= \frac{1}{mn}\sum_{i=1}^{b}\sum_{j=1}^{n} e_{i,j}^2 + \frac{1}{m}\sum_{i=1}^{b} O_p(1) \\
&= \frac{1}{mn}\sum_{i=1}^{b}\sum_{j=1}^{n} e_{i,j}^2 + O_p\left(\frac{1}{m}\right).
\end{aligned}$$

So, if we add the above sum of squares to those from the other intervals, we will obtain the desired result.

Theorem 4.2 *Under Assumption* 3.2 *and* 4.1*, the following statistics have the same asymptotic distribution:*

$$(1) \quad t(\hat{\sigma}^2, N_{mn}, D_{mn}); \tag{4.13}$$

$$(2) \quad t(\hat{\sigma}_b^2, N_{mn,0}, D_{mn,0}); \tag{4.14}$$

$$(3) \quad t(\hat{\sigma}_b^2, N_{mn,0}, N_{mn,b}, D_{mn,0}, D_{mn,b}). \tag{4.15}$$

Proof of Theorem 4.2

We have already shown the properties of $N_{mn}, N_{mn,0}, N_{mn,b}, D_{mn}, D_{mn,0}$, and $D_{mn,b}$. Hence, by Lemma 4.2 the three statistics asymptotically have the same distribution. Q.E.D.

An appealing feature in our approach is that $\rho-$ and $t-$ statistics do not use any information on the size and location of structural breaks while the conventional approaches assume the break locations are known or have to search for them. For example, if we consder two known breaks in level, then the conventional t-statistic is additionally a function of $\hat{\boldsymbol{\theta}}(\lambda_1, \lambda_2), \lambda_1$, and λ_2 where λ_1 and λ_2 denote the break points, and $\hat{\boldsymbol{\theta}}(\lambda_1, \lambda_2)$ is the estimator of $\boldsymbol{\theta}$ using the information on the break fractions (see Perron, 1989). If the break points are unknown, then we should also estimate those points as well as break sizes (α_1, α_2) , *e.g.,* $\hat{\boldsymbol{\theta}}(\hat{\lambda_1}, \hat{\lambda_2}, \hat{\alpha_1}, \hat{\alpha_2},)$ (see Zivot and Andrews, 1992). In our approach, the information on the strucutural change such as break point and size, however, is not required since the effect of structural change is restricted to a few intervals containing break points whose effects will eventually disappear as m gets large.

4.3 Spurious rejection with structural change and size distortion

Since Perron's (1989) paper, the unit root test with structural change has been a popular topic. In most cases, the presence of an undetected break makes an otherwise stationary process appear to be a unit root process. However, in contrast to this, Leybourne, Mills and Newbold (1998) were concerned with the case in which a unit root process with break can appear to be a stationary process if standard Dickey-Fuller tests are applied. They showed that the location of an unknown break point in the sample is very important to the behavior of unit root tests. Especially, they

illustrated a peculiar phenomenon that if a break occurs in the first 20% of the sample, then the t-statistic often falsely rejects the null hypothesis at the nominal 0.05% level when calculated from without modeling the break.

As a simple example, they consider the regression model with break in level given by

$$Y_t = \alpha_1 s_t(\tau) + W_t, \quad W_t = \rho W_{t-1} + \epsilon_t, W_0 = 0, \quad t = 1, 2, \cdots, T,$$

where $\epsilon_t \sim i.i.d.(0, 1)$. In this model the break fraction is

$$s_t(\tau) = \begin{cases} 0, & t \leq \tau T \\ 1, & t > \tau T, \quad \tau \in (0, 1) \end{cases} . \tag{4.16}$$

and α_1 is the break size as a multiple of the standard deviation of the ϵ_t. It is well known that the null distribution is asymptotically invariant to a break of size $\alpha_1(T) = o(T^{-1/2})$ which corresponds to Condition B (slowly evolving trend) in Elliot et al. (1996). However, Leybourne et al. argued that even if α_1 grows at rate $T^{1/2}$, i.e. $\alpha_1(T) = O(T^{1/2})$, then the limiting distributions of the test statistics depend on the break size. To show this, they assume that $\alpha_1 = kT^{1/2}$ for k fixed, and derive the limiting distribution of the t-statistics given by

$$t_\mu \Rightarrow \frac{\int_0^1 W_1(r)dW(r) - k\int_0^1 W(r)dr + k(1-\pi)W(1) - \pi k^2}{(1+k^2)^{1/2}\{\int_0^1 W_1^2(r)dr - 2\pi k\int_0^1 W(r)dr + 2k\int_\tau^1 W(r)dr + \pi(1-\pi)k^2\}^{1/2}},$$

where $\Rightarrow$ denotes weak convergence, $W_1(r) = W(r) - \int_0^1 W(s)ds$, $\pi = 1 - \tau$. This limiting distribution is asymmetric around $\tau = 0.5$.

In the panel data framework, if there is a structural change at the same point for all

individuals, then we have the following intermediate results as $n \to \infty$,

$$t_\mu \Rightarrow \frac{\frac{1}{\sqrt{m}} \sum_{i=1}^{m} \left(\int_0^1 W_{1i}(r) dW_i(r) - k \int_0^1 W_i(r) dr + k\tau_i W_i(1) - \pi_i k^2 \right)}{(1+k^2)^{1/2} \{ \frac{1}{m} \sum_{i=1}^{m} \left(\int_0^1 W_{1i}^2(r) dr - 2\pi_i k \int_0^1 W_i(r) dr + 2k \int_{\tau_i}^1 W(r) dr + \pi_i \tau_i k^2 \right) \}^{1/2}},$$

where $\pi_i = 1 - \tau_i$ denotes the fraction in the i^{th} interval. Hence, the position of the break in an interval can affect the limiting distribution of t-statistic, which implies that in the panel data the spurious rejection in the presence of break still occurs. As well, the break in level in the linear trend model has a similar asymptotic result as in Leybourne *et al.* (1998) but the result depends on i.

Figure 4.5 illustrate how a structural change leads to spurious rejection in the unit root test using the graphical method in Davidson and MacKinnon (1998) and Cook (2001). For the total sample size of 100, we consider the type of break in level with three different break sizes, $\alpha_1 = 2.5, 5$, and 10 as in Leybourne *et al.* (1998). In the plot, the horizontal axis denotes the break point, which ranges from 2 to 100 by 1, and the vertical axis represents the empirical rejection frequencies under the nominal size 0.05. Tests are calculated without modeling structural changes. As can be seen in Figure 4.5, the size distortion is very serious in the early sample, but rapidly disappears as the break occurs later in the sample. That means that if there is a structural break under the null hypothesis, the proportion of spurious rejections in favor of stationarity increases as the break size gets large and as n decreases.

Lee (2000) has shown that the spurious rejection problem in the presence of a break under the null hypothesis is restricted to the Dickey-Fuller test in the sense that the test is based on the conditional likelihood estimator neglecting the information

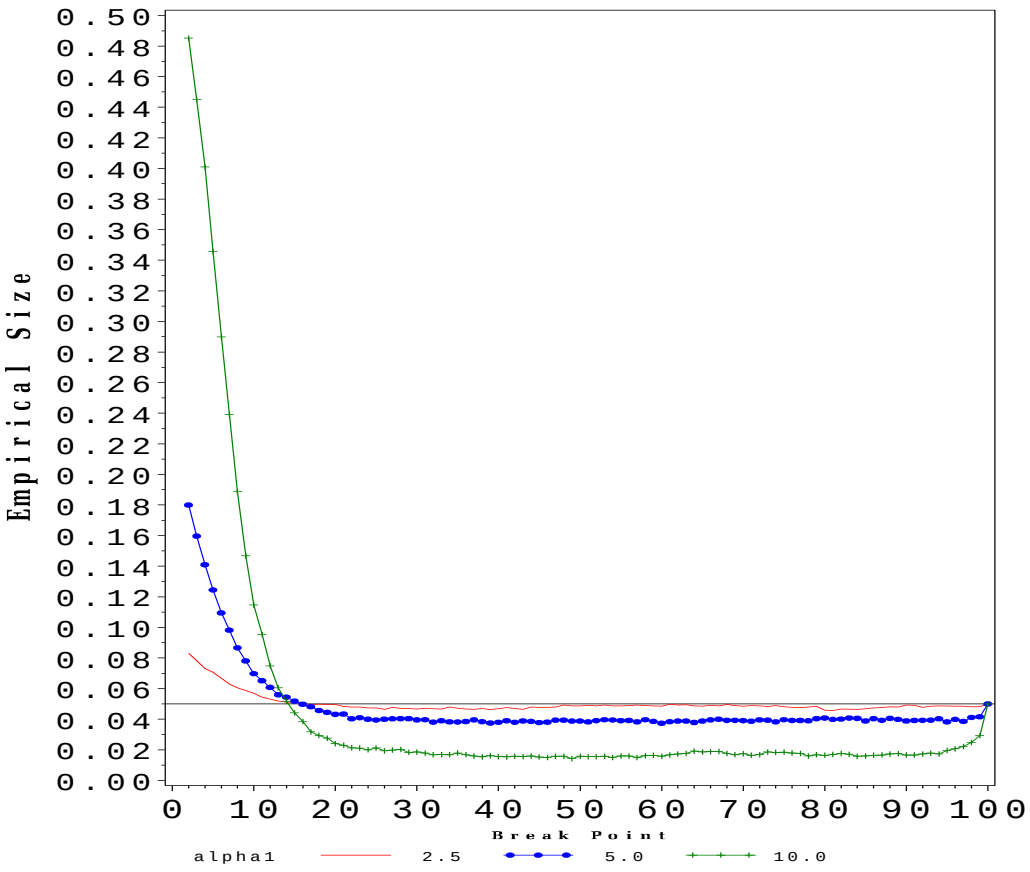


Figure 4.5: Empirical Size of $t(0.05)$ with Break in Level in M2-G [T=100, m=5]

on the first observation. He also illustrated by simulation that the unconditional likelihood estimator of Gonzalez-Farias (1992), uniformly most powerful (UMP) test of Elliott and Stock (1994), and Lagrangian multiplier (LM or score) test of Schmidt and Phillips (1992) have no spurious problem in the early sample.

In our framework, the structural change in the original data is a kind of outlier in the transformed data because the break occurs in only one interval while the break in the panel data occurs at the same point of every individual, which implies that we should be careful of unit root tests in the panel data. Theorem 4.2 indicates that in

the univariate time series a single point break does not affect the t-statistics as the sample size increases.

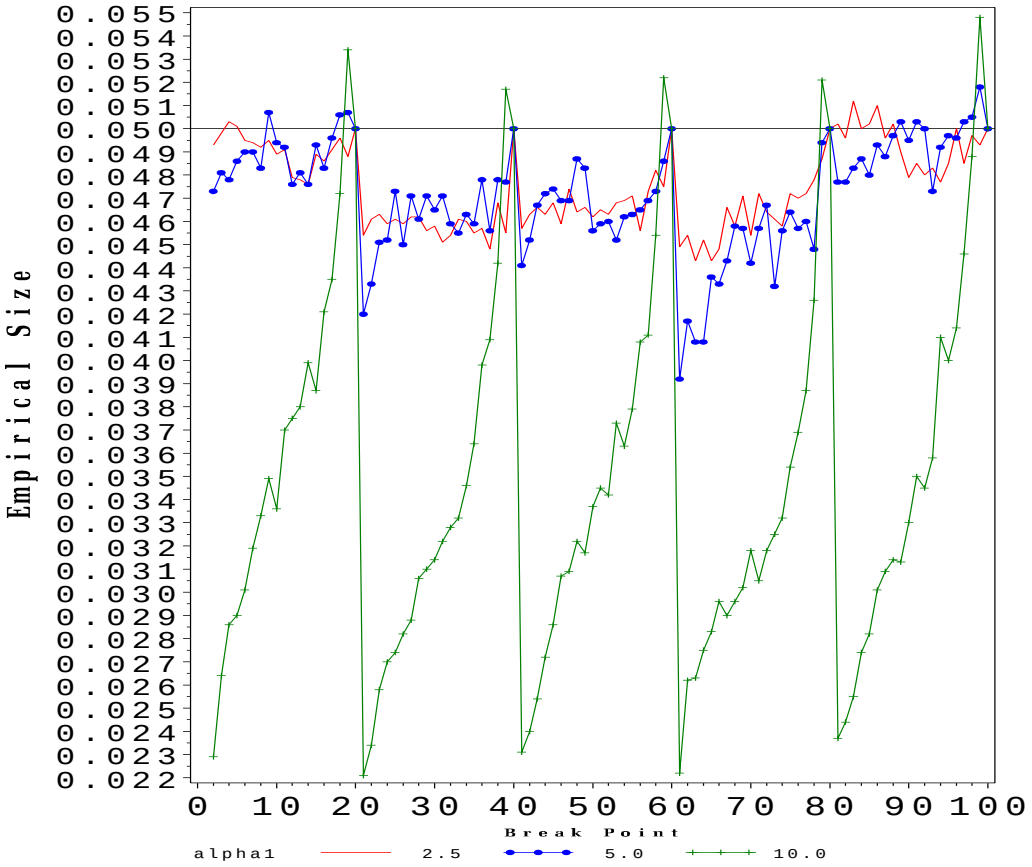


Figure 4.6: Actual Size of $t(0.05)$ with Break in Level in M2-G [T=100, m=5]

In Figure 4.6, we illustrate the pattern of size distortion in Model M2-G with $(m, n) = (5, 20)$ and break sizes of $2.5, 5, 10$. Unlike the conventional DF test ($m = 1$), the size distortion is not serious in the early part of each interval. The empirical sizes fall in the intervals$0.045 - 0.051, 0.040 - 0.052, 0.022 - 0.055$ for break sizes $2.5, 5, 10$, respectively. Hence, the graph indicates as in Figure 4.5 that the bigger break size, the worse size distortion.

Up to $\alpha_1 = 5$, the size distortion is overall less than the simulation standard error, 0.014, which implies that we can use the no-break critical values without serious size distortion. That is, our approach is robust to the structural change problem with moderate break sizes. However, size distortion becomes serious as the number of observations in each interval n gets small. Therefore, we should choose a combination of (m, n) such that $\sqrt{m}/n$ is close to zero.

4.4 Finite sample behavior in the presence of structural change

In this section we study the finite sample behavior of unit root tests in the presence of structural changes. We consider a relatively small sample size, $T = 100$, since as T increases the power of both our approach and DF tests approaches to 1 even in the presence of structural changes. We also consider three different break sizes, $\alpha_1 = 2.5, 5$, and 10 for breaks in both level and trend as in the previous section. Since the error term $\epsilon_t \sim N(0,\ 1)$, $\alpha_1 = 10$ is a very large value, ten times the standard error of white noise. However, extremely large breaks are sometimes observed in the real world. For example, the exchange rates of Asian countries during the Asian Crisis have a change over ten times the sample standard deviation. For the slope change in trend, we use two values of slope change, $\beta_2 = 0.2$ and 0.5.

We generate a break at each point excepting the initial observation before the data transformation, and then we calculate the power of the tests for each break where we change m and n for fixed T. In the simulation study we are not interested in

detecting where the structural change occurs since the test statistics do not asymptotically involve the information on the break. Instead, we examine the effect of the independent transformation, under the null hypothesis, on the power of test.

Now we analyze the power of the test in the presence of structural change according to the independent data transformation. The simulation results for empirical powers are illustrated through Figure $7.14 - 7.42$ in Appendix 7.4 and Appendix 7.5. There are lots of plots so we briefly summarize as follows:

- Total sample size before transformation: $T = 100$
- Transformation: $(m, n) = (1, 100), (5, 20), (10, 10), (20, 5)$
- Break size in level and trend: $\alpha_1 = 2.5, 5, 10$
- Multiple breaks in level: $(\alpha_1, \alpha_2) = (2.5, 2.5), (2.5, -5)$
- Distance between two breaks: 37 observations
- Slope change in trend: $\beta_2 = 0.2, 0.5$
- Regression models used: M1, M2-G, M2-I, M3-G, M3-GD, M3-I
- Test statistics: t-statistics obtained by POLS
- Significance level: 5% simulated critical value without break
- Size adjustment: 5% simulated critical value with break

As can be seen, for $m = 1$ every plot has very high power in the early part of the sample and low power in the other part, and the magnitude is more severe. This

phenomenon is consistent with the under-size and over-size distortion when the no-break critical values are used. For $m = 5, 10$ and 20, the effect of the early sample break disappears.

The power of the test monotonically increases as the location of the break moves from the first point to the last point of each interval, and the power is close to the no-break power in the case that the break occurs at the last observation in each interval since the impact of the break on the power is the smallest in that interval. Another feature is that after the independent transformation the power pattern is almost the same across intervals while for $m = 1$ the power plot is an asymmetric U-shape around the mid-point of the sample. This implies two things: First, we do not need to check the power pattern through the whole sample in the transformed sample. Instead, we just need to calculate the power only in one interval. For example, if we choose $m = 5$ for $T = 100$, then it is sufficient to calculate the power for the break in the 20 observations in any single interval. In other words, for the global mean model, the power corresponding to the break at observation $y_{i,j}$ depends only on j, not i where $y_{i,j}$ denotes the j-th observation in the i-th interval. Second, in the presence of structural change the highest power we can achieve with the transformed data cannot exceed the no-break power using the same transformation, say, the same m and n.

In Chapter 3, we classified the transformed models into MG and MI models where MG denotes the global parameter models and MI denotes the interval-specific parameter models. We also considered MGD and MID models where D denotes the deletion of the constant term from MG and MI. First, we investigate the power patterns of model M2-G and M2-GD. In model M2-GD, the power pattern of the original data

is monotonically increasing in the break point, which is explained by the misspecification effect of the model. That is, if the break occurs at the early part of the sample, then the true model seems to be close to the intercept model. So the M2-GD regression model tends to misspecify the true model, and the misspecification effect gradually disappears as the break occurs later, which implies that the testing power for the break at the last point is quite close to that of the no-break case.

In Figure 7.14, for each $m = 1, 5, 10$ and 20, we illustrate the testing power at the 5% size of model M2-GD (or M1) with break in level in which we consider $T = 100, \alpha_1 = 2.5$, and $\rho = 0.9$. For $m = 1$, the power monotonically increases from 18% to 70% as the break point moves from the early point to the last point. The testing power ranges between $25-35\%$, $20-25\%$, and $15-19\%$ for $m = 5, 10$, and 20, respectively. In this plot the power for the original data ($m = 1$) dominates the power for the transformed data, excepting for the first 40% sample of $m = 5$. However, as the break size α_1 becomes large, the power of $m = 1$ gets close to zero faster than that of $m = 5, 10, 20$, and hence, the power of the original data test is dominated by that of the transformed data test.

Next, we examine the power pattern of model M2-G which considers the intercept term constant across the intervals in the regression model. In this case, the power plot of the original data is U-shaped in break point, which is different from the power plot of M2-GD, a monotonically increasing plot. However, the power plots for the transformed data have the same pattern as in model M2-GD. Overall, the power of M2-G is worse than that of M2-GD, which is due to an additional nuisance parameter. The power of the original test with $m = 1$ is dominated by the power of $m = 5$ if

the break occurs in the mid-sample. As mentioned in Section 4.3, our approach does not lead to serious size distortion, the size-adjusted empirical power is only slightly improved. For M2-I, we present empirical power in Figure 7.23 - 7.25, but this model has very poor performance, which suggests that we use M2-G or m2-GD if we suspect a break in level.

We demonstrate the pattern of empirical power in the presence of trend break in Figure 7.26-7.34. From investigation of the plots, we see that there is a spurious rejection problem for the conventional DF test, and the power of $m = 5$ overall dominates that of $m = 1$ by $4 - 6\%$. The empirical power of M3-G is almost the same as that of M2-G while the conventional DF test has lower power if a nuisance parameter is added. The result is also confirmed in the situation without structural change.

In Figure 7.35-7.42, we report the empirical power for slope change in trend. As can be seen, if we use the M3-G model, the power pattern of transformed models is opposite to that of $m = 1$. This is due to model misspecification. On the other hand, size adjusted empirical power is quite flat but very low. M3-I has no misspecification problem, but it has very low empirical power.

Table 4.1 presents the power patters of three unit root tests, $i.e.$, DF, Perron and KD (Kim and Dickey) tests. The DF and KD tests, which neglect the detection of break point (λ), depend on break size (α) while the Perron test does not if the break point is correctly detected but does otherwise. As can be seen in Table 4.1, a small value of α, say, 2.5, leads to the similar power pattern for these three tests. However,

Table 4.1: Power Comparison of DF, Perron, and KD Approach

λ denotes break fraction. KD denotes the approach of this paper with $m = 5$. Subscripts 1 and 2 denote break size 2.5 and 5, respectively. $Perron_1$ and $Perron_2$ represent misspecification of break with $\lambda + 0.05$.

λ	ρ	KD_1	KD_2	DF_1	DF_2	$Perron$	$Perron_1$	$Perron_2$
	0.80	0.48	0.29	0.53	0.17	0.49	0.43	0.33
0.3	0.90	0.20	0.15	0.16	0.09	0.14	0.13	0.12
	0.95	0.10	0.08	0.07	0.06	0.07	0.07	0.07
	0.80	0.49	0.30	0.57	0.28	0.56	0.41	0.40
0.5	0.90	0.21	0.15	0.17	0.11	0.18	0.15	0.16
	0.95	0.09	0.09	0.08	0.06	0.09	0.09	0.08
	0.80	0.48	0.29	0.55	0.22	0.51	0.47	0.38
0.7	0.90	0.20	0.15	0.17	0.11	0.15	0.15	0.14
	0.95	0.09	0.08	0.08	0.07	0.08	0.08	0.08

if α changes from 2.5 to 5, DF and KD tests have a severe loss of power. It is more serious in the DF test. The Perron test also suffers power reduction if the break point is misspecified. In this case, the loss of power comes from two sources, break size and misspecification magnitude. In the table, we report two break sizes under 5% misspecification magnitude. This misspecification issue in the Perron type unit root test is studied by Montañés and Ollogui (1999). Another feature is that as ρ gets close to 1, the KD test has slightly better performance than DF and Perron's.

Finally, we investigate the empirical power in the presence of multiple breaks, which is allowed in Assumption 4.1. For simplicity, we generate two breaks with $(\alpha_1, \alpha_2) = (2.5, 2.5), (2.5, -5)$ where the distance between two breaks is 37. In Figure 7.43-7.46, the empirical power of M2-G is reported. In each plot, the horizontal axis

represents the first break point, α_1, which varies from 2 to 63. Thus, if the first break occurs at the 63-th observation, then the second break occurs at the last observation. For the break combination of $(2.5, 2.5)$, the power of $m = 1$ is very similar to that of $\alpha_1 = 5$ in the single break case. However, the power of transformed model is close to that of $\alpha_1 = 2.5$. Therefore, we expect that our approach may outperform the conventional unit root tests as the number of break points increases.

Chapter 5

Empirical Data Analysis

Through the previous chapters we have focused on the data transformation, model setup, estimation, and asymptotic behavior both with uncontaminated and contaminated data. In this chapter, we apply our method to some real world data and compare it to other approaches. To this end, we consider two different data sets: US GDP as an example of a small sample and US/Canada exchange rate as a large sample.

When we transformed the univariate data into m sets of n observations, we chose m and n both in theoretical and practical ways. The theoretical restriction is that the growth rate $\sqrt{m}/n \to 0$, which is a reasonable restriction in the sense that the power of the test depends more seriously on n than m. In practice, if each interval is too small, the data transformation can make the originally significant deterministic term insignificant, which suggests that we may want to increase interval size until the statistical properties of the deterministic terms are preserved. Thus, we shall choose a sufficiently large n relative to m in practice.

5.1 A small sample case: US GDP data

We first apply our method to a family of US GDP data which is one of the most controversial macroeconomic data sets in the unit root test literature since Nelson and Plosser (1982). The family has the nominal and real GDP, and GDP deflator which are collected both quarterly and annually. GDP data typically have a kind of trend, which could be linear or nonlinear and/or deterministic or stochastic. Beside the issue of linearity, the main interest in the unit root test is whether the GDP data has a linear deterministic trend or linear stochastic trend due to a drift term in the random walk process. This has far-reaching implications both in economic theory and statistical analysis.

The deterministic trend implies that a macroeconomic policy has a temporary effect so that the process converges to the long-run equilibrium, and the series could be made stationary by detrending. On the other hand, the stochastic trend view is that a policy produces a permanent shock, and the series is made stationary by differencing. For details, see Enders (2004, Chap. 4), Hamilton (1994, Chap. 16) and Murray and Nelson (2000).

Annual US Real GDP

In order to perform the unit root test in the annual US GDP data, researchers have typically used the Nelson and Plosser (1982) data, or the extended Nelson and Plosser data (Schotman and van Dijk, 1991), or Maddison (1995) data since they can easily compare their own method to others using these popular data sets. We use the Maddison data to compare our results to the ADF, Perron, and Zivot and Andrews

test results in Murray and Nelson (2000) where the ADF test does not consider any kind of structural breaks in the deterministic part, the Perron test considers an exogeneous structural break, and the Zivot and Andrews (ZA) test considers an endogenous break. Although there are numerous unit root papers, it is sufficient to discuss the topic of structural change and unit root test using these three seminal papers as they cover the basics of the main approaches to the problem.

Figure 5.1 plots the annual US real GDP from 1870-1994. As can be seen, the series suddenly crashed down around the Great Depression era then grew at a rapid rate from the early 1930's to the mid 1940's. In the other periods, the series appears to have a linear trend, either deterministic or stochastic.

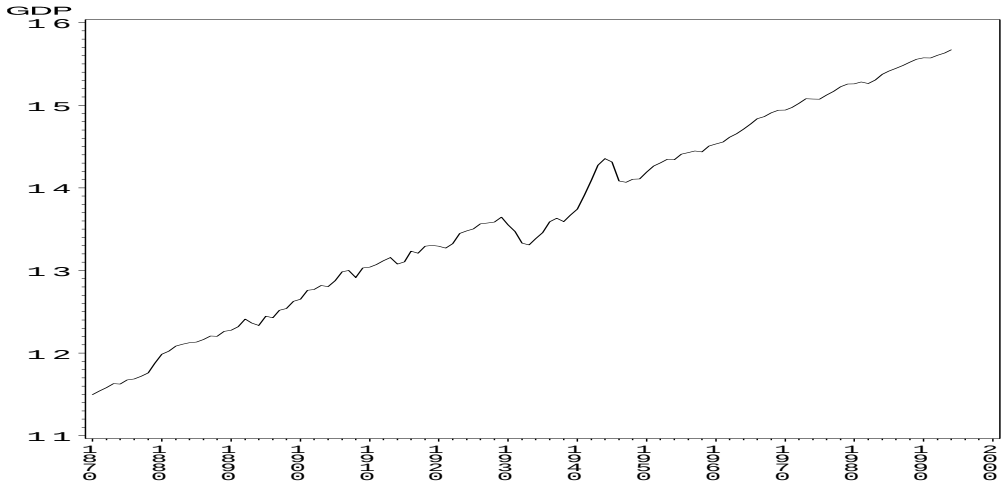
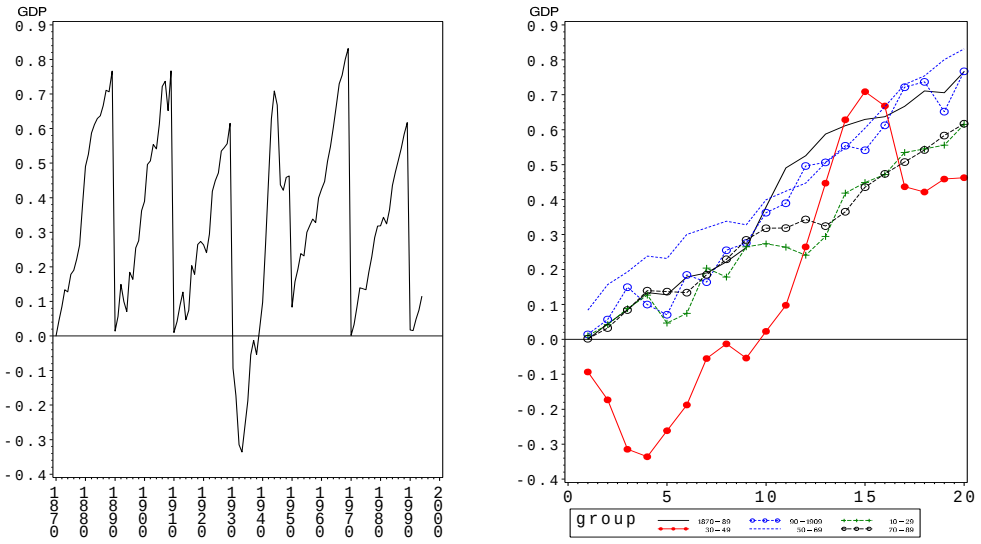


Figure 5.1: The Log Valued Annual US Real GDP: 1870-1994

In Figure 5.2, we first transform the original data to the double-indexed data using $m = 6$ and $n = 20$ which means that each interval has 20 years of observations.

In the unit root test, the first step is to select an appropriate lag for the target series



(a) Over Years (b) Within Interval

Figure 5.2: Transformation of Annual US GDP: m = 6, n = 20

Table 5.1: Unit Root Tests for Annual US Real GDP: M3-G

	Parameter	Estimate	Std Err	t Value	Pr > |t|
	Intercept	0.0047	0.0146	0.32	0.7507
m=6	Trend	0.0054	0.0020	2.71	0.0078
n=20	$y_{i,j1}$	-.1259	0.0417	**−3.02**	0.0032
	$\Delta y_{i,j-1}$	0.3142	0.0934	3.36	0.0011
	Intercept	0.0157	0.0126	1.24	0.2166
m=4	Trend	0.0044	0.0015	2.86	0.0051
n=30	$y_{i,j1}$	-.1239	0.0389	**−3.19**	0.0019
	$\Delta y_{i,j-1}$	0.3030	0.0902	3.36	0.0011

Table 5.2: Unit Root Tests for Annual US Real GDP: M3-I

	Parameter	Estimate	Std Err	t Value	Pr > |t|
	mean1	-.0257	0.0307	-0.84	0.4054
	mean2	-.0345	0.0305	-1.13	0.2616
	mean3	-.0300	0.0297	-1.01	0.3147
	mean4	-.1989	0.0498	-4.00	0.0001
m=6	mean5	0.0165	0.0289	0.57	0.5699
n=20	mean6	-.0042	0.0292	-0.14	0.8855
	trend1	0.0197	0.0040	4.87	< .0001
	trend2	0.0198	0.0039	5.05	< .0001
	trend3	0.0151	0.0033	4.61	< .0001
	trend4	0.0251	0.0050	5.04	< .0001
	trend5	0.0175	0.0036	4.89	< .0001
	trend6	0.0137	0.0032	4.26	< .0001
	$\Delta y_{i,j-1}$	0.4464	0.0911	4.90	< .0001
	$y_{i,j1}$	-.4468	0.0748	**-5.97**	< .0001
				(**−1.56**)	
	mean1	0.0095	0.0223	-0.84	0.6723
	mean2	0.0231	0.0220	-1.13	0.2956
m=4	mean3	0.0915	0.0325	-1.01	0.0058
n=30	mean4	0.0380	0.0220	-4.00	0.0891
	trend1	0.0122	0.0026	4.87	< .0001
	trend2	0.0104	0.0022	5.05	< .0001
	trend3	0.0138	0.0029	4.61	< .0001
	trend4	0.0103	0.0024	5.04	< .0001
	$\Delta y_{i,j-1}$	0.3879	0.0886	4.38	< .0001
	$y_{i,j1}$	-.3252	0.0886	**-5.24**	< .0001
				(**−1.74**)	

because the true lag is usually unknown and chosen by a data-dependent method. In practice, we use two selection criteria, the general-to-specific (GS) method and Schwartz's Bayesian information criterion (SBC, BIC, or SIC). Hall (1994) and Ng and Perron (1995) show that SIC chooses the correct lag in the limit, and GS tends to choose the lag at least as large as the true lag where the GS strategy is that we first choose an arbitrary maximum value of lag *a priori*, and then delete lags sequentially until the absolute value of t-statistic is greater than 1.645. They recommend GS in the sense that many lags result in a moderate decrease in power of test but a substantial improvement in size distortion. Sims, Stock, and Watson (1990) argued that if the deterministic part of the true model is correctly included, then SIC is preferable, but they suggested to use both methods in practice. Thus, we consider both criteria in our analysis.

Table 1 in Murray and Nelson (2000) shows that for SIC three tests commonly select lag 1 while for GS, the ADF selected lag 6 and both Perron and ZA selected lag 8. The Perron test exogenously picked up the break point as of 1929 and the ZA also chose year 1929 as the year that minimizes the unit root t-statistic. After selecting the lag and fitting the break, these unit root tests all reject the null hypothesis in favor of the deterministic trend.

Now we turn to our method of analysis. We consider two models, M3-G and M3-I. Our choice of m and n are very restricted because the original sample size $T = 125$. We choose two scenarios: (1) $m = 6$, $n = 20$, and (2) $m = 4$, $n = 30$ so that n is sufficiently large compared to m and thus $\sqrt{m}/n$ is near 0. As well, we selected lag 1 as the lag of the stationary part by the significant t-value. Here we did not consider

the GS method because the sample size of each interval is too small to choose a very large lag.

In Table 5.1, we reported the regression results using PROC Mixed in SAS software because it easily gives the various test statistics for the interval-specific models. As we expected, the intercept estimate is insignificant in both cases while the trend is significant. The t-values of the unit root coefficient are -3.02 and -3.19 for each case, which implies that the null hypothesis is rejected at the 5% level of the standard normal limiting distribution. Even if we use the finite sample critical values, the t-values are in the rejection region where the 5% critical value for $m = 5$ and $n = 20$, is -2.42. Thus, our M3-G model results are consistent with the three benchmarking tests in the Maddison data.

For M3-I, the interval-specific means are overall insignificant and the interval-specific trends are all significant. For the two scenarios, the t-values of the unit root coefficient estimate are -5.97 and -5.24, which illustrates that the bias of the estimates in the interval-specific parameter model depends on n. The values in the parentheses are the bias-corrected values using the formula in Theorem 3.5. The bias-corrected values are around the 5% critical value of the limiting distribution. We expect less power here as we are fitting a lot of parameters to a moderate size data set. One nice feature of the interval-specific model is that the mean and trend parameter estimates are seen to be fairly similar except for interval 4 so this approach is able to identify unusual intervals.

Quarterly US Real GDP

Murray and Nelson(2000) use the postwar quarterly real GDP data to perform the unit root test using a larger sample than the annual data, reasoning that the larger the sample size, the better power.

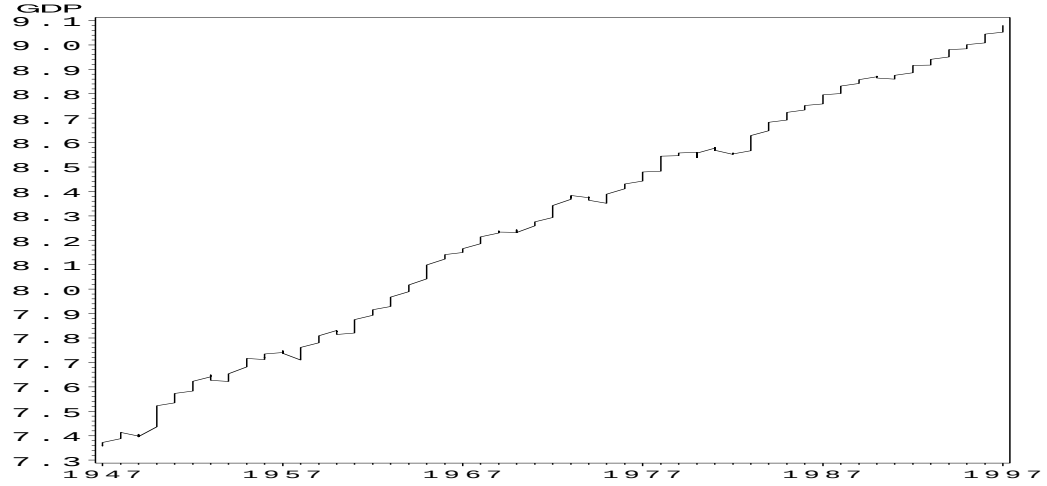
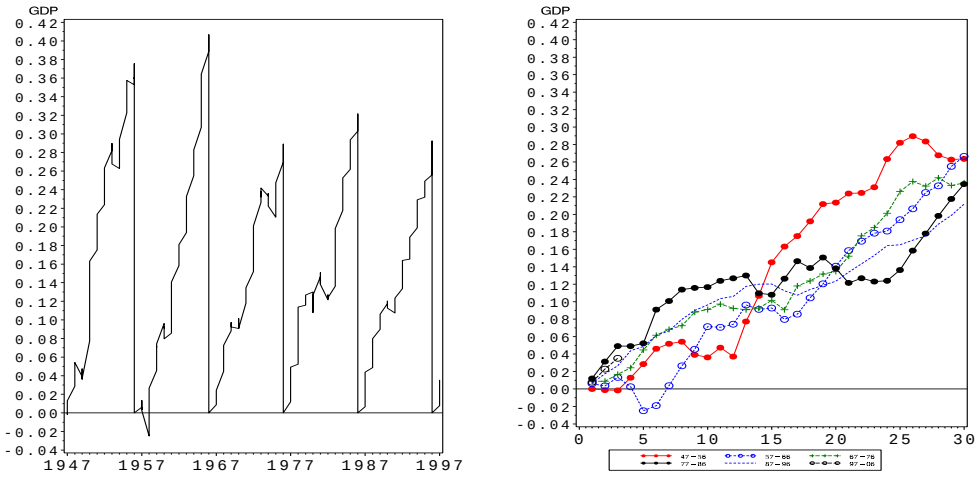


Figure 5.3: The Log Valued Quarterly US Real GDP: 1947:1-2006:4

Table 5.3: Unit Root Tests for Quarterly US Real GDP: M3-G

	Parameter	Estimate	Std Err	t Value	Pr > |t|
	Intercept	0.00392	0.0017	2.23	0.0266
m=5	Trend	0.0004	0.0002	2.37	0.0190
n=40	$y_{i,j1}$	-.0495	0.0212	**−2.33**	0.0207
	$\Delta y_{i,j-1}$	0.3714	0.0694	5.35	<.0001
	Intercept	0.0055	0.0020	2.78	0.0060
m=10	Trend	0.0004	0.0003	1.66	0.0938
n=20	$y_{i,j1}$	-.0593	0.0260	**−2.28**	0.0237
	$\Delta y_{i,j-1}$	0.3758	0.0741	5.07	<.0001

For this data, the three benchmarking tests strongly support, *i.e.,* are nowhere near rejecting, the unit root hypothesis where the Perron test captures the slope



(a) Over Years

(b) Within Interval

Figure 5.4: Transformation of Quarterly US GDP: m = 6, n = 40

Table 5.4: Unit Root Tests for Quarterly US Real GDP: M3-I

	Parameter	Estimate	Std Err	t Value	Pr > $\|t\|$
m=5	$\Delta y_{i,j-1}$	0.3870	0.0681	5.68	< .0001
n=40	$y_{i,j1}$	-.1631	0.0330	**-4.94**	< .0001
				(**−0.78**)	
m=10	$\Delta y_{i,j-1}$	0.3751	0.0722	5.19	< .0001
n=20	$y_{i,j1}$	-.3040	0.0471	**-6.45**	< .0001
				(**−0.41**)	

change in 1973:1 and the ZA test endogenously chooses the slope change in 1972:2. The DF-GLS test also favors the unit root hypothesis. In Table 5.3, for the combination of $(m, n) = (5, 40)$, our test using Model M3-G also does not reject the null hypothesis at the simulated 5% significant level, -2.39 although it is rejected at the asymptotic critical value, -1.645. We see that for $(10, 20)$, the t-value of -2.28 is below at the simulated 5% critical value, -2.10. However, we should notice that the estimated value of trend is insignificant, which implies that we may lose features of the data by over-dividing, as we mentioned in the introduction part of this chapter. Hence, we increase n until the trend estimate becomes statistically significant. As well, while this situation could have occured by chance, it is also possible that a break in the series had less of an effect on the transformed data than on the original data as our theory suggests.

Table 5.5: Variation in n with fixed m: M3-G

m=5,n=	38	39	40	41	42
	-2.56	-2.41	-2.32	-2.38	-2.02

In order to examine how sensitive the unit root test statistic is to the choice of n for fixed m we perform the test, varying the number of observations in each interval. In Table 5.5, we chose $m = 5$ and $n = 38$ through 42. The standard error for each case is approximately 0.02. The t-statistics seem to be sensitive to the interval size, n. This phenomenon is due to the small size of n. As n goes to ∞, the sensitivity must disappear. Also note that the $n = 38$ case omits 20 data points that are included in the $n = 42$ case.

5.2 A large sample case: US/Canada daily exchange rate

In this section we perform our unit root tests using the US/Canada daily exchange rate as an example of a larger sample. As in the previous section, we plot the whole series and the transformed series in Figure 5.5, 5.6 and 5.7. The whole sample period is January 4, 1971-May 19, 2006 with 8884 observations. We transformed the data with $n = 250$ which is roughly one year of observations.

Figure 5.5: Nominal US/Canada Exchange Rate: 1/4/1971-5/19/2006

The ADF test for the whole sample indicates strong evidence (nowhere near rejection) in favor of the unit root hypothesis. We have not found literature on the unit root test with structural change for the US/Canada exchange rate. Hence, we did not consider the Perron and ZA tests in this example.

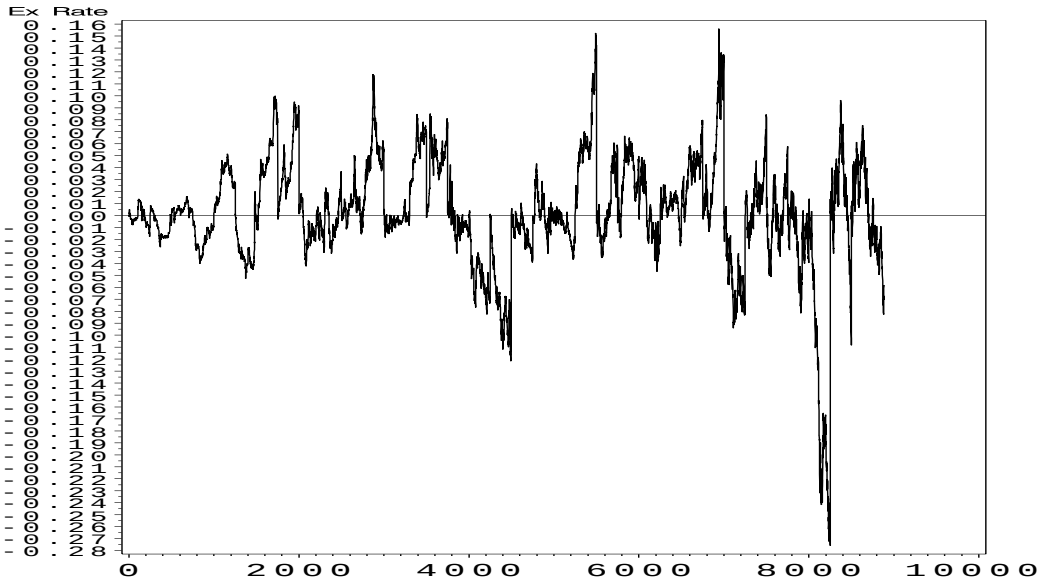


Figure 5.6: Transformation (a) of US/Canada Exchange Rate: m = 35, n = 250

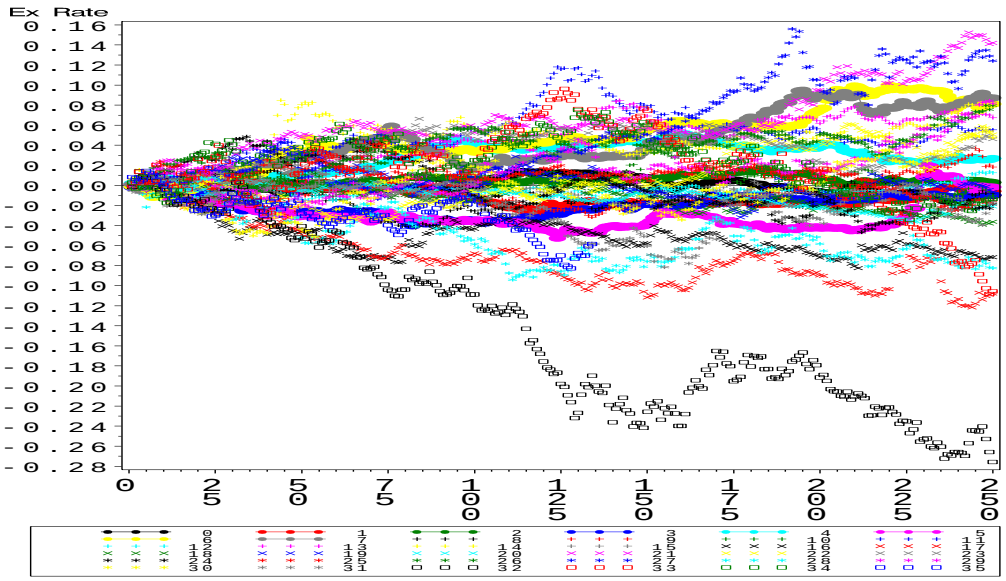


Figure 5.7: Transformation (b) of US/Canada Exchange Rate: m = 35, n = 250

Table 5.6: Unit Root Tests for US/Canada Exchange Rate

Model	(m,n)	Estimate	Std Err	t Value	Pr > |t|
	(177,50)	0.00092	0.00203	0.45	0.6526
	(88,100)	0.00080	0.00149	0.59	0.5542
M1	(35,250)	0.00155	0.00091	1.71	0.0867
		-.00006	0.00108	-0.05	0.9573
	(17,500)	0.00150	0.00056	2.67	0.0076
		0.00065	0.00056	0.91	0.3626
	(177,50)	-.06336	0.00363	-17.45	< .0001
				(-1.29)	
	(88,100)	-.03106	0.00265	-11.74	< .0001
M2-I				(-0.28)	
	(35,250)	-.00111	0.00165	-6.71	< .0001
				(0.60)	
	(17,500)	-.00493	0.00111	-4.47	< .0001
				(0.65)	

In Table 5.6, we present the test results using two models, M1 and M2-I, and four pairs of (m, n). In the M1 model, we see that the null is not rejected for $n = 50$ and 100, but is rejected for $n = 250$ and 500, which is not good news in the sense of robustness. Hence, we performed the two-sided ADF test for each interval, and found that we reject the null (in favor of the explosive alternative) in one interval without mean and trend, which is shown as an outlier in Figure 5.7. After removing this interval, we obtain results supporting the unit root for $n = 250$ and 500. In this example, the effect of one interval is not negligible, this is, the double-indexed transformation has not sufficiently reduced the effect of the contaminated interval using this n. Also, this phenomenon could be explained by the concept of locally stationary process as studied by Dahlhaus (1997) and Davis, Lee, and Rodriquez-Yam

(2006) where the former uses frequency and wavelet analysis to detect the locality, and the later adopts a genetic algorithm to estimate the break points. Another feature is that the estimates of the unit root coefficient are positive. Especially, for $n = 500$, the value of the estimate is 0.0015 and is significant. If this is true, then we should have seen in the plot that the process grows exponentially. Fortunately, as we increase m (decrease n) from 35 to 88 or 177, the effect of the outlier interval diminishes, and hence the tests support the unit root hypothesis, which is consistent with our theoretical and simulation results.

M2-I models strongly support the unit root hypothesis for every pair of (m, n) under the bias correction. In Table 5.6, we also confirm that the bias size of test statistics highly depends on the number of observations in each interval. As n becomes small, the t-value quickly decreases even though m gets large. The values in parentheses are bias-corrected using the formula in Theorem 3.3 where the correction was made by the asymptotic moments, which implies that the bias could be over-corrected in a finite sample, *e.g.*, in Table 5.6, the bias-corrected value is 0.60. In the table the computer generated $p-$values of M2-I are invalid since those are calculated under the assumption of the standard normal distribution.

Chapter 6

Conclusion

6.1 Findings

In this paper, our main goal was to devise a new unit root test robust to structural changes. To this end, we have introduced a panel data approach for the univariate case via a data transformation. As a key idea, we have first made the data transformation by double-indexing and last value subtraction such that $\{Y_t\}_{t=1}^{T} \mapsto \{y_{i,j} : 1 \leq i \leq m,\ 1 \leq j \leq n,\ mn = T\}$. This data transformation allows us to obtain the independent intervals for an AR (1) process and fast decaying covariance structure across intervals for a general ARMA (p,q) process.

Using the panel unit root approach in Levin and Lin(1992) and the sequential limit concept, we have found that the transformation leads to different asymptotic behavior than the standard unit root test as long as the number of intervals, m, grows. We have achieved asymptotic normality for the unit root test statistic using the growth rate $n\sqrt{m}$, which is slower than the growth rate T where the case of $m = 1$

reduces to the conventional Dickey-Fuller test.

Although both the Dickey-Fuller test and our approach asymptotically achieve power of 1, the simulation results showed that compared to the Dickey-Fuller test, in the finite sample the power of the unit root test in our framework is worse. However, in the presence of structural change our methodology has better power than the Dickey-Fuller test. Another advantage is that we do not need to detect a break point and measure a break size, unlike the conventional unit root tests considering structural changes.

6.2 Future Studies

When transforming the univariate data into m sets of n observations, we have chosen m and n theoretically as well as practically. However, we did not look at whether or not some optimal combination of (m, n) exists. Hence, further study may focus on this topic. Since we have employed the POLS estimation method, other estimation methods such as GMM can be considered. Also, the incidental parameters problem, which cause a serious bias of the estimate of the unit root coefficient could be extended beyond the cases we considered.

Bibliography

[1] Anderson, T.W., and Hsiao, C. (1981), "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47-82.

[2] Baltagi, B. H. (2005), *Econometric Analysis of Panel Data*, John Wiley & Sons, 3rd ed.

[3] Ben-David, D., Lumsdaine, R. L., and Papell, D. H. (2003), "Unit Roots, Post-war Slowdowns and and Long-Run Growth: Evidence from Two Structural Breaks," *Empirical Economics*, 28, 303-319.

[4] Benerjee, A., Lumsdaine, R. L., and Stock, J. H. (1992), "Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence," *Journal of Business and Economic Statistics*, 10, 271-289.

[5] Breitung, J., and Meyer, W. (1994), "Testing for Unit Roots in Panel Data: Are Wages on Different Bargaining Levels Cointegrated?," *Applied Econometrics*, 26, 353-361.

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

[7] _____ (1988), "Limiting Distributions of Least Square Estimates of Unstable Autoregressive Process," *The Annals of Statistics*, 16, 367-401.

[8] Chang, Y., and J.Y. Park (2004), "Taking a New Contour: A Novel View Unit Root Test," mimeographed, Department of Economics, Rice University

[9] Cook, S. (2001), "Asymmetric Unit Root Tests in the Presence of Structural Breaks under the Null," *Economics Bulletin*, 3, 1-10.

[10] _____ (2001), "Unit Root Testing in the Presence of Innovation Variance Breaks: A Simple Solution with Increased Power," *Journal of Applied Mathematics*, 2, 233-240.

[11] Dahlhaus, R. (1997), "Fitting Time Series Models to Nonstationary Process," *The Annals of Statistics*, 25, 1-37.

[12] Davidson, J. (1994), "*Stochastic Limit Theory,*" Oxford University Press, New York.

[13] Davidson, R. and MacKinnon, J. G. (1998), "Graphical Methods for Investigating the Size and Power of Hypothesis Tests," *Manchester School*, 66, 1-26.

[14] Davis, R.A., Lee, T.C.M, and Rodriquez-Yam, G.A. (2006), "Structural Break Estimation for Nonstationary Models," *Journal of the American Statistical Association*, 101, 229-239.

[15] Dickey, D. A. (1976), "Estimation and Hypothesis Testing in Nonstationary Time Series," Ph. D. thesis, Iowa State University, Ames, Iowa.

[16] Dickey, D.A. and Fuller, W. A. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statis-*

tical Association, 74, 427-431.

[17] Dickey, D.A., Hasza, D. P., and Fuller, W. A.(1984), "Testing for Unit Roots in Seasonal Time Series," *Journal of the American Statistical Association*, 79, 335-367.

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

[19] Enders, W. (2004), *Applied Econometric Time Series*, John Wiley & Sons, New York, 2nd ed.

[20] Evans, G. W. (1989), "Output and Unemployment Dynamics in the United States: 1950-1985," *Journal of Applied Econometrics*, 4, 213-237.

[21] Fuller, W. A. (1996), *Introduction to Statistical Time Series*, John Wiley & Sons, New York, 2nd ed.

[22] Fuller, W. A., and Park, H. J. (1995), "Alternative Estimators and Unit Root Tests for the Autoregressive Process," *Journal of Time Series Analysis*, 16, 415-429.

[23] Garcia, R. and Perron, P. (1996), "An Analysis of the Real Interest Rate Under Regime Shifts," *Review of Economics and Statistics*, 78, 111-125.

[24] Gonzalez-Farias, G.M. (1992), "A New Unit Root Test for Autoregressive Time Series," Ph. D. thesis, Department of Statistics, North Carolina State University.

[25] Hall, A. (1994), "Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection," *Journal of Business and Economic Statistics*, 12, 461-470.

[26] Hall, P., and Heyde, C. C. (1980), *Martingale Limit Theorem and its Application*, Academic Press, New York.

[27] Hamilton, J. D. (1994), "Time Series Analysis," Princeton University Press, Princeton, New Jersey

[28] Harris, Richard D.F., and Tzavalis, E. (1999), "Inference for Unit Roots in Dynamic Panels Where the Time Dimension is Fixed," *Journal of Econometrics*, 91, 201-226.

[29] Hsiao, C. (2002), *Analysis of Panel Data*, Cambridge University Press, 2 ed.

[30] Hylleberg, S., Engle, R. F., Granger, C. W. J., and Yoo, B. S. (1990), "Seasonal Integration and Cointegration," *Journal of Econometrics*, 44, 215-238.

[31] Im, K., Pesaran, H., and Shin, Y. (2003), "Testing for Unit Roots in Heterogenous Panels," *Journal of Econometrics*, 115, 53-74.

[32] Kanas, A. (1998), "Testing for a Unit Root in ERM Exchange Rates in the Presence of Structural Breaks: Evidence from the Bootstrap," *Applied Economics Letters*, 5, 407-410.

[33] Kapetanios, G. (2005) "Unit-Root Testing against the Alternative Hypothesis of up to m Structural Breaks," *Journal of Time Series Analysis*, 26, 123-133.

[34] Kim, T.-H, Leybourne, S., and Newbold, P. (2002), "Unit Root Tests with a Break in Innovation Variance," *Journal of Econometrics*, 109, 365-387.

[35] Kunst, R. M. (1997), "Testing for Cyclical Non-stationarity in Autoregressive Processes," *Journal of Time Series Analysis*, 18, 123-135.

[36] Lai, T.L., Siegmund, D. (1983), "Fixed Accuracy Estimation of an Autoregressive Parameter," *The Annals of Statistics*, 11, 478-485.

[37] Lee, J. (2000), "On the End-point Issue in Unit Root Test in the Presence of a Structural Break," *Economic Letters*, 68, 7-11.

[38] Levin, A., and Lin, C. F. (1992), "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties," Discussion Paper 92-93, Department of Economics, University of California at San Diego.

[39] _____ (1993), "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties," *mimeo*, Department of Economics, University of California, San Diego.

[40] Levin, A., Lin, C. F., and Chu, C.-S. (2002), "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties," *Journal of Econometrics*, 108, 1-24.

[41] Leybourne, S.J., Mills, T.C., and Newbold, P. (1998), "Spurious Rejections by Dickey-Fuller Tests in the Presense of a Break under the Null," *Journal of Econometrics*, 87, 191-203.

[42] Lumsdaine, R.L., and Papell, D. H. (1997), "Multiple Trend Breaks and the Unit Root Hypothesis," *The Review of Economics and Statistics*, 79, 212-218.

[43] Lütkepohl, H., Müller, C., and Saikkonen, P. (2001), "Unit Root Tests for Time Series with a Structural Break when the Break Point is Known," in Hsiao, C., Morimune, K., and Powell, J. (eds) *Nonlinear Statistical Modeling: Essays in Honor of Takeshi Amemiya*, Chapter 12, Cambridge University Press,

Cambridge, 327-348.

[44] Maddala, G. S. and Kim, I-. M. (1998), "*Unit roots, Cointegration, and Structural Change*," Chambridge University Press.

[45] Maddala, G. S. and S. Wu (1996), "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test," *Oxford Bulletin of Economics and Statistics*, 61, 631-652.

[46] Maddison, A. (1995), "Monitoring the World Economy 1820-1992," OECD, Paris.

[47] Montañés, A., and Olloqui, I. (1999), "Misspecification of the breaking date in segmented trend variables: effect on the unit root tests," *Economics Letters*, 65, 301-307.

[48] Murray, C. J. and Nelson, C. R. (2000), "The Uncertain Trend in U.S. GDP," *Journal of Monetary Economics*, 46, 79-95.

[49] Nabeya, S. (1999), "Asymptotic Moments of Some Unit Root Test Statistics in the Null Case," *Econometric Theory*, 15, 139-149.

[50] Nelson, C. R., and Plosser, C. (1982), "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics* 10, 139-162.

[51] Ng, S., and Perron, P. (1995), "Unit Root Test in ARMA Models with Data Dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association*, 90, 268-281.

[52] Nickell, S. (1981), "Biases in Dynamic Models with Fixed Effects," *Econometrica*,

49, 1399-1416.

[53] Pantula, S., Gonzalez-Farias, G., and Fuller, W. A. (1994), "A comparison of unit root test criteria," *Journal of Business and Economic Statistics*, 12, 449-459.

[54] Park, H. J., and Fuller, W. A. (1995), "Alternative Estimators and Unit Root Tests for the Autoregressive Process," *Journal of Time Series Analysis*, 16, 415-429.

[55] Perron, P. (1989), "The great crash, the oil-price shock, and the unit-root hypothesis," *Econometrica*, 57, 1361-1402.

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

[57] Phillips, P. C. B., and Moon, H. R. (1999), "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67, 1057-1111.

[58] Phillips, P. C. B., Moon, H. R., and Xiao, Z. (2001), "How to Estimate Autoregressive Roots Near Unity," *Econometric Theory*, 17, 29-69.

[59] Quah, D., (1994), "Exploiting Cross-Section VariatioN for Unit Root Inference in Dynamic data," *Economics Letters*, 44, 9-19.

[60] Saikkonen, P., and Lü*tkepohl*, H. (2002), "Testing for a Unit Root in a Time Series with a Level Shift at Unknown Time," *Econometric Theory*, 18, 313-348.

[61] Schimidt, P. and Phillips, P.C.B. (1992), "Testing for a Unit Root in the Presence of Deterministic Trend," *Oxford Bulletin of Economics and Statistics*, 54,

257-288.

[62] Schotman, P. C., and van Dijk, H. K. (1991), "On Bayesian Rourtes to Unit Roots," *Journal of Applied Econometrics*, 6, 387-401.

[63] Sims, C. Stock, J., and Watson, M. (1990), "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58, 113-144.

[64] Zivot, E. and D. W. K. Andrews (1992), "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis," *Journal of Business and Economic Statistics*, 10, 251-270.

Chapter 7

Appendices

7.1 Critical Values and Power of Basic Models

Table 7.1: Critical values of ρ and t-statistics in M1

T	m	$\rho(0.01)$	$\rho(0.05)$	$\rho(0.10)$	$t(0.01)$	$t(0.05)$	$t(0.10)$
	1	-13.17	-7.74	-5.63	-2.58	-1.92	-1.61
100	5	-6.49	-4.13	-3.00	-2.49	-1.84	-1.48
	10	-5.34	-3.14	-2.59	-2.47	-1.76	-1.41
	20	-4.62	-3.12	-2.40	-2.41	-1.71	-1.36
	1	-13.32	-7.89	-5.60	-2.55	-1.93	-1.60
	5	-6.55	-4.11	-3.03	-2.47	-1.82	-1.49
250	10	-5.33	-3.43	-2.56	-2.47	-1.77	-1.42
	25	-4.53	-3.01	-2.28	-2.42	-1.73	-1.35
	50	-4.19	-2.96	-2.27	-2.37	-1.71	-1.34
	1	-13.63	-8.00	-5.71	-2.57	-1.93	-1.61
	5	-6.68	-4.15	-3.04	-2.52	-1.84	-1.49
500	10	-5.44	-3.44	-2.54	-2.47	-1.79	-1.47
	25	-4.61	-2.92	-2.23	-2.47	-1.71	-1.35
	50	-4.13	-2.83	-2.16	-2.38	-1.72	-1.34
	100	-4.10	-2.84	-2.19	-2.39	-1.70	-1.32
	1	-13.56	-7.95	-5.66	-2.55	-1.94	-1.61
	5	-6.80	-4.23	-3.09	-2.54	-1.88	-1.52
	10	-5.59	-3.47	-2.62	-2.54	-1.80	-1.46
	25	-4.36	-2.94	-2.24	-2.39	-1.73	-1.38
2500	50	-4.04	-2.73	-2.09	-2.37	-1.71	-2.09
	100	-3.84	-2.61	-2.02	-2.36	-1.67	-1.32
	250	-3.67	-2.58	-2.01	-2.34	-1.65	-1.30
	500	-3.86	-2.72	-2.11	-2.34	-1.68	-1.31

Table 7.2: Critical values of ρ_μ and t_μ-statistics in M2-G

T	m	$\rho(0.01)$	$\rho(0.05)$	$\rho(0.10)$	$t(0.01)$	$t(0.05)$	$t(0.10)$
	1	-19.12	-13.39	-10.79	-3.50	-2.87	-2.57
100	5	-7.98	-5.49	-4.19	-2.92	-2.29	-1.91
	10	-6.23	-4.07	-3.16	-2.77	-2.03	-1.67
	20	-5.04	-3.50	-2.72	-2.57	-1.88	-1.52
	1	-20.09	-13.80	-11.01	-3.46	-2.87	-2.55
	5	-8.17	-5.44	-4.17	-2.94	-2.27	-1.90
250	10	-6.12	-4.10	-3.14	-2.74	-2.05	-1.68
	25	-4.93	-3.32	-2.57	-2.58	-1.88	-1.51
	50	-4.41	-3.15	-2.46	-2.47	-1.82	-1.45
	1	-20.17	-13.73	-11.04	-3.43	-2.85	-2.56
	5	-8.39	-5.63	-4.22	-2.96	-2.29	-1.92
500	10	-6.35	-4.13	-3.15	-2.77	-2.06	-1.68
	25	-5.00	-3.24	-2.52	-2.64	-1.87	-1.51
	50	-4.36	-3.02	-2.34	-2.51	-1.82	-1.44
	100	-4.24	-2.96	-2.30	-2.46	-1.76	-1.39
	1	-20.23	-14.01	-11.22	-3.46	-2.86	-2.56
	5	-8.55	-5.69	-4.27	-2.97	-2.31	-1.94
	10	-6.40	-4.15	-3.21	-2.80	-2.07	-1.71
	25	-4.69	-3.29	-2.56	-2.57	-1.91	-1.54
2500	50	-4.25	-2.92	-2.27	-2.47	-1.81	-1.45
	100	-3.99	-2.73	-2.14	-2.44	-1.74	-1.39
	250	-3.75	-2.65	-2.09	-2.37	-1.70	-1.35
	500	-1.39	-2.76	-2.15	-2.38	-1.71	-1.34

Table 7.3: Critical values of t_{μ_i} and $t^{BC}_{\mu_i}$-statistics in M2-I

T	m	$t^B_{\mu_i}(.01)$	.05	.10	$t^{\Delta}_{\mu_i}(.01)$	.05	.10
	1	-3.57	-2.87	-2.57	-2.62	-1.84	-1.50
100	5	-4.98	-4.38	-4.04	-2.50	-1.84	-1.46
	10	-6.13	-5.49	-5.13	-2.53	-1.81	-1.41
	20	-7.92	-7.12	-6.69	-2.74	-1.83	-1.36
	1	-3.43	-2.85	-2.56	-2.47	-1.81	-1.49
250	5	-4.94	-4.34	-4.01	-2.46	-1.79	-1.43
	10	-6.09	5.43	-5.12	-2.47	-1.75	-1.40
	25	-8.31	-7.66	-7.28	-2.45	-1.71	-1.30
	50	-10.81	-10.05	-9.59	-2.40	-1.56	-1.04
	1	-3.34	-2.83	-2.53	-2.37	-1.79	-1.46
	5	-4.91	-4.30	-3.98	-2.43	-1.75	-1.39
500	10	-5.98	-5.44	-5.12	-2.36	-1.75	-1.39
	25	-8.20	-7.63	-7.30	-2.30	-1.69	-1.31
	50	-10.74	-10.11	-9.78	-2.33	-1.62	-1.25
	100	-14.20	-13.37	-12.98	-2.18	-1.25	-0.82
	1	-3.44	-2.85	-2.57	-2.47	-1.82	-1.51
	5	-4.89	-4.33	-4.01	-2.40	-1.78	-1.43
2500	10	-6.01	-5.45	-5.11	-2.39	-1.76	-1.38
	25	-8.23	-7.70	-7.38	-2.36	-1.78	-1.40
	50	-10.80	-10.18	-9.84	-2.39	-1.70	-1.32
	100	-14.32	-13.64	-13.35	-2.31	-1.56	-1.24
	250	-21.10	-20.49	-20.14	-1.94	-1.26	-0.87
	500	-28.38	-27.62	-27.25	-1.11	-0.26	0.15

Table 7.4: Critical values of ρ_τ and t_τ-statistics in M3-G

T	m	$\rho(0.01)$	$\rho(0.05)$	$\rho(0.10)$	$t(0.01)$	$t(0.05)$	$t(0.10)$
	1	-27.07	-20.21	-17.27	-4.10	-3.45	-3.16
100	5	-8.98	-5.95	-4.57	-3.12	-2.42	-2.02
	10	-6.48	-4.22	-3.26	-2.95	-2.18	-1.78
	20	-5.15	-3.56	-2.74	-2.91	-2.13	-1.71
	1	-27.88	-21.19	-17.78	-3.96	-3.43	-3.14
	5	-9.34	-5.99	-4.52	-3.12	-2.38	-1.98
250	10	-6.35	-4.24	-3.24	-2.84	-2.12	-1.73
	25	-4.95	-3.36	-2.60	-2.75	-1.99	-1.60
	50	-4.44	-3.17	-2.47	-2.77	-2.03	-1.62
	1	-29.16	-21.50	-17.82	-3.98	-3.41	-3.12
	5	-9.44	-6.15	-4.63	-3.10	-2.40	-2.00
500	10	-6.62	-4.29	-3.23	-2.85	-2.11	-1.72
	25	-5.06	-3.28	-2.54	-2.71	-1.93	-1.55
	50	-4.35	-3.04	-2.35	-2.65	-1.91	-1.52
	100	-4.24	-2.97	-2.31	-2.74	-1.97	-1.56
	1	-29.02	-21.41	-18.03	-3.93	-3.41	-3.12
	5	-9.50	-6.21	-4.69	-3.06	-2.39	-2.01
	10	-6.68	-4.33	-3.30	-2.87	-2.11	-1.74
2500	25	-4.78	-3.33	-2.57	-2.58	-1.92	-1.55
	50	-4.26	-2.93	-2.27	-2.51	-1.82	-1.46
	100	-4.00	-2.75	-2.14	-2.50	-1.78	-1.42
	250	-3.74	-2.66	-2.09	-2.49	-1.79	-1.42
	500	-3.89	-2.76	-2.16	-2.66	-1.91	-1.50

Table 7.5: Critical values of t_{τ_i} and $t^{BC}_{\tau_i}$-statistics in M3-I

T	m	$t^{B}_{\tau_i}(.01)$	.05	.10	$t^{\Delta}_{\tau_i}(.01)$	.05	.10
	1	-4.05-3.45	-3.14	-2.69	-1.92	-1.53	
100	5	-6.57	-5.90	-5.56	-2.84	-1.99	-1.57
	10	-8.59	-7.80	-7.44	-3.14	-2.13	-1.68
	20	-11.51	-10.54	-10.04	-3.62	-2.39	-1.76
	1	-3.88	-3.39	-3.11	-2.48	-1.85	-1.49
250	5	-6.42	-5.77	-5.47	-2.65	-1.83	-1.45
	10	-8.14	-7.57	-7.29	-2.56	-1.84	-1.48
	25	-12.02	-11.26	-10.86	-2.97	-2.01	-1.50
	50	-16.41	-15.35	-14.83	-3.46	-2.11	-1.44
	1	-3.93	-3.40	-3.11	-2.54	-1.87	-1.49
	5	-6.23	-5.73	-5.44	-2.42	-1.78	-1.41
500	10	-8.10	-7.55	-7.25	-2.51	-1.81	-1.43
	25	-11.72	-11.09	-10.81	-2.59	-1.79	-1.43
	50	-15.92	-15.17	-14.79	-2.84	-1.88	-1.40
	100	-21.54	-20.72	-20.24	-2.76	-1.72	-1.12
	1	-3.91	-3.41	-3.12	-2.51	-1.87	-1.51
	5	-6.33	-5.71	-5.43	-2.54	-1.76	-1.41
2500	10	-7.93	-7.46	-7.21	-2.30	-1.70	-1.38
	25	-11.55	-11.03	-10.74	-2.38	-1.71	-1.34
	50	-15.62	-15.04	-14.76	-2.45	-1.71	-1.36
	100	-21.32	-20.76	-20.41	-2.48	-1.71	-1.33
	250	-32.44	-31.79	-31.42	-2.31	-1.49	-1.02
	500	-44.51	-43.69	-43.19	-1.54	-0.49	0.14

Table 7.6: Empirical Power of t-statistic in M1

T	m	$\rho = 0.8$			$\rho = 0.9$			$\rho = 0.95$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
100	1	0.93	1.00	1.00	0.32	0.78	0.93	0.08	0.33	0.54
	5	0.42	0.72	0.84	0.14	0.37	0.53	0.05	0.18	0.31
	10	0.28	0.57	0.71	0.09	0.26	0.40	0.04	0.14	0.23
	20	0.16	0.39	0.53	0.06	0.18	0.29	0.03	0.11	0.19
250	1	1.00	1.00	1.00	0.99	1.00	1.00	0.51	0.91	0.98
	5	0.94	0.99	1.00	0.58	0.83	0.91	0.21	0.48	0.63
	10	0.88	0.98	0.99	0.40	0.70	0.82	0.13	0.34	0.49
	25	0.66	0.88	0.94	0.21	0.46	0.61	0.06	0.20	0.33
	50	0.40	0.66	0.79	0.11	0.29	0.42	0.04	0.14	0.24
500	1	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00
	5	1.00	1.00	1.00	0.93	0.99	1.00	0.55	0.82	0.91
	10	1.00	1.00	1.00	0.88	0.97	0.99	0.40	0.70	0.83
	25	0.99	1.00	1.00	0.65	0.88	0.94	0.20	0.48	0.62
	50	0.94	0.99	1.00	0.42	0.69	0.81	0.12	0.30	0.44
	100	0.71	0.90	0.95	0.45	0.72	0.83	0.06	0.19	0.31
2500	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	50	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00
	100	1.00	1.00	1.00	1.00	1.00	1.00	0.92	0.98	0.99
	250	1.00	1.00	1.00	0.99	1.00	1.00	0.58	0.81	0.89
	500	1.00	1.00	1.00	0.86	0.96	0.98	0.28	0.52	0.67

Table 7.7: Empirical Power of t-statistic in M2-G

		$\rho = 0.8$			$\rho = 0.9$			$\rho = 0.95$		
T	m	1%	5%	10%	1%	5%	10%	1%	5%	10%
100	1	0.51	0.88	0.96	0.08	0.32	0.51	0.02	0.12	0.22
	5	0.28	0.54	0.70	0.07	0.23	0.37	0.03	0.10	0.19
	10	0.19	0.46	0.61	0.05	0.19	0.31	0.02	0.10	0.17
	20	0.13	0.33	0.47	0.04	0.15	0.24	0.02	0.08	0.15
250	1	1.00	1.00	1.00	0.74	0.97	1.00	0.14	0.45	0.68
	5	0.86	0.97	0.99	0.41	0.70	0.82	0.11	0.32	0.48
	10	0.81	0.95	0.98	0.30	0.59	0.74	0.08	0.25	0.39
	25	0.60	0.84	0.92	0.16	0.40	0.55	0.05	0.17	0.28
	50	0.36	0.62	0.75	0.10	0.25	0.38	0.04	0.12	0.21
500	1	1.00	1.00	1.00	1.00	1.00	1.00	0.74	0.97	0.99
	5	1.00	1.00	1.00	0.85	0.96	0.99	0.40	0.69	0.82
	10	1.00	1.00	1.00	0.80	0.95	0.98	0.30	0.60	0.75
	25	0.99	1.00	1.00	0.59	0.85	0.92	0.16	0.42	0.56
	50	0.93	0.99	1.00	0.38	0.65	0.78	0.09	0.27	0.41
	100	0.69	0.88	0.94	0.18	0.42	0.57	0.05	0.18	0.29
2500	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	50	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00
	100	1.00	1.00	1.00	1.00	1.00	1.00	0.90	0.98	0.99
	250	1.00	1.00	1.00	0.99	1.00	1.00	0.56	0.80	0.88
	500	1.00	1.00	1.00	0.85	0.96	0.98	0.26	0.51	0.66

Table 7.8: Empirical Power of t-statistic in M3-G

		$\rho = 0.8$			$\rho = 0.9$			$\rho = 0.95$		
T	m	1%	5%	10%	1%	5%	10%	1%	5%	10%
100	1	0.28	0.67	0.83	0.04	0.20	0.34	0.01	0.09	0.16
	5	0.25	0.52	0.68	0.07	0.22	0.35	0.02	0.10	0.19
	10	0.18	0.44	0.60	0.05	0.18	0.30	0.02	0.09	0.17
	20	0.12	0.32	0.46	0.04	0.14	0.24	0.02	0.08	0.15
250	1	1.00	1.00	1.00	0.51	0.84	0.94	0.08	0.27	0.44
	5	0.83	0.96	0.99	0.36	0.67	0.81	0.09	0.30	0.46
	10	0.80	0.95	0.98	0.29	0.59	0.73	0.08	0.25	0.39
	25	0.59	0.84	0.92	0.16	0.40	0.55	0.04	0.17	0.28
	50	0.36	0.62	0.75	0.09	0.25	0.38	0.04	0.12	0.21
500	1	1.00	1.00	1.00	1.00	1.00	1.00	0.46	0.83	0.94
	5	0.99	1.00	1.00	0.83	0.96	0.99	0.36	0.65	0.80
	10	1.00	1.00	1.00	0.79	0.95	0.98	0.28	0.59	0.74
	25	0.99	1.00	1.00	0.59	0.85	0.92	0.16	0.41	0.56
	50	0.92	0.99	1.00	0.37	0.65	0.78	0.09	0.27	0.40
	100	0.67	0.88	0.94	0.18	0.42	0.56	0.05	0.18	0.28
2500	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	50	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00
	100	1.00	1.00	1.00	1.00	1.00	1.00	0.90	0.98	0.99
	250	1.00	1.00	1.00	0.99	1.00	1.00	0.57	0.80	0.88
	500	1.00	1.00	1.00	0.85	0.96	0.98	0.26	0.51	0.65

Table 7.9: Empirical Power of t-statistic in M2-I

		$\rho = 0.8$			$\rho = 0.9$			$\rho = 0.95$		
T	m	1%	5%	10%	1%	5%	10%	1%	5%	10%
100	1	0.46	0.87	0.96	0.07	0.32	0.51	0.02	0.12	0.22
	5	0.15	0.39	0.58	0.04	0.16	0.27	0.02	0.09	0.17
	10	0.10	0.28	0.43	0.04	0.14	0.23	0.02	0.09	0.16
	20	0.05	0.19	0.31	0.03	0.11	0.19	0.02	0.08	0.14
250	1	1.00	1.00	1.00	0.74	0.97	1.00	0.14	0.46	0.66
	5	0.92	0.99	1.00	0.22	0.55	0.74	0.05	0.18	0.33
	10	0.64	0.91	0.97	0.12	0.36	0.53	0.04	0.15	0.26
	25	0.27	0.56	0.72	0.07	0.21	0.34	0.03	0.11	0.20
	50	0.13	0.34	0.51	0.04	0.14	0.26	0.02	0.09	0.17
500	1	1.00	1.00	1.00	1.00	1.00	1.00	0.79	1.00	1.00
	5	1.00	1.00	1.00	0.91	1.00	1.00	0.21	0.56	0.74
	10	1.00	1.00	1.00	0.65	0.89	0.96	0.13	0.35	0.52
	25	0.93	0.99	1.00	0.27	0.54	0.69	0.07	0.21	0.33
	50	0.60	0.83	0.90	0.14	0.33	0.47	0.04	0.15	0.25
	100	0.28	0.57	0.71	0.07	0.23	0.35	0.03	0.12	0.20
2500	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	25	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00
	50	1.00	1.00	1.00	1.00	1.00	1.00	0.76	0.93	0.98
	100	1.00	1.00	1.00	0.99	1.00	1.00	0.42	0.72	0.82
	250	1.00	1.00	1.00	0.71	0.89	0.94	0.19	0.41	0.54
	500	0.96	0.99	1.00	0.37	0.65	0.77	0.09	0.26	0.38

Table 7.10: Empirical Power of t-statistic in M3-I

		$\rho = 0.8$			$\rho = 0.9$			$\rho = 0.95$		
T	m	1%	5%	10%	1%	5%	10%	1%	5%	10%
100	1	0.28	0.65	0.82	0.04	0.19	0.33	0.02	0.09	0.16
	5	0.03	0.15	0.30	0.01	0.07	0.14	0.01	0.06	0.11
	10	0.02	0.09	0.16	0.01	0.06	0.12	0.01	0.05	0.10
	20	0.01	0.08	0.12	0.01	0.06	0.10	0.01	0.05	0.10
250	1	1.00	1.00	1.00	0.54	0.85	0.95	0.08	0.27	0.45
	5	0.45	0.80	0.90	0.05	0.22	0.36	0.02	0.09	0.16
	10	0.16	0.41	0.56	0.03	0.12	0.20	0.02	0.06	0.13
	25	0.03	0.12	0.22	0.01	0.06	0.12	0.01	0.05	0.11
	50	0.01	0.07	0.15	0.01	0.06	0.11	0.01	0.05	0.10
500	1	1.00	1.00	1.00	1.00	1.00	1.00	0.49	0.83	0.94
	5	1.00	1.00	1.00	0.49	0.78	0.90	0.07	0.23	0.36
	10	0.89	0.98	1.00	0.14	0.39	0.54	0.03	0.12	0.21
	25	0.23	0.51	0.65	0.03	0.14	0.23	0.02	0.07	0.13
	50	0.05	0.18	0.31	0.02	0.08	0.14	0.01	0.06	0.11
	100	0.02	0.09	0.17	0.01	0.06	0.12	0.01	0.05	0.10
2500	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	25	1.00	1.00	1.00	0.99	1.00	1.00	0.44	0.72	0.85
	50	1.00	1.00	1.00	0.81	0.95	0.98	0.10	0.29	0.43
	100	0.99	1.00	1.00	0.21	0.48	0.63	0.03	0.12	0.21
	250	0.30	0.56	0.70	0.04	0.14	0.22	0.02	0.07	0.13
	500	0.06	0.17	0.29	0.02	0.07	0.14	0.01	0.05	0.11

7.2 Histogram of Statistics

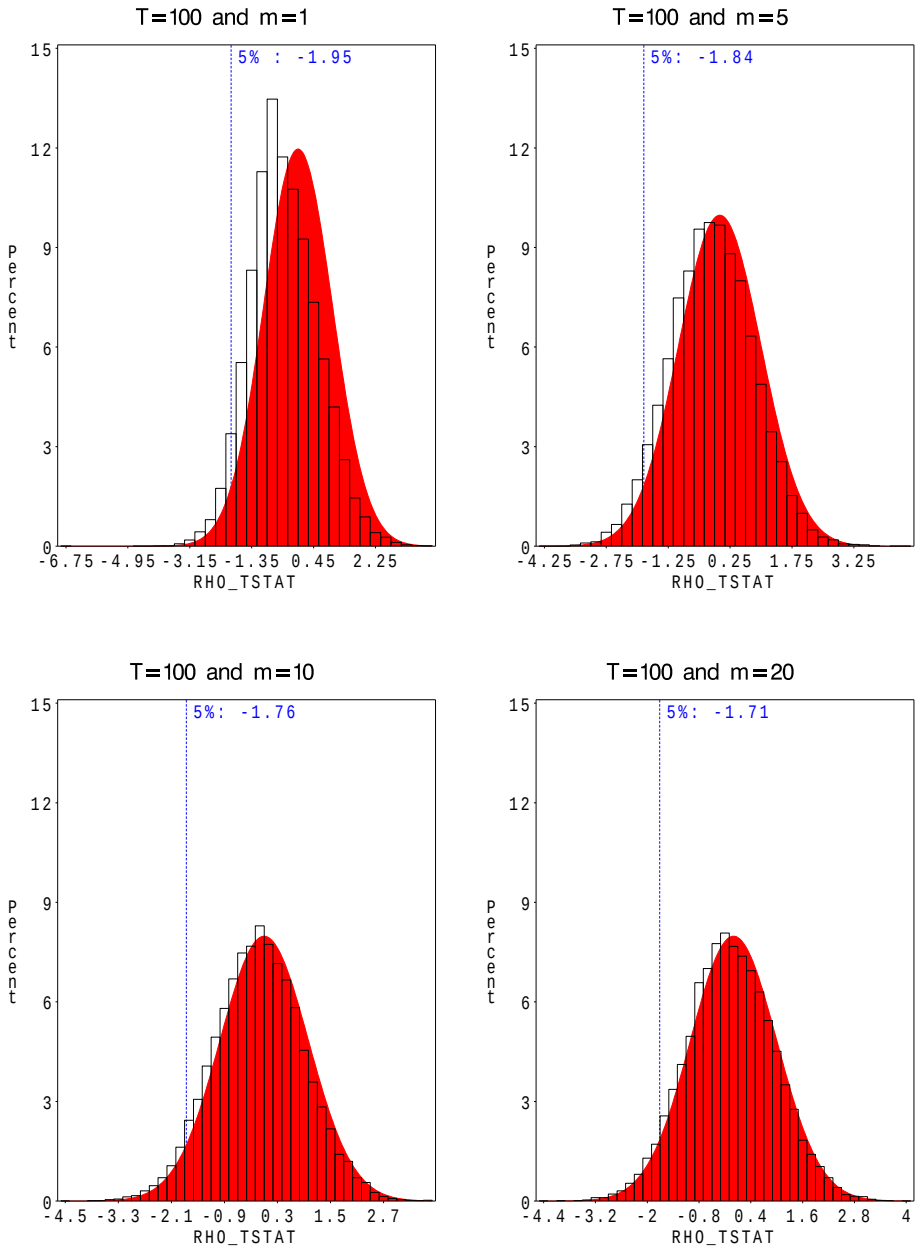


Figure 7.1: Density of t-statistic in M1 (T=100)

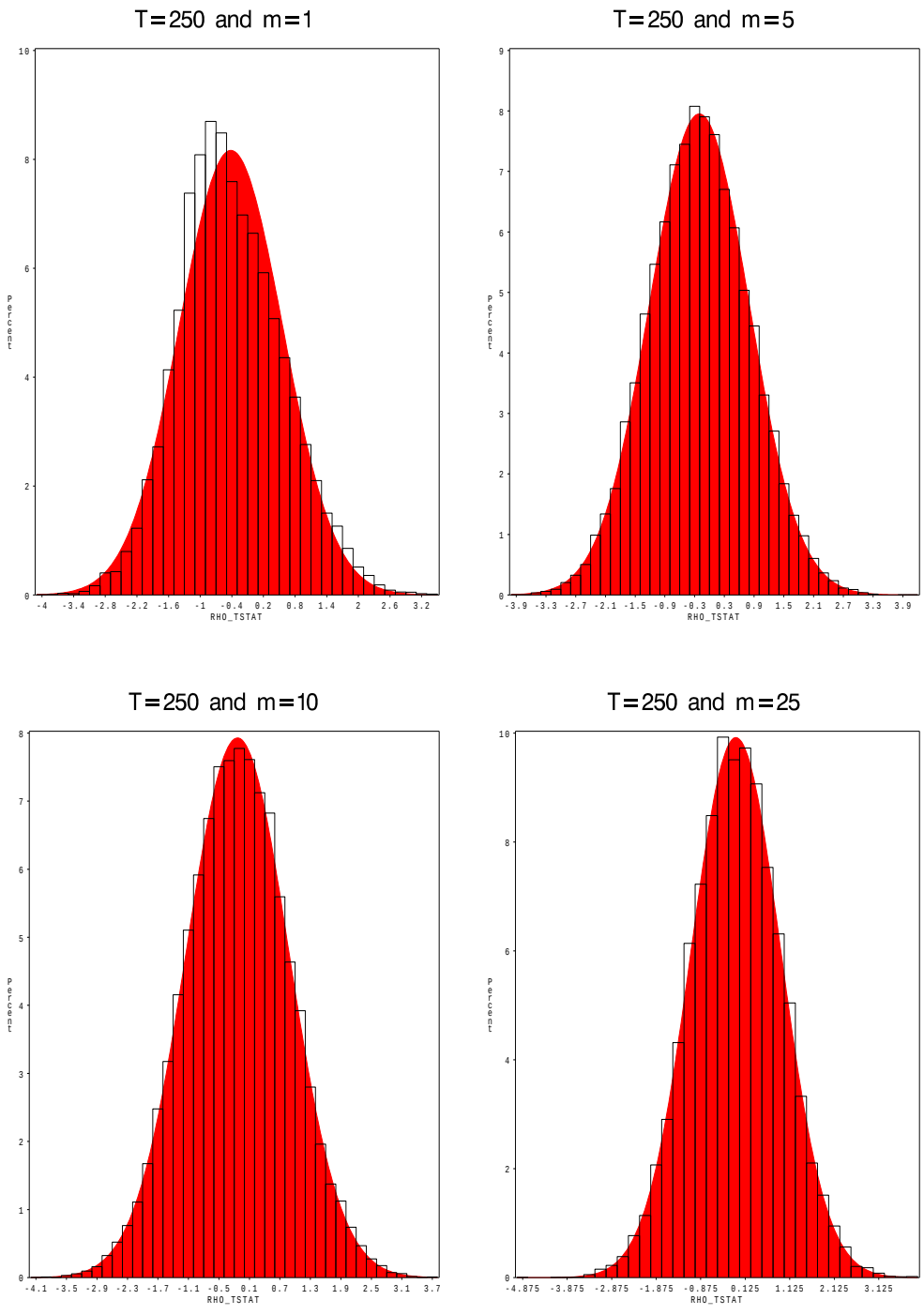


Figure 7.2: Density of t-statistic in M1 (T=250)

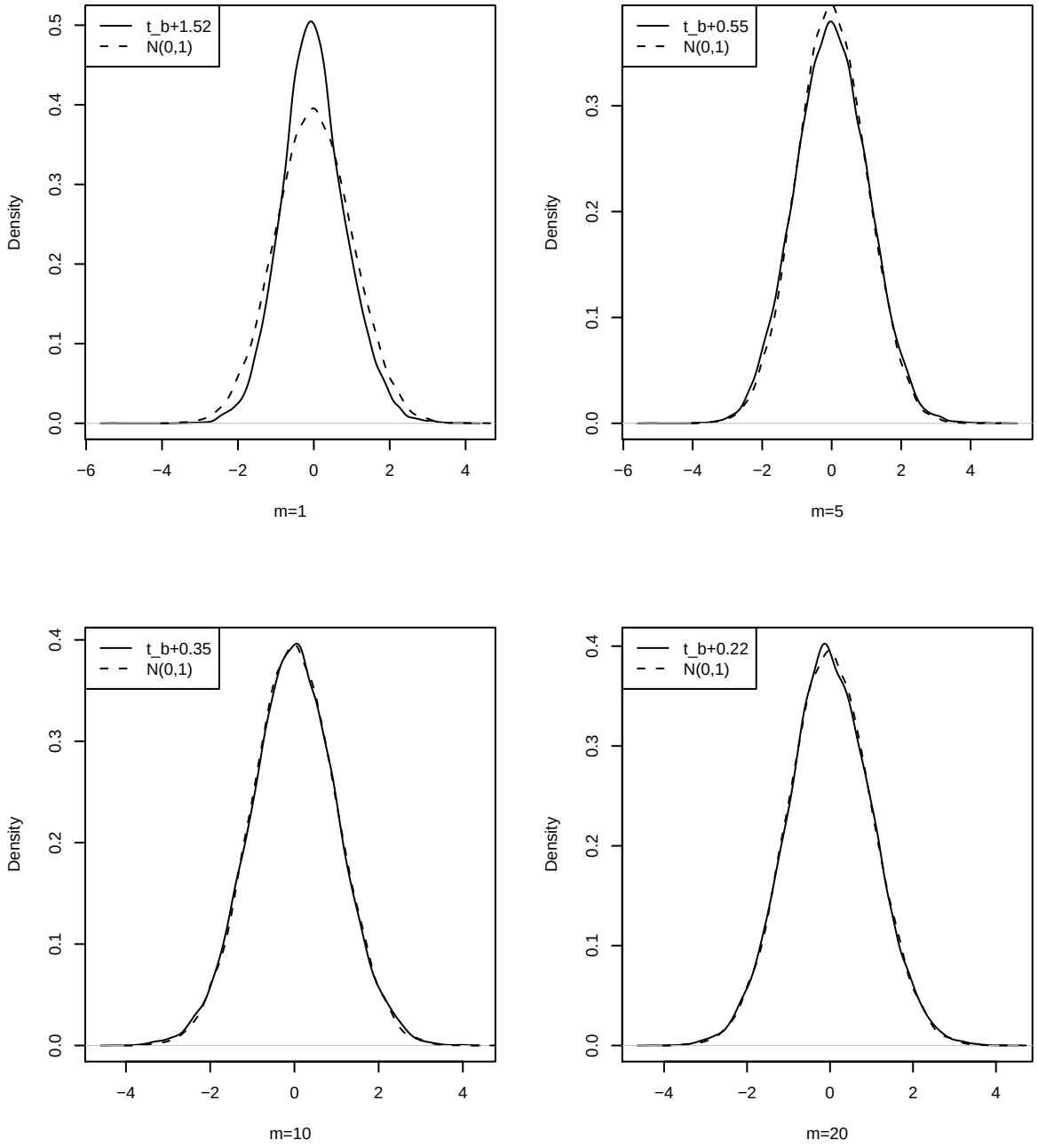


Figure 7.3: Density of t-statistic in M2-G (T=100)

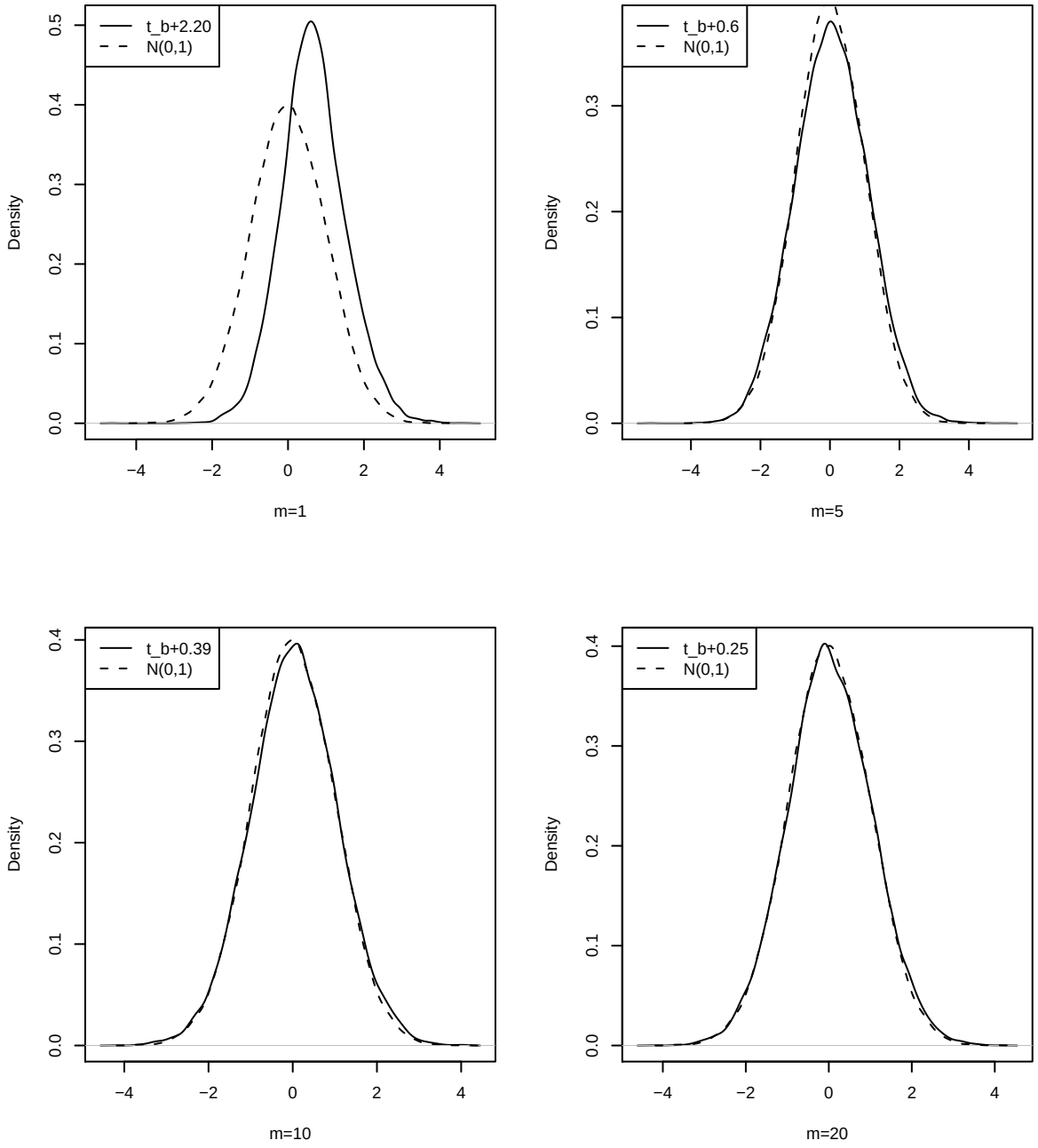


Figure 7.4: Density of t-statistic in M3-G (T=100)

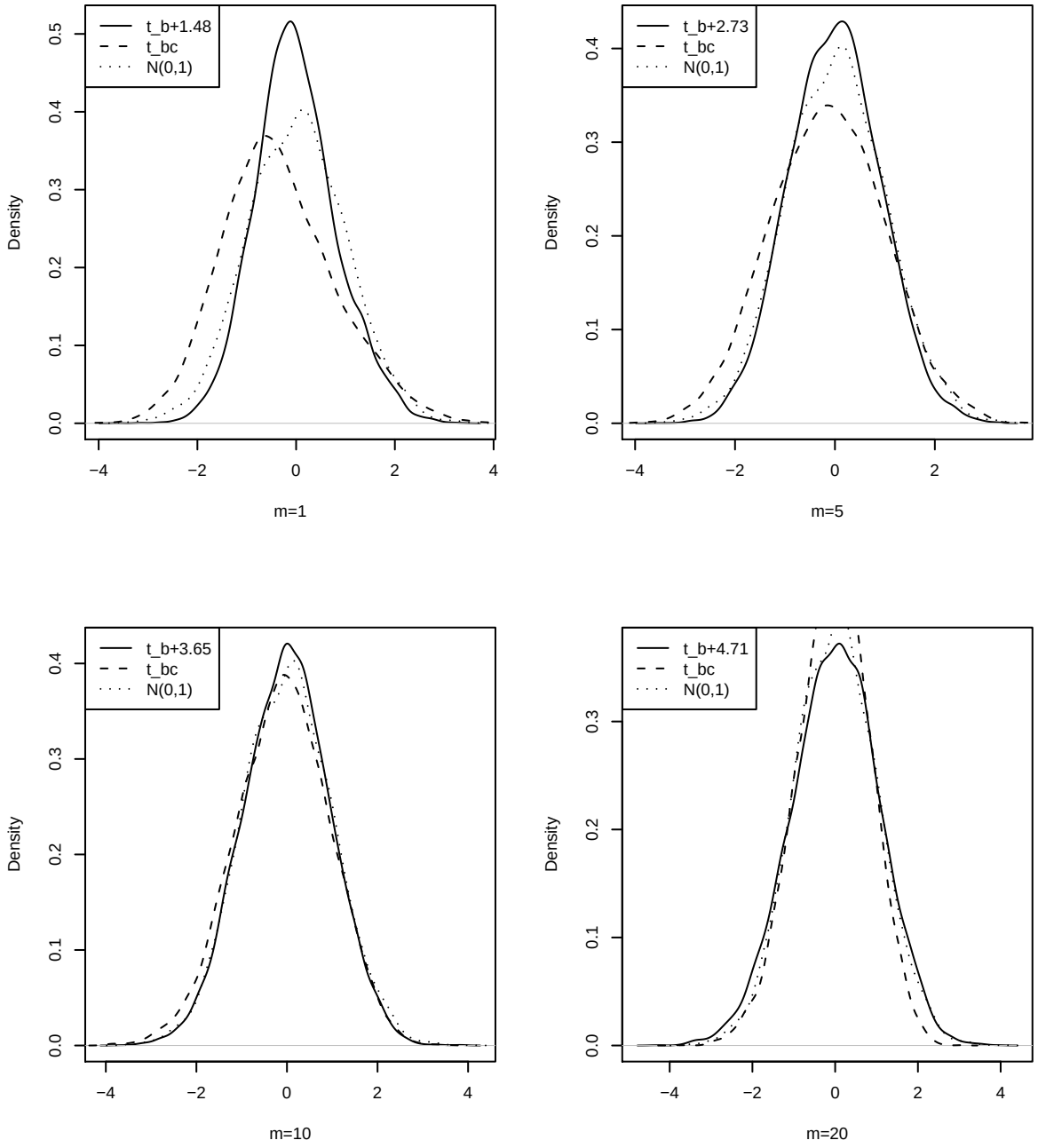


Figure 7.5: Density of t-statistic in M2-I (T=100)

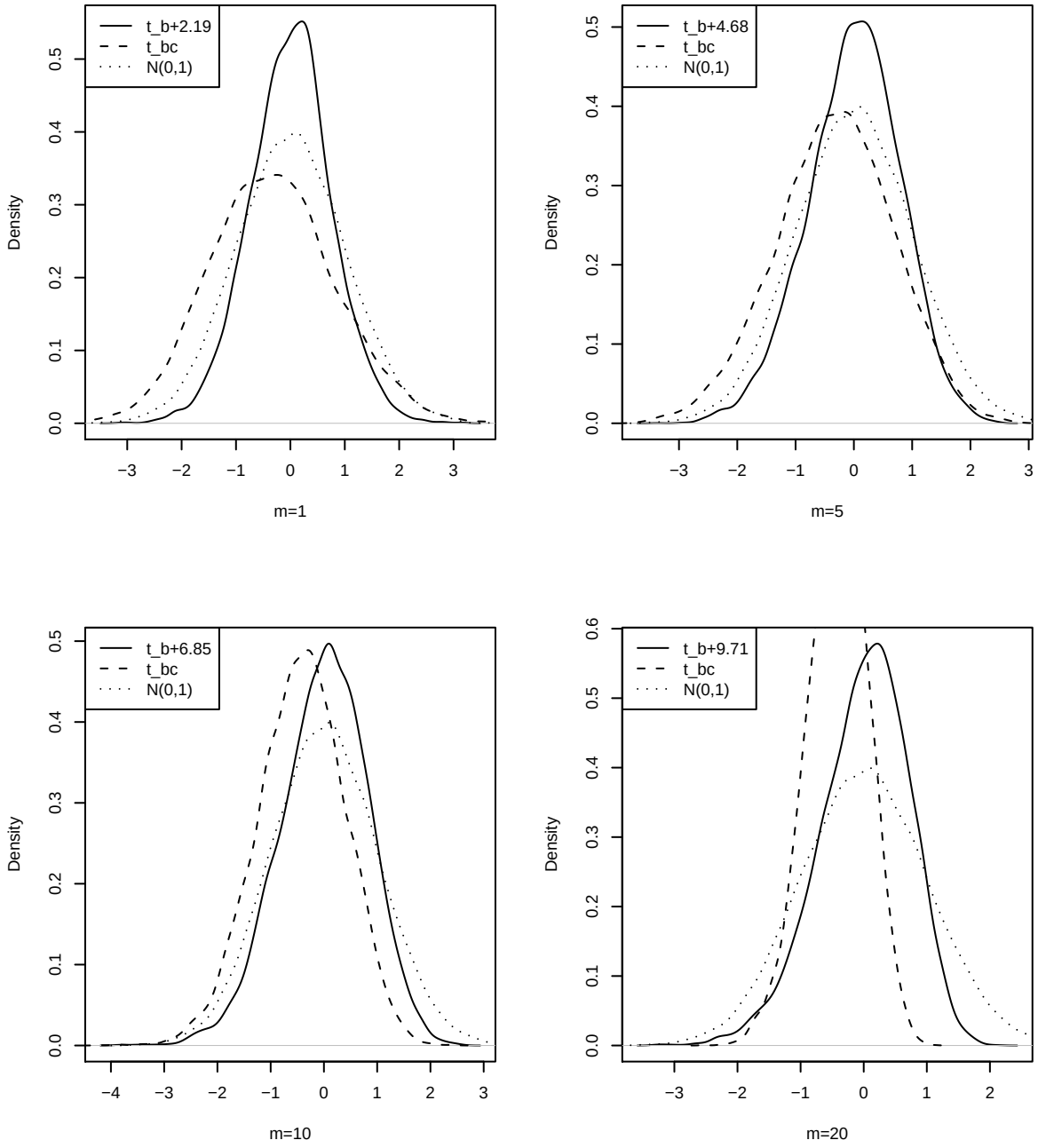


Figure 7.6: Density of t-statistic in M3-I (T=100)

7.3 Empirical Power of Basic Models

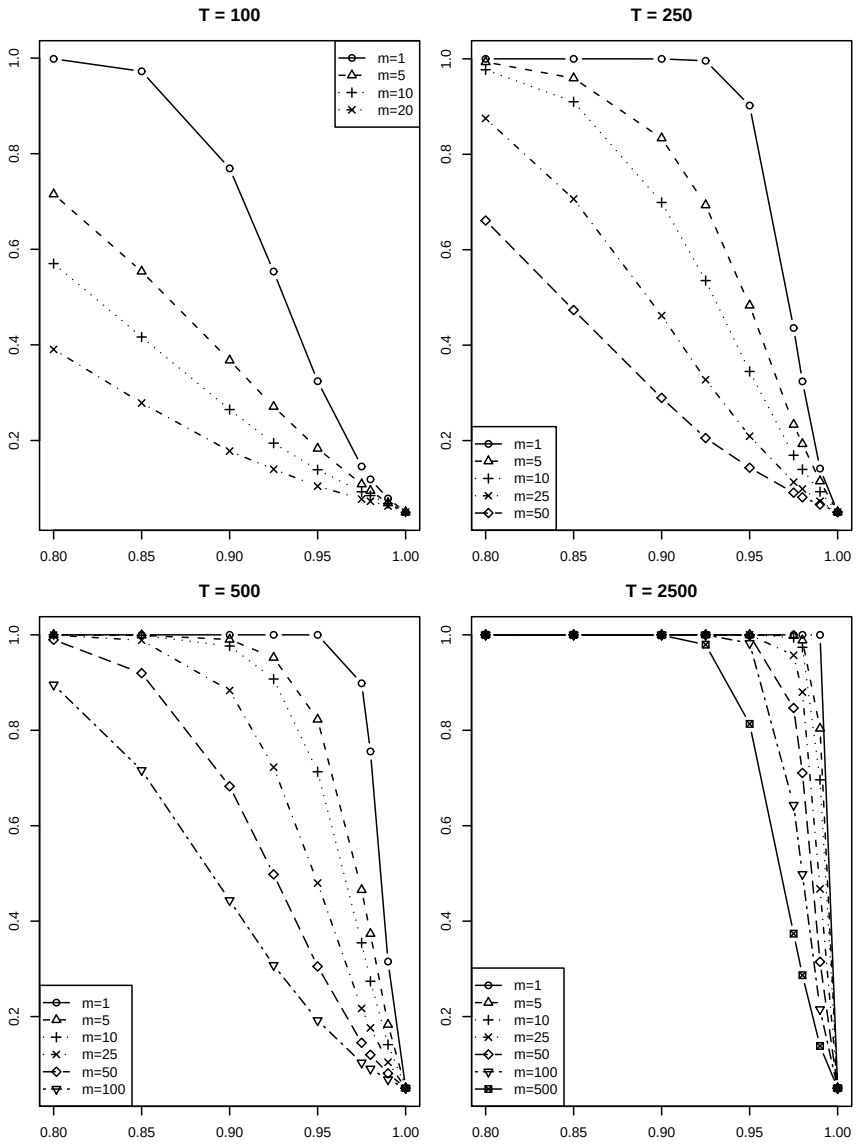


Figure 7.7: Empirical Power of t-statistic in M1

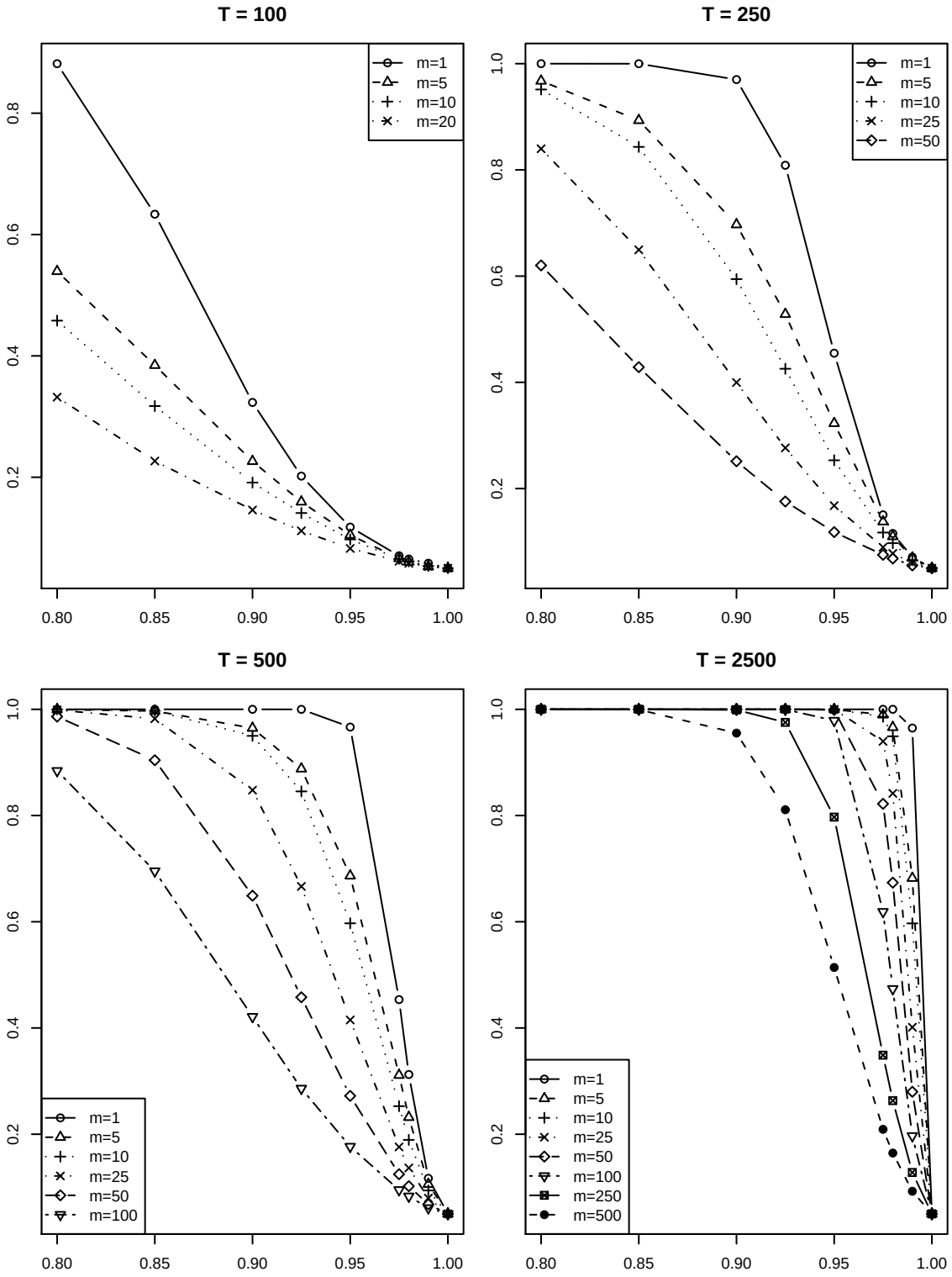


Figure 7.8: Empirical Power of $t-$statistic in M2-G

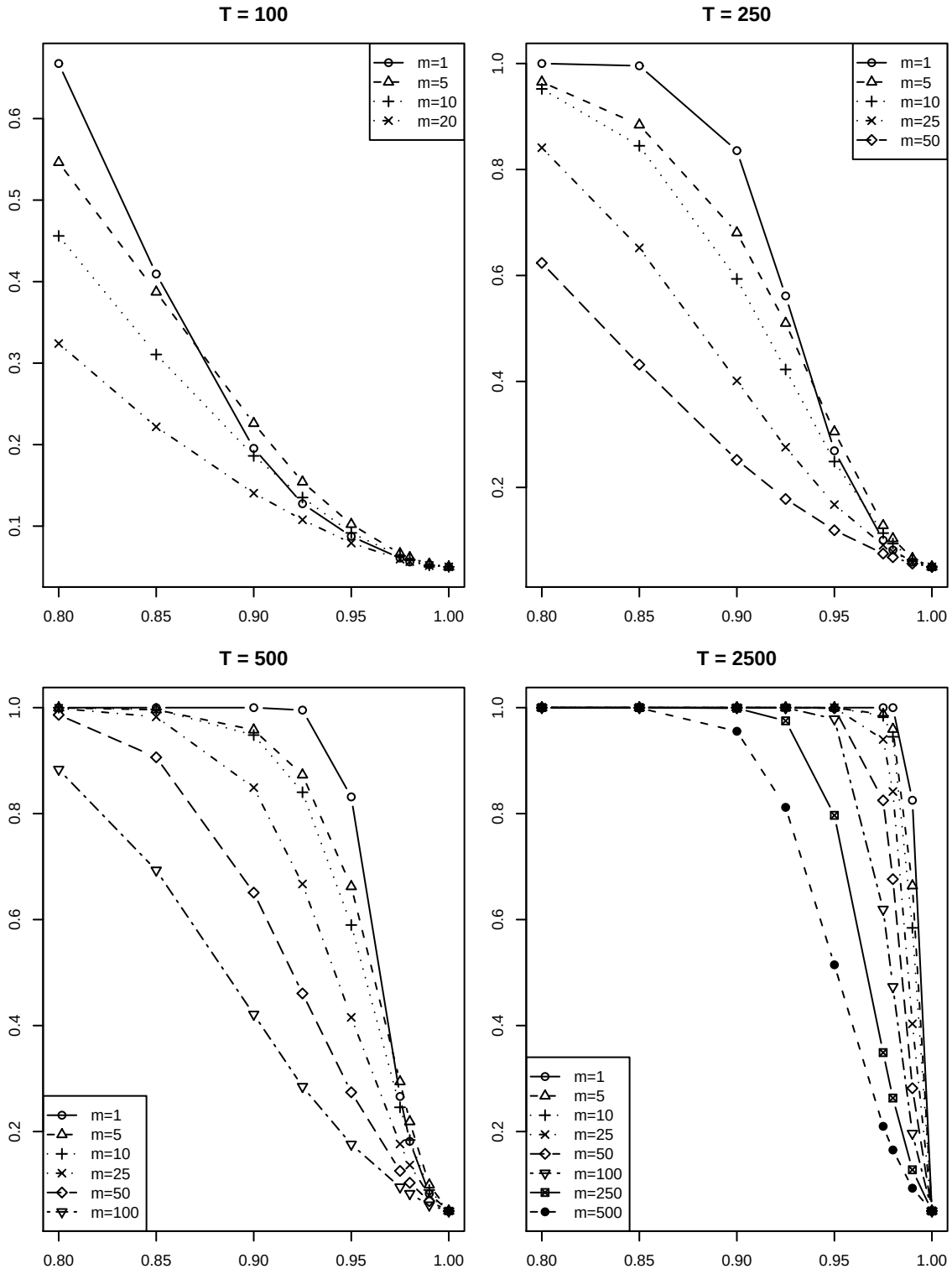


Figure 7.9: Empirical Power of $t-$statistic in M3-G

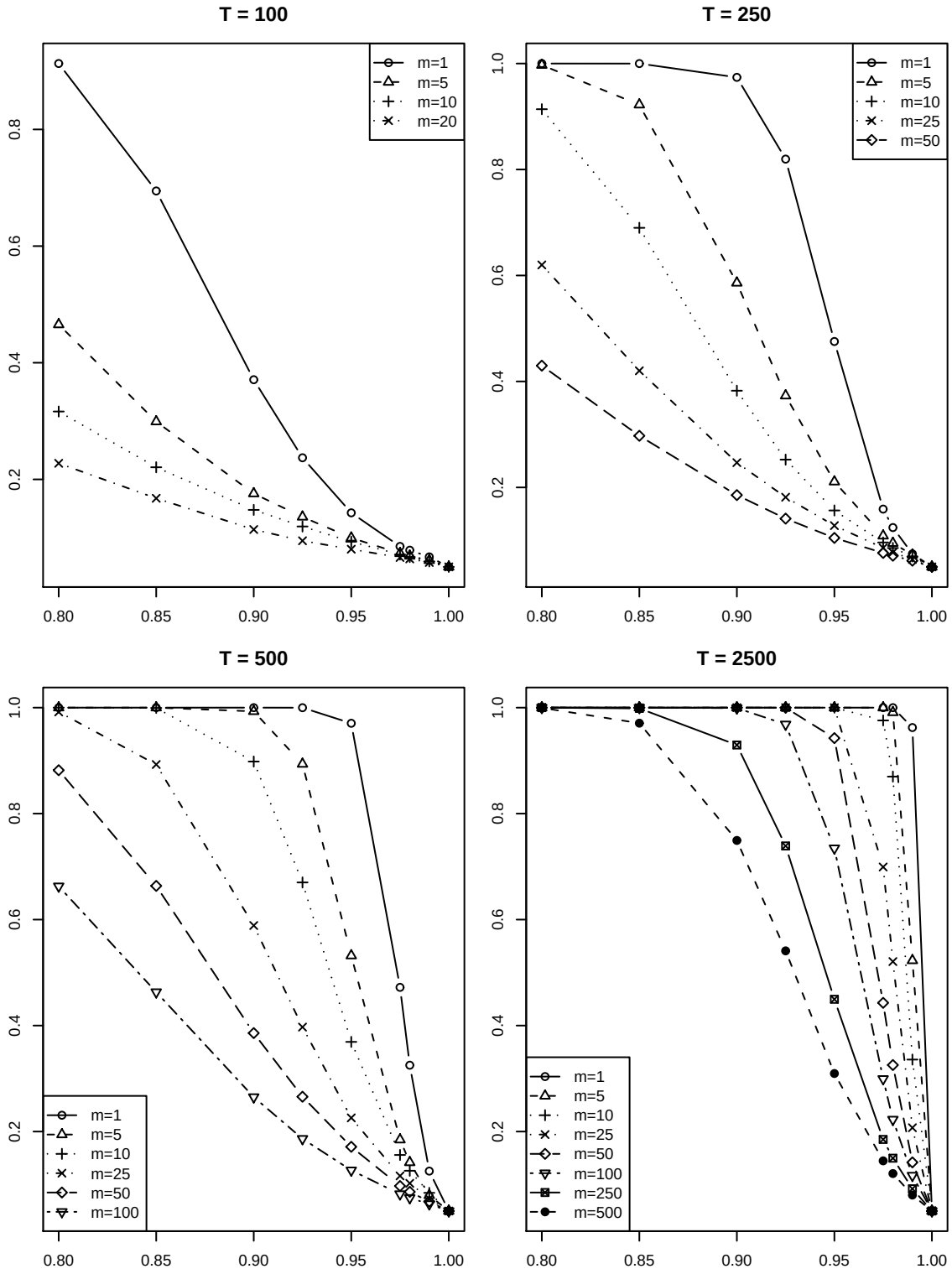


Figure 7.10: Empirical Power of t_{biased} in M2-I

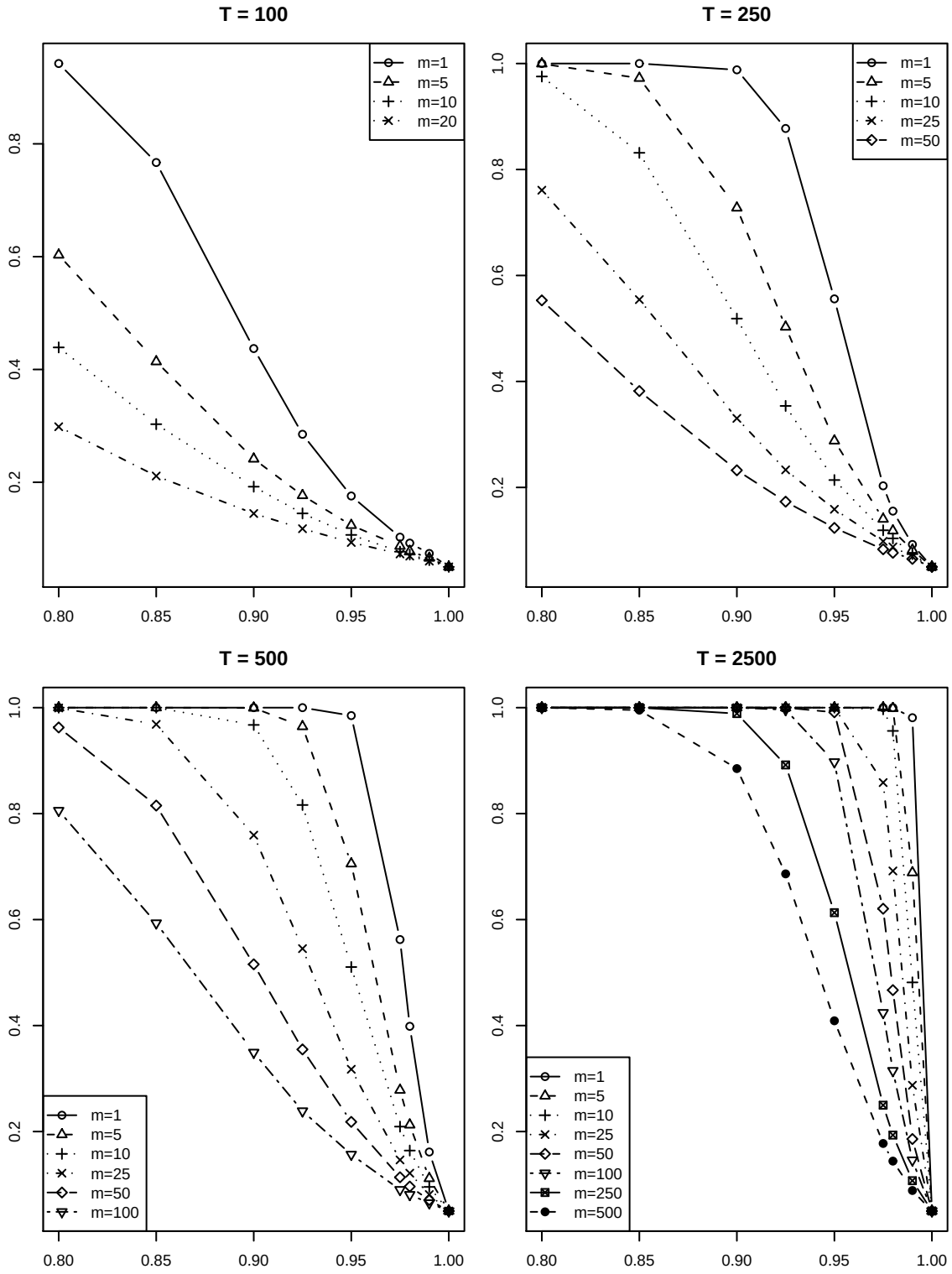


Figure 7.11: Empirical Power of t_{bc} in M2-I

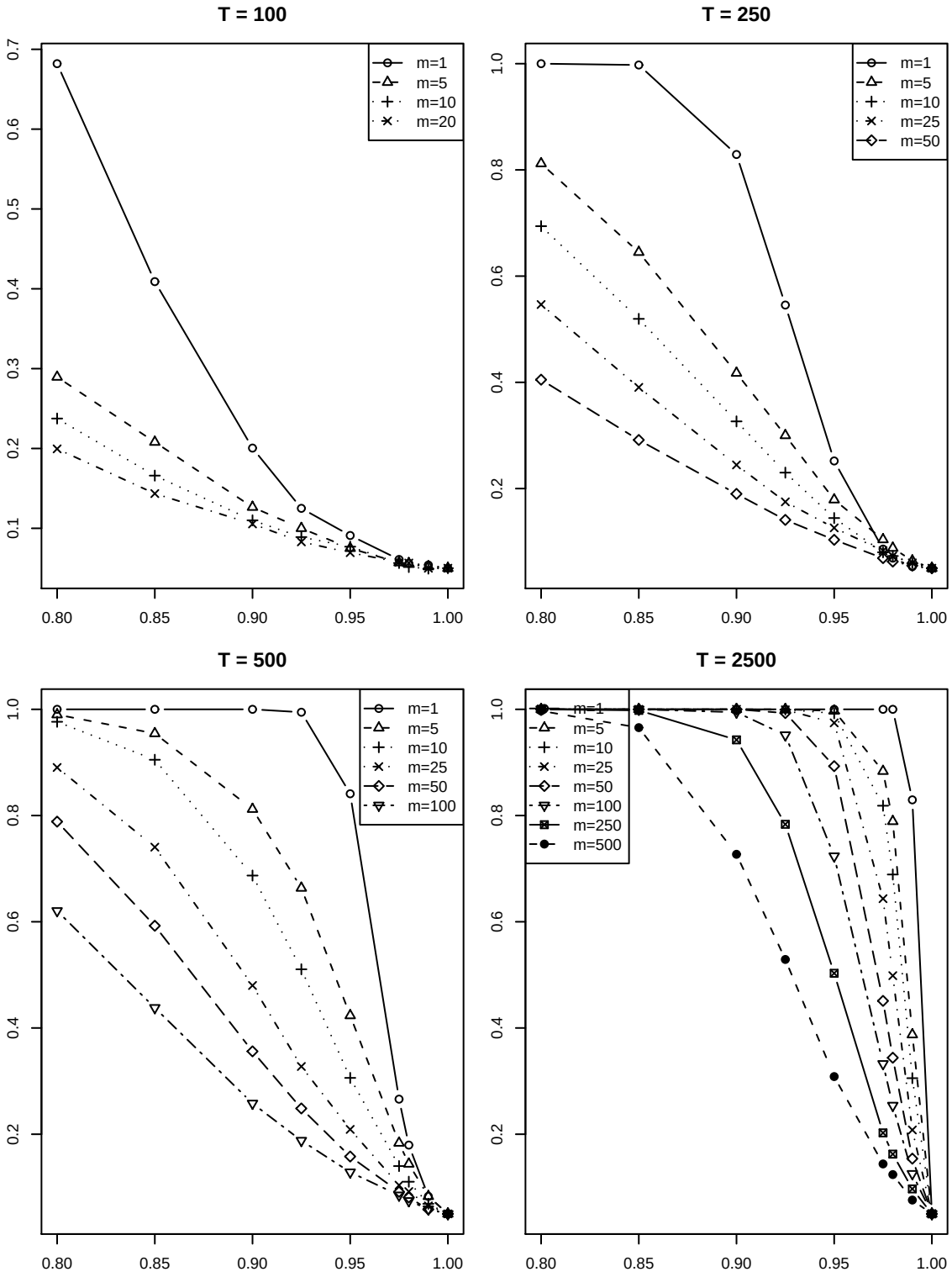


Figure 7.12: Empirical Power of t_{biased} in M3-I

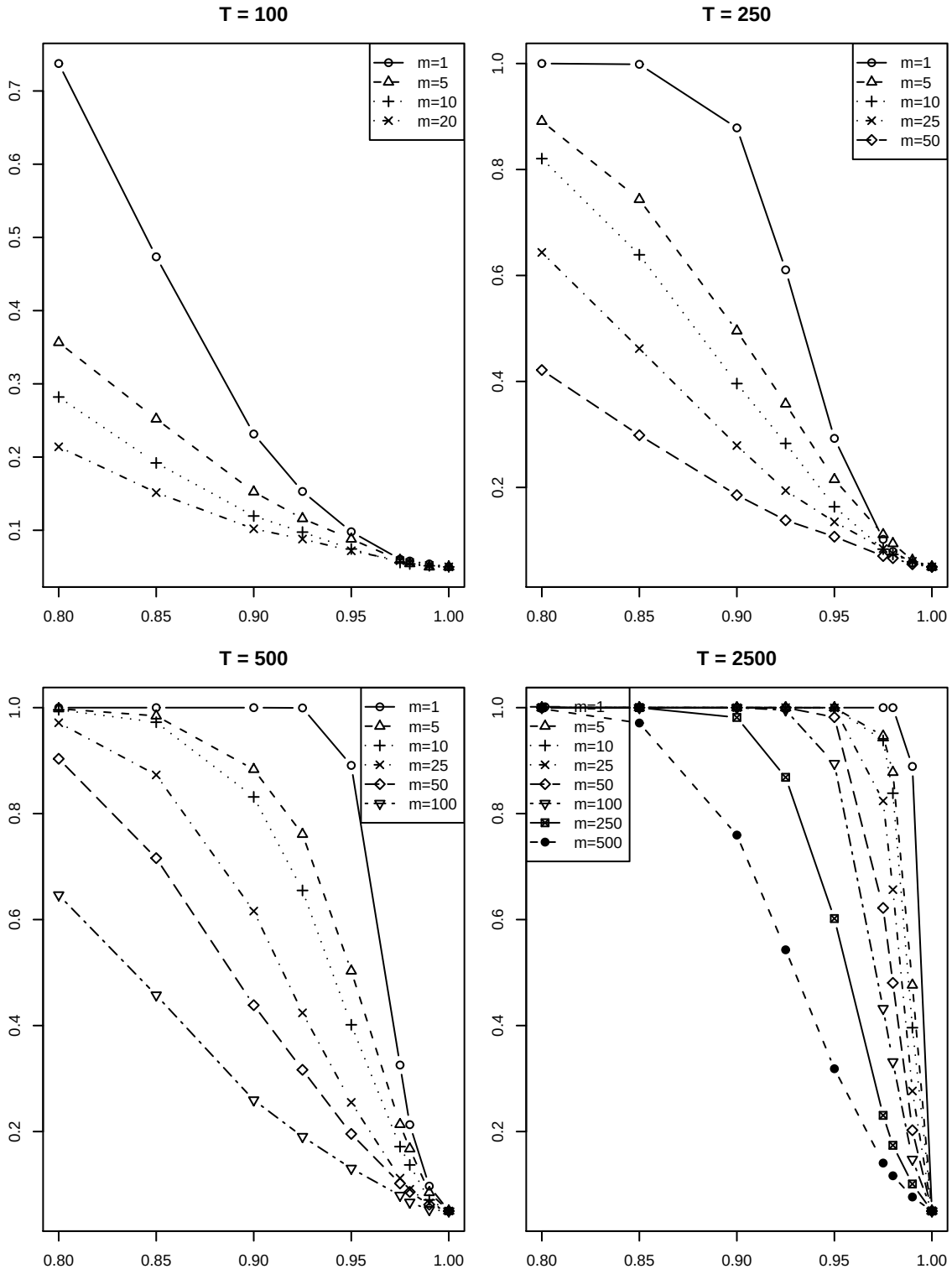


Figure 7.13: Empirical Power of t_{bc} in M3-I

7.4 Empirical Power in the Presence of Single Break

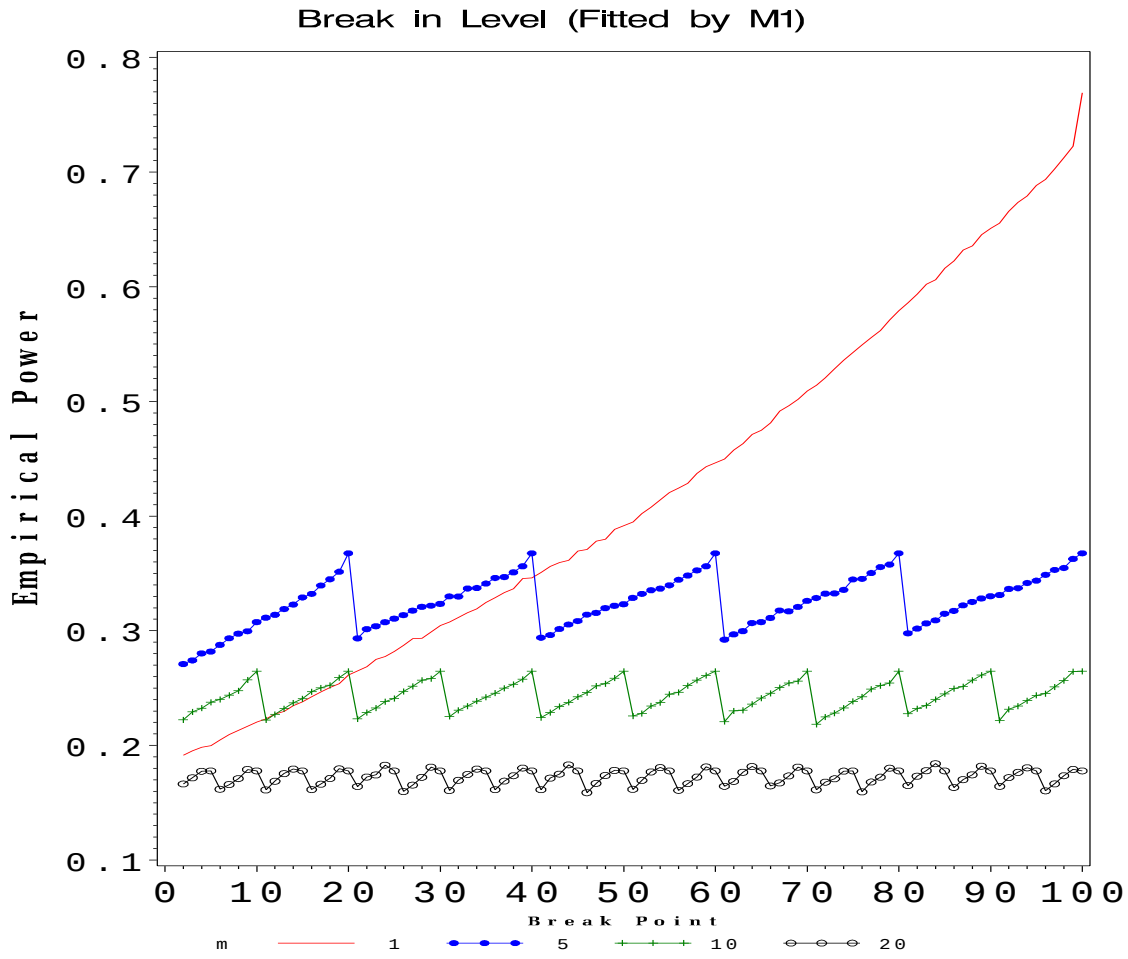


Figure 7.14: Empirical Power of $t(0.05)[T = 100, \rho = 0.9, \alpha_1 = 2.5]$

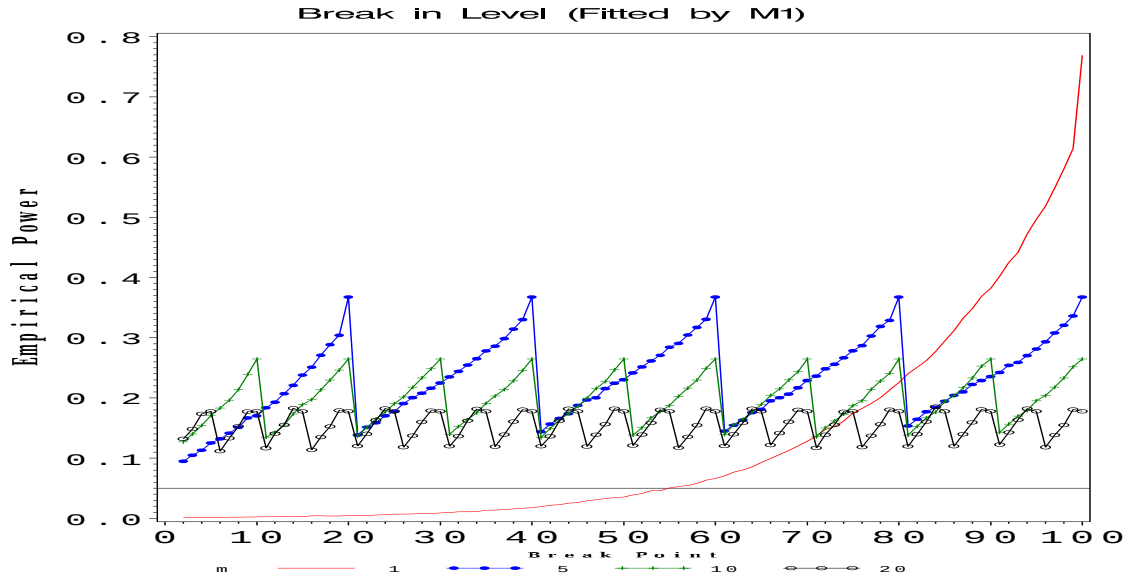


Figure 7.15: Empirical Power of $t(0.05)[T = 100, \rho = 0.9, \alpha_1 = 5]$

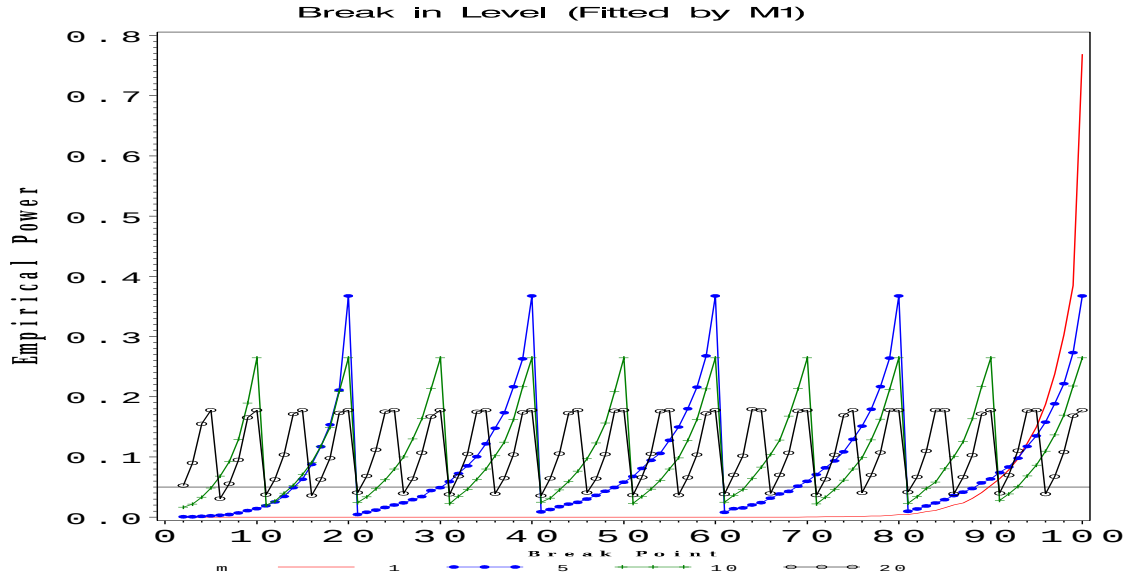


Figure 7.16: Empirical Power of $t(0.05)[T = 100, \rho = 0.9, \alpha_1 = 10]$

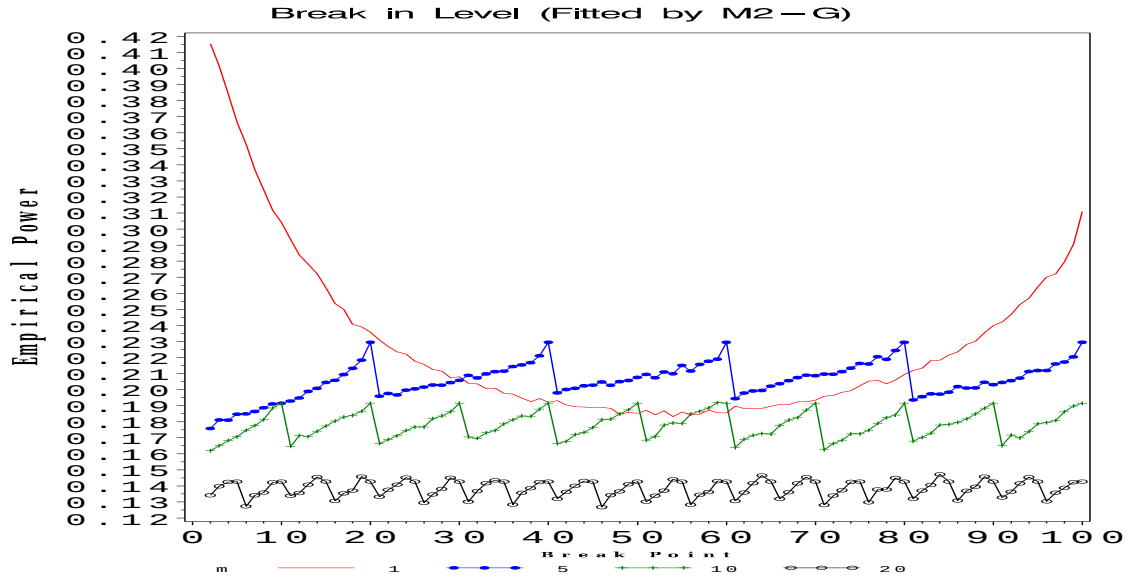


Figure 7.17: Empirical Power of $t_\mu(0.05)[T = 100, \rho = 0.9, \alpha_1 = 2.5]$

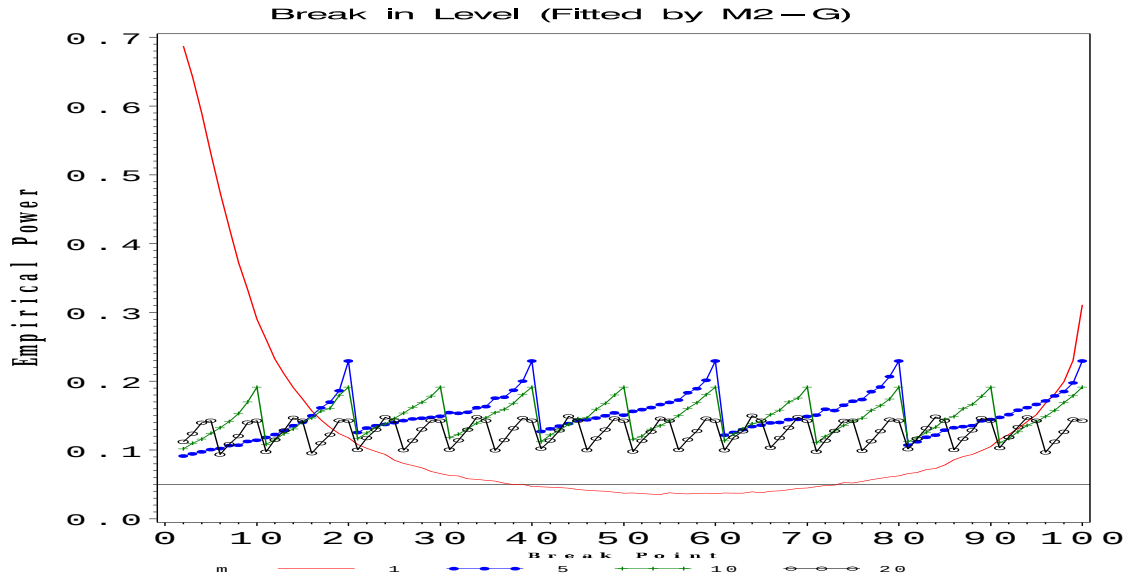


Figure 7.18: Empirical Power of $t_\mu(0.05)[T = 100, \rho = 0.9, \alpha_1 = 5]$

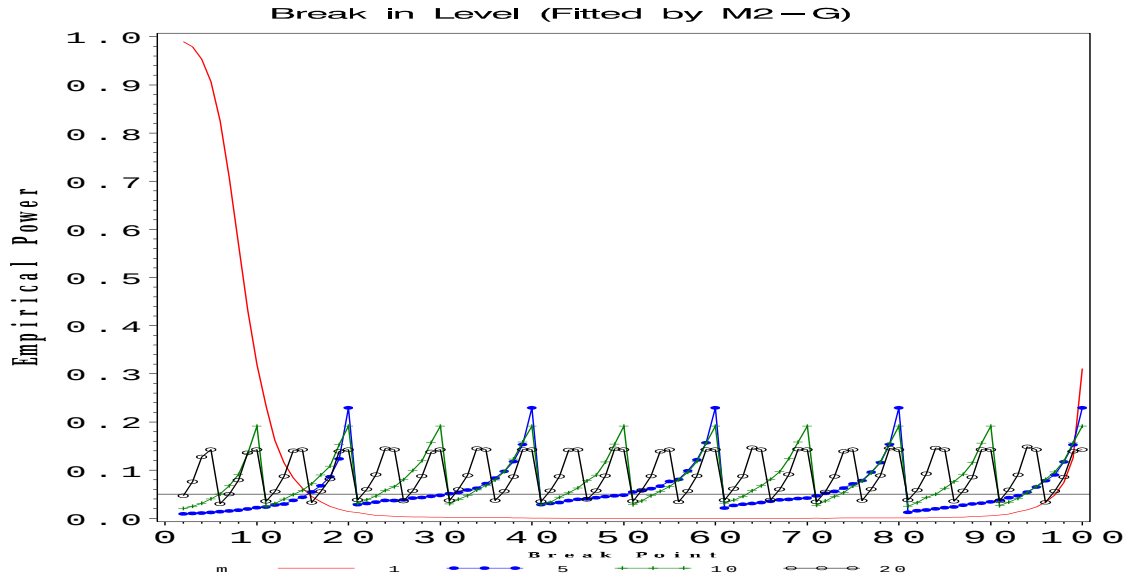


Figure 7.19: Empirical Power of $t_\mu(0.05)[T = 100, \rho = 0.9, \alpha_1 = 10]$

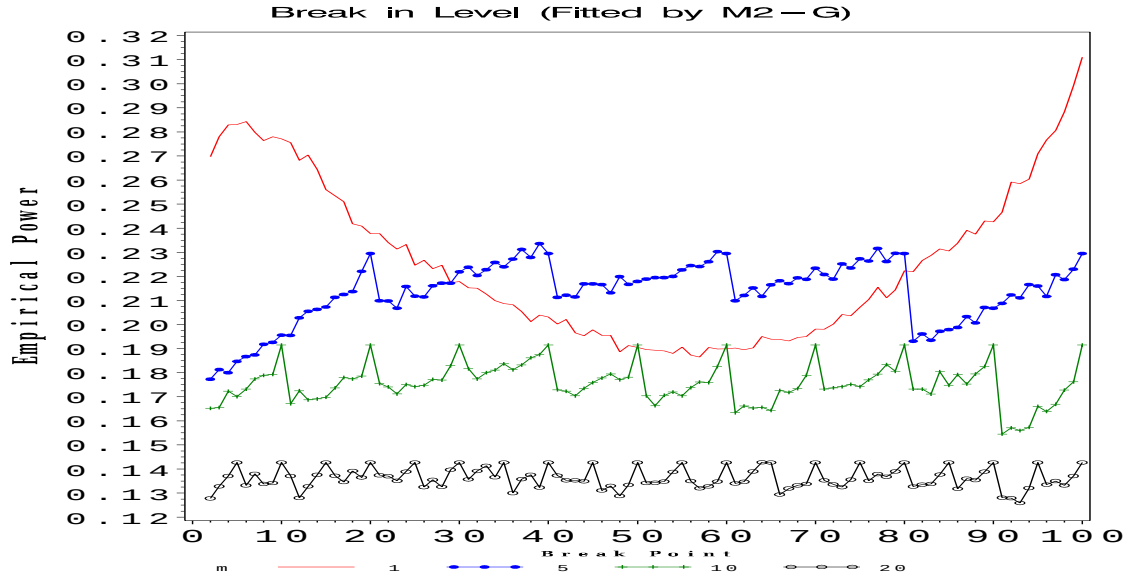


Figure 7.20: Size-Adjusted Empirical Power of $t_\mu(0.05)[T = 100, \rho = 0.9, \alpha_1 = 2.5]$

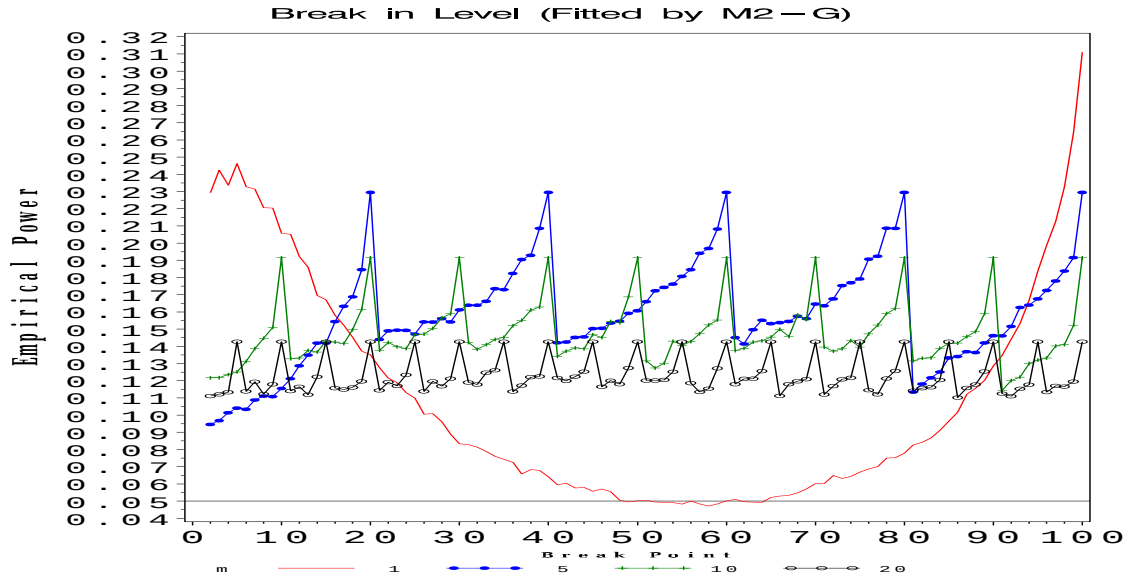


Figure 7.21: Size-Adjusted Empirical Power of $t_\mu(0.05)[T = 100, \rho = 0.9, \alpha_1 = 5]$

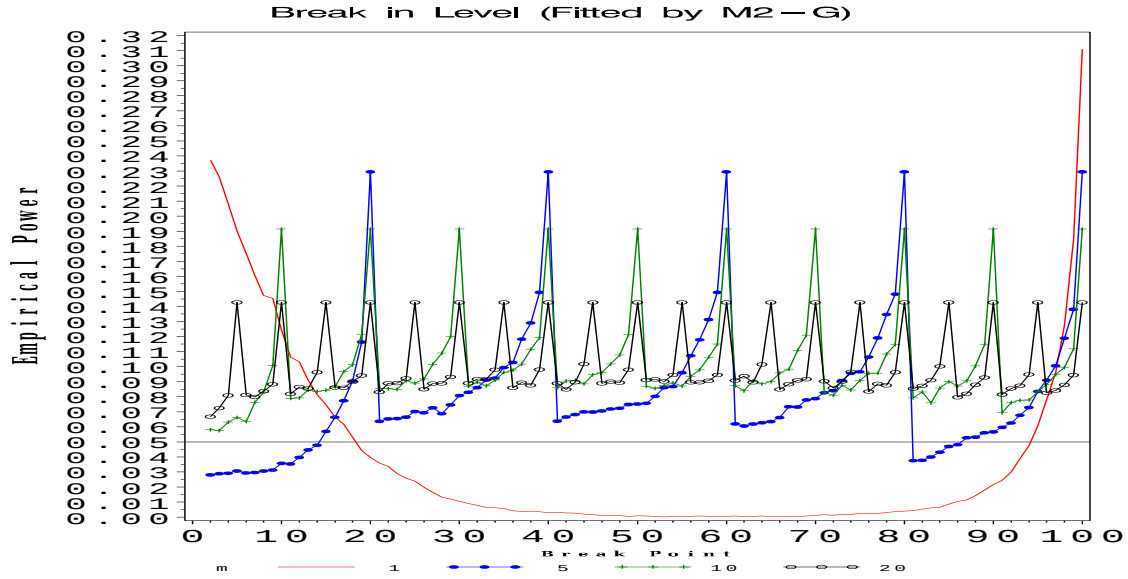


Figure 7.22: Size-Adjusted Empirical Power of $t_\mu(0.05)[T = 100, \rho = 0.9, \alpha_1 = 10]$

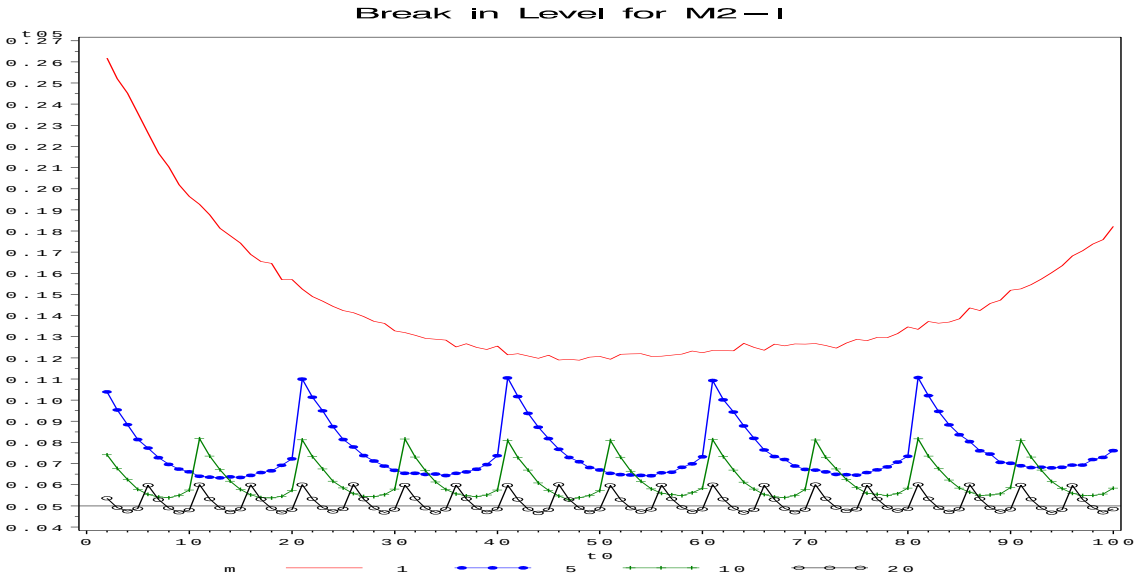


Figure 7.23: Empirical Power of $t_{\mu_i}(0.05)[T = 100, \rho = 0.9, \alpha_1 = 2.5]$

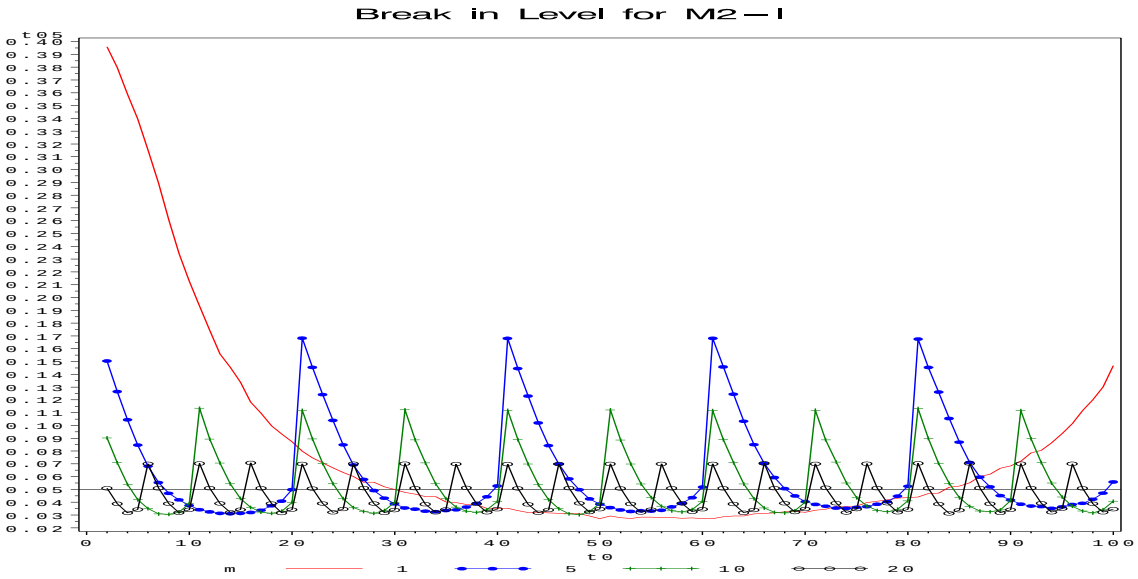


Figure 7.24: Empirical Power of $t_{\mu_i}(0.05)[T = 100, \rho = 0.9, \alpha_1 = 5]$

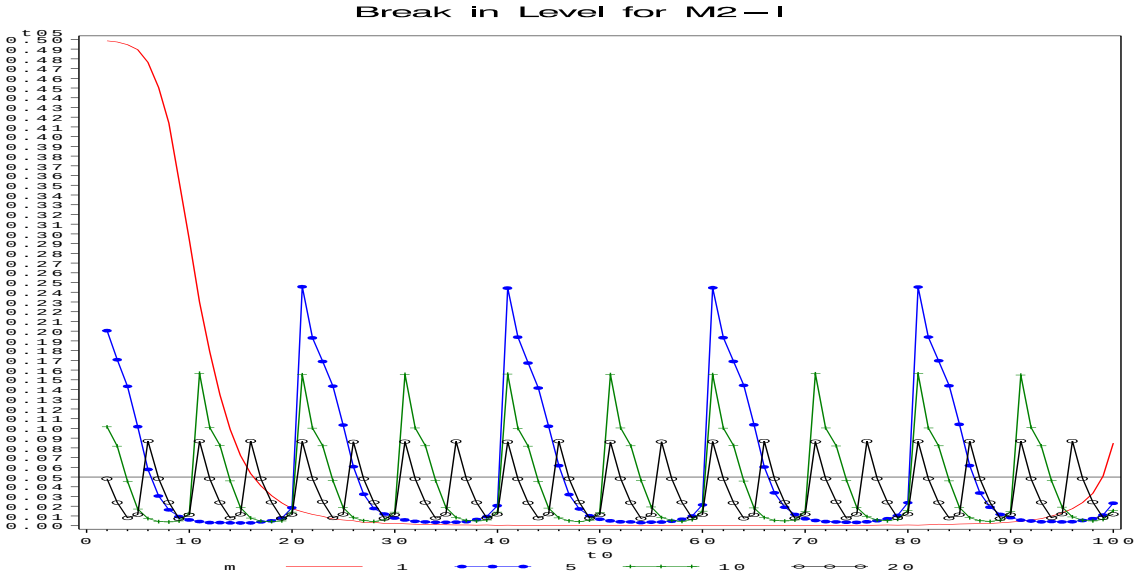


Figure 7.25: Empirical Power of $t_{\mu_i}(0.05)[T = 100, \rho = 0.9, \alpha_1 = 10]$

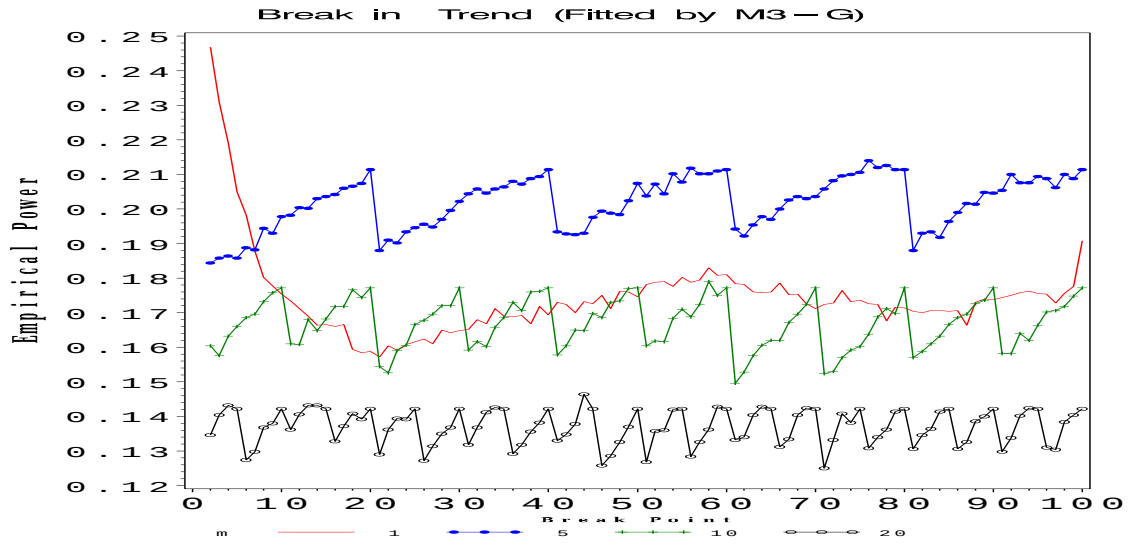


Figure 7.26: Empirical Power of $t_\tau(0.05)[T = 100, \rho = 0.9, \alpha_1 = 2.5]$

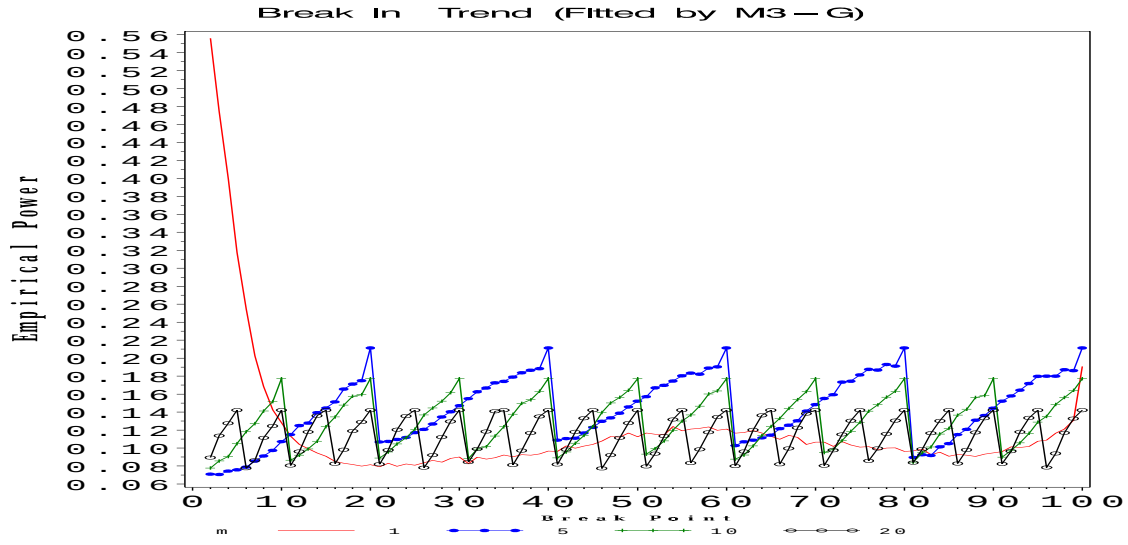


Figure 7.27: Empirical Power of $t_\tau(0.05)[T = 100, \rho = 0.9, \alpha_1 = 5]$

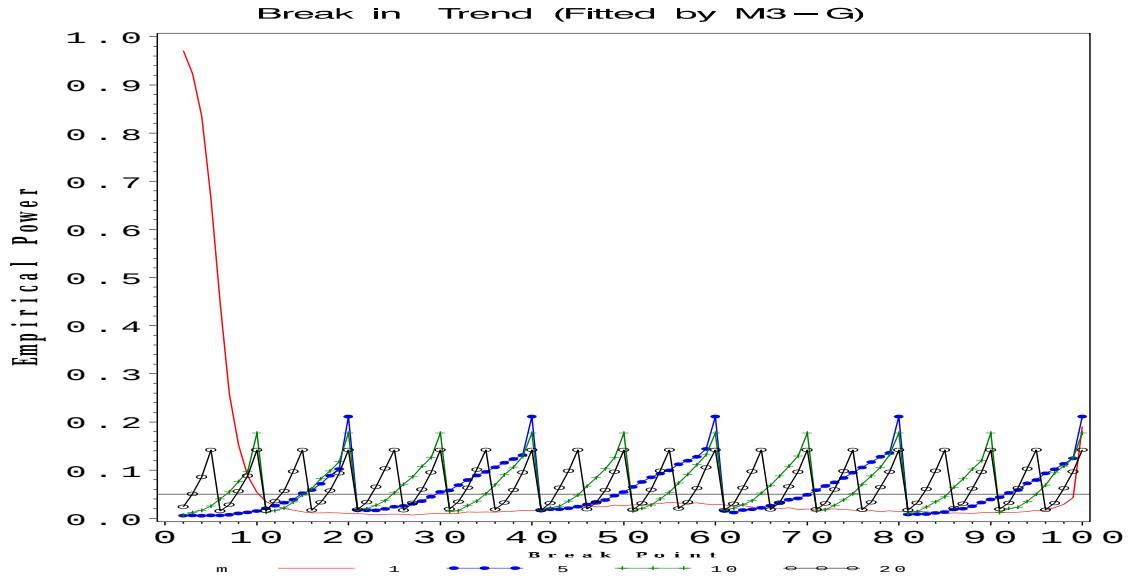


Figure 7.28: Empirical Power of $t_\tau(0.05)[T = 100, \rho = 0.9, \alpha_1 = 10]$

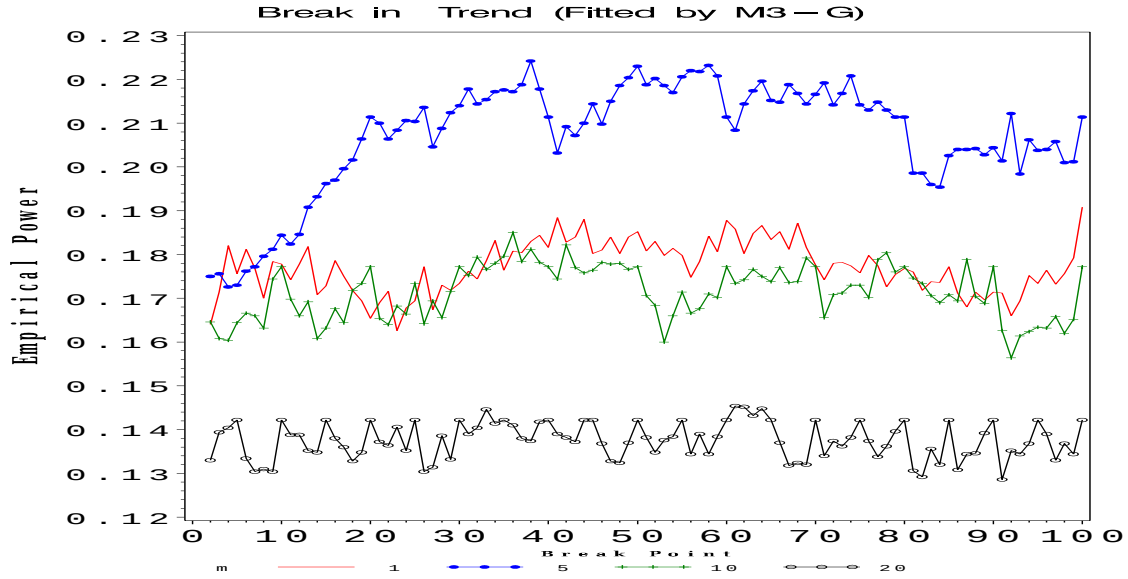


Figure 7.29: Size-Adjusted Empirical Power of $t_\tau(0.05)[T = 100, \rho = 0.9, \alpha_1 = 2.5]$

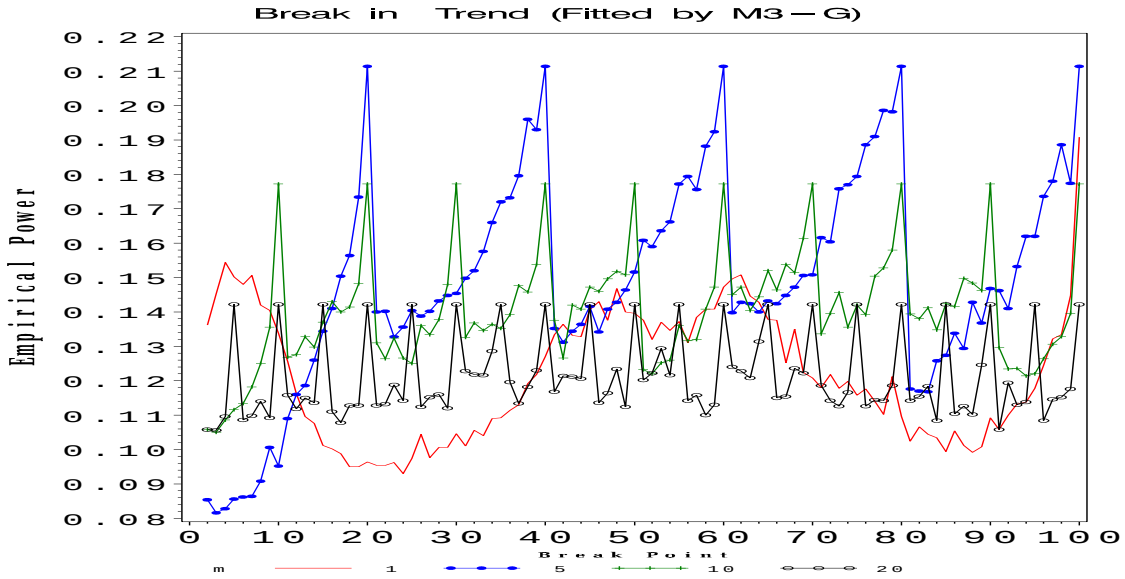


Figure 7.30: Size-Adjusted Empirical Power of $t_\tau(0.05)[T = 100, \rho = 0.9, \alpha_1 = 5]$

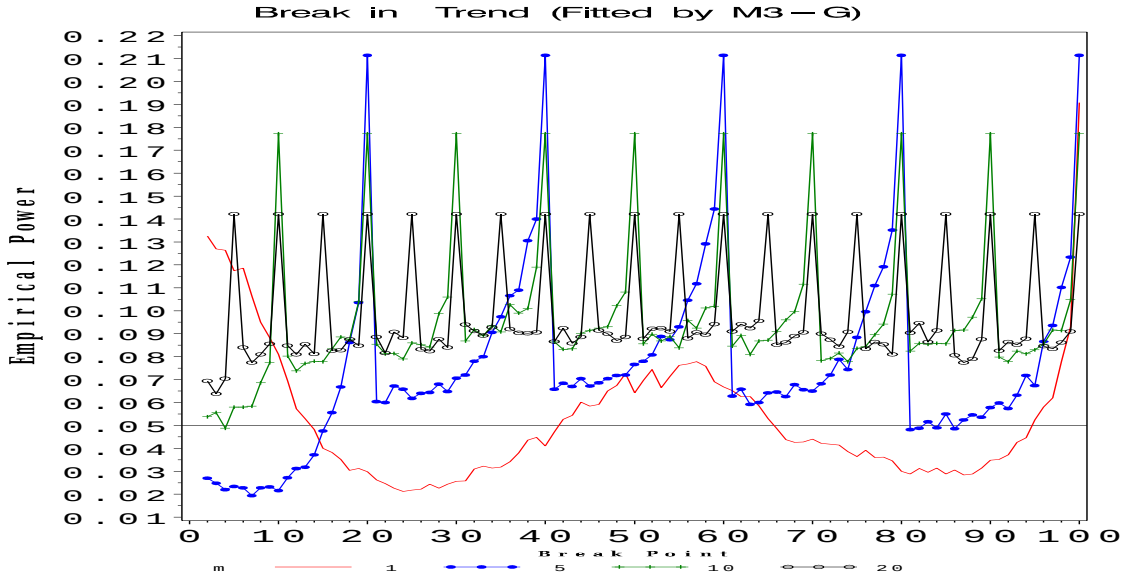


Figure 7.31: Size-Adjusted Empirical Power of $t_\tau(0.05)[T = 100, \rho = 0.9, \alpha_1 = 10]$

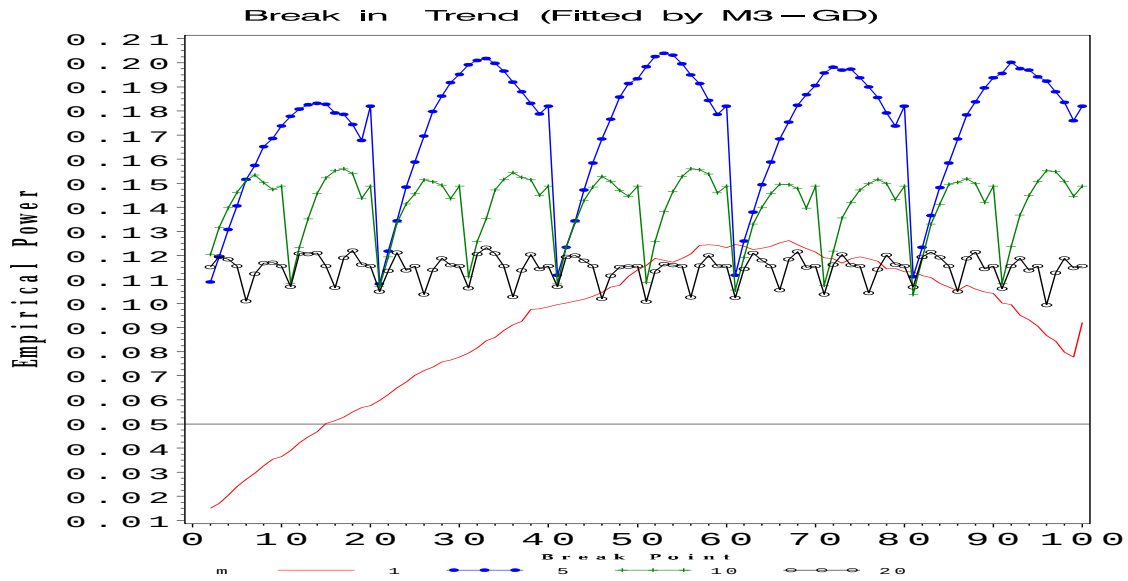


Figure 7.32: Empirical Power of $t_\tau(0.05)[T = 100, \rho = 0.90, \alpha_1 = 2.5]$

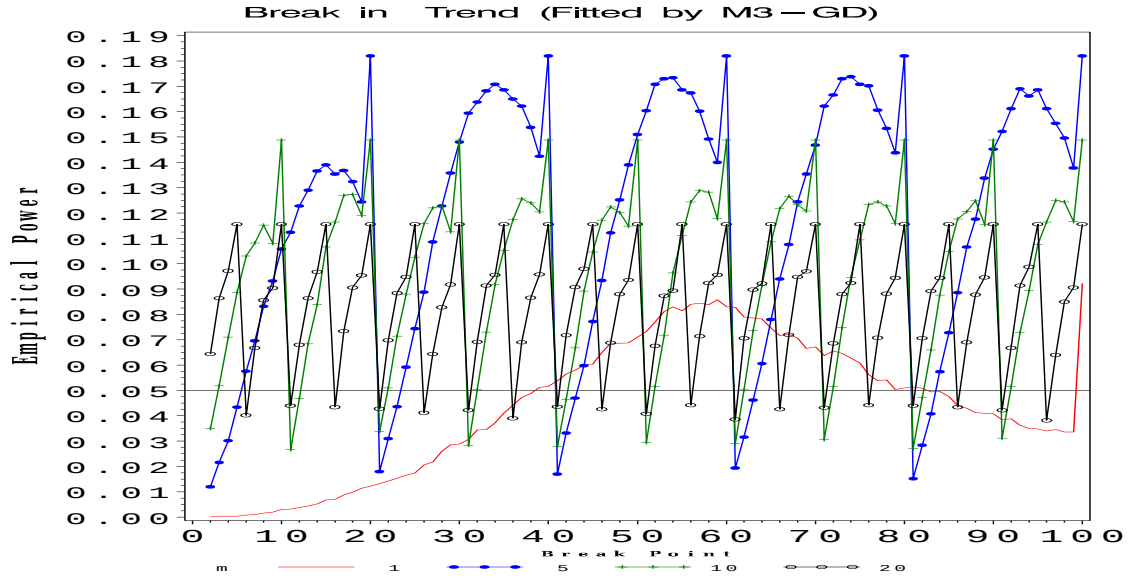


Figure 7.33: Empirical Power of $t_\tau(0.05)[T = 100, \rho = 0.90, \alpha_1 = 5]$

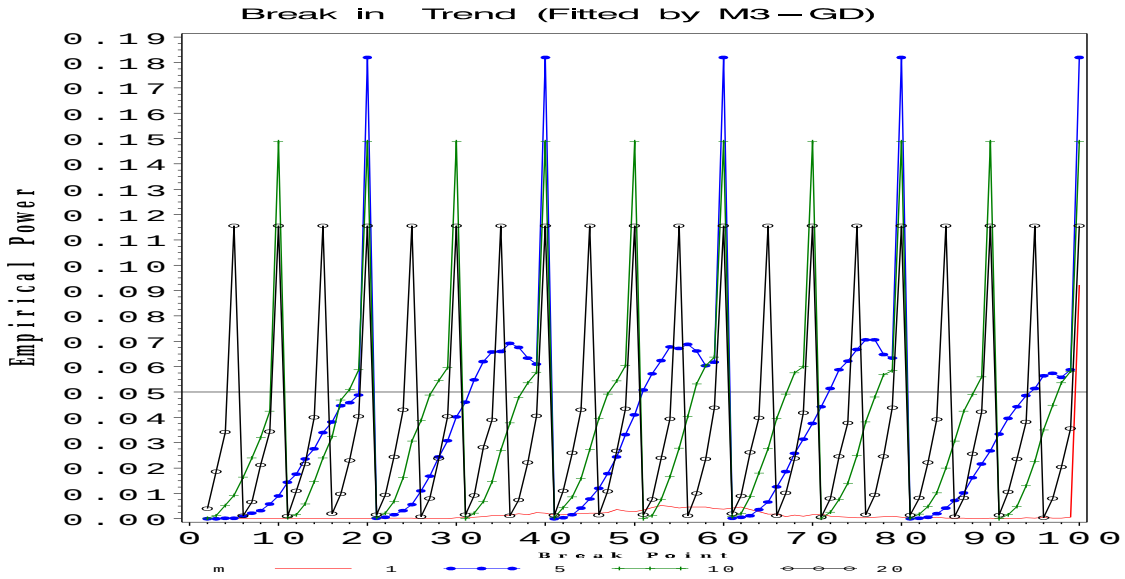


Figure 7.34: Empirical Power of $t_\tau(0.05)[T = 100, \rho = 0.90, \alpha_1 = 10]$

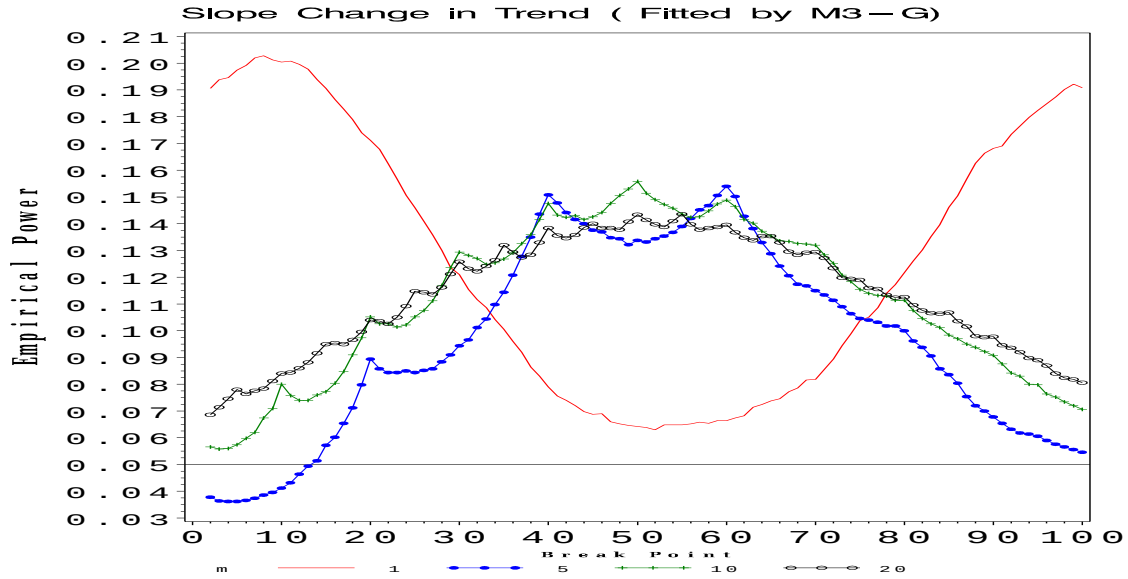


Figure 7.35: Empirical Power of $t_\tau(0.05)[T = 100, \rho = 0.9, \beta_1 = 1, \beta_2 = .2]$

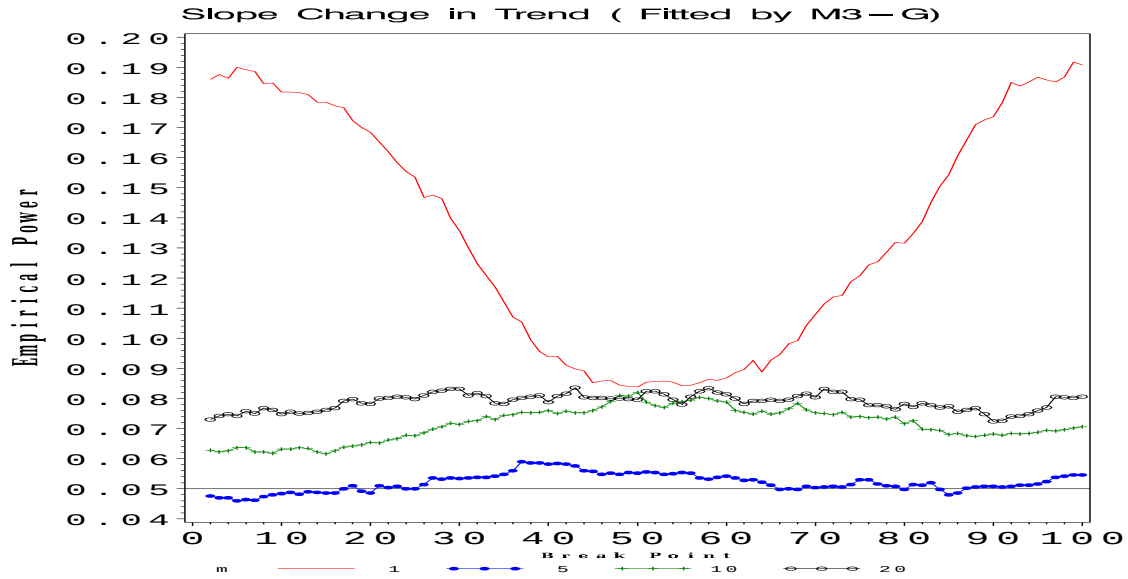


Figure 7.36: Size Adjusted Power of $t_\tau(0.05)[T = 100, \rho = 0.9, \beta_1 = 1, \beta_2 = .2]$

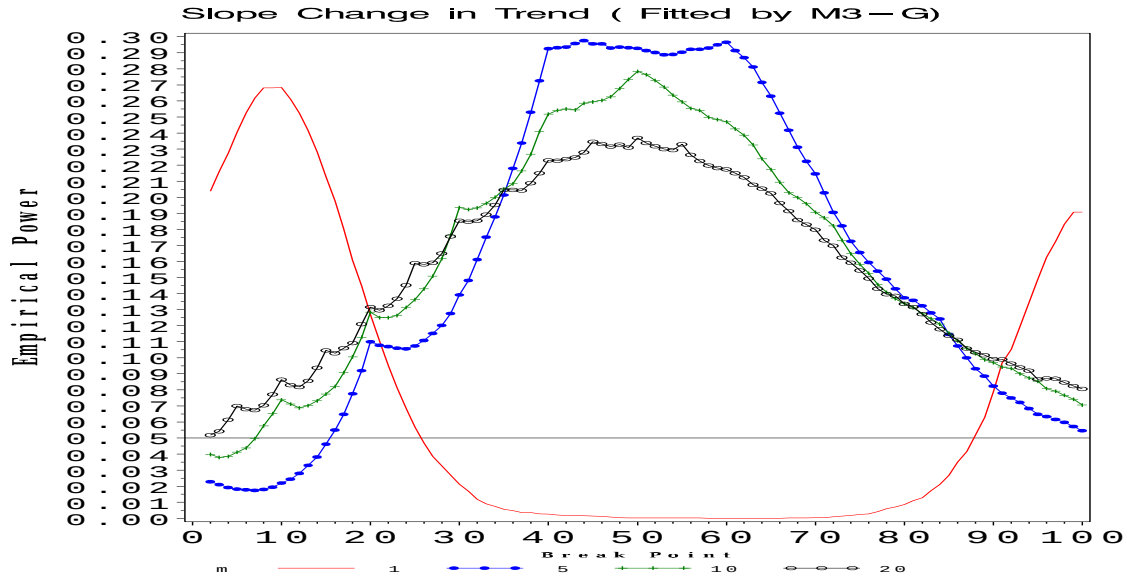


Figure 7.37: Empirical Power of $t_\tau(0.05)[T = 100, \rho = 0.9, \beta_1 = 1, \beta_2 = .5]$

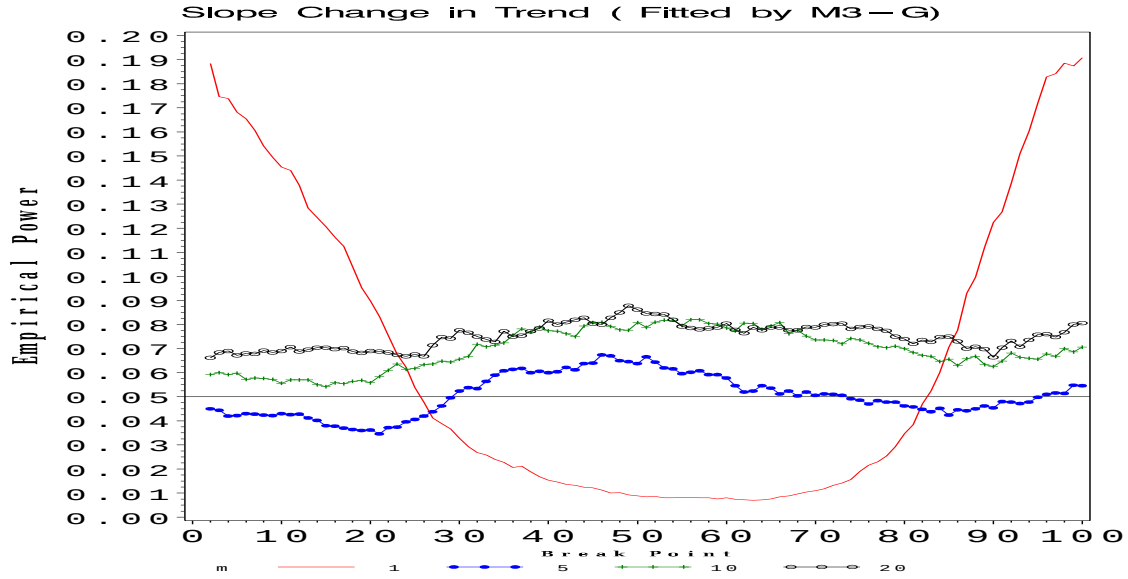


Figure 7.38: Size Adjusted Power of $t_\tau(0.05)[T = 100, \rho = 0.9, \beta_1 = 1, \beta_2 = .5]$

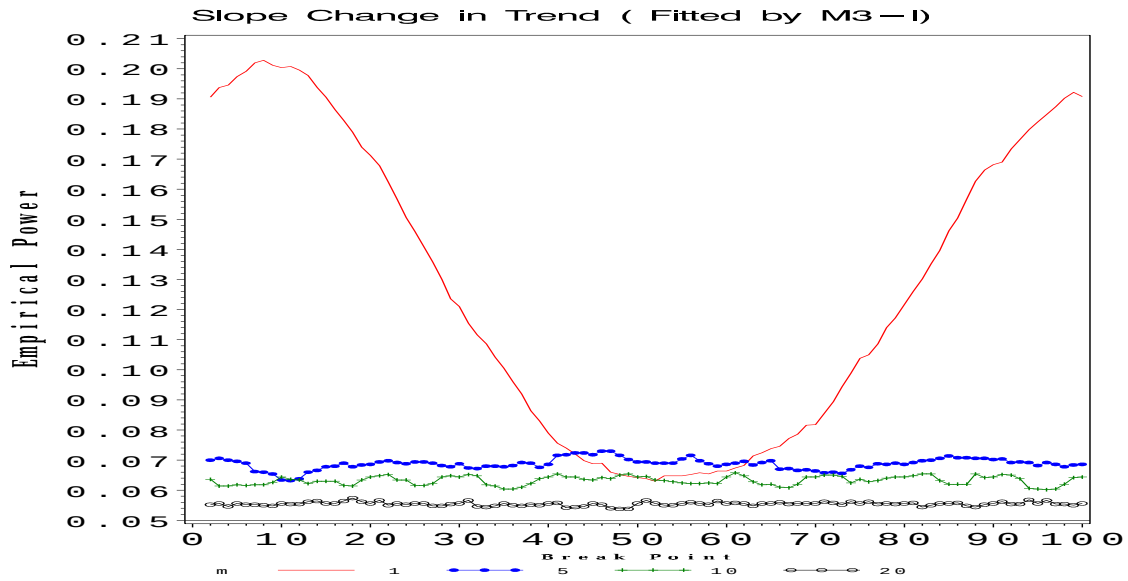


Figure 7.39: Empirical Power of $t_{\tau_i}(0.05)[T = 100, \rho = 0.9, \beta_1 = 1, \beta_2 = .2]$

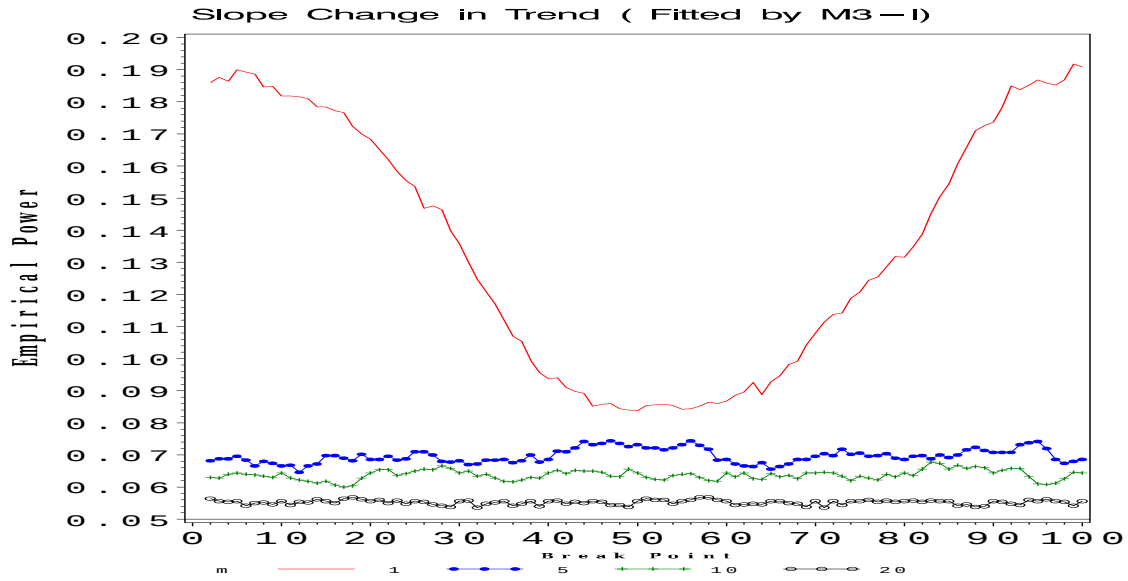


Figure 7.40: Size Adjusted Power of $t_{\tau_i}(0.05)[T = 100, \rho = 0.9, \beta_1 = 1, \beta_2 = .2]$

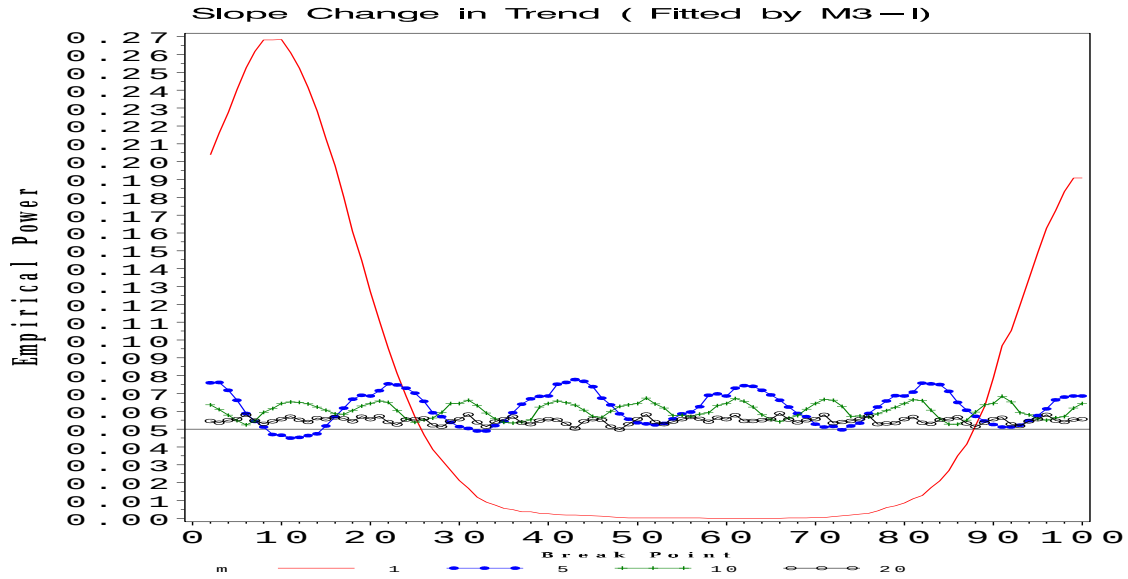


Figure 7.41: Empirical Power of $t_{\tau_i}(0.05)[T = 100, \rho = 0.9, \beta_1 = 1, \beta_2 = .5]$

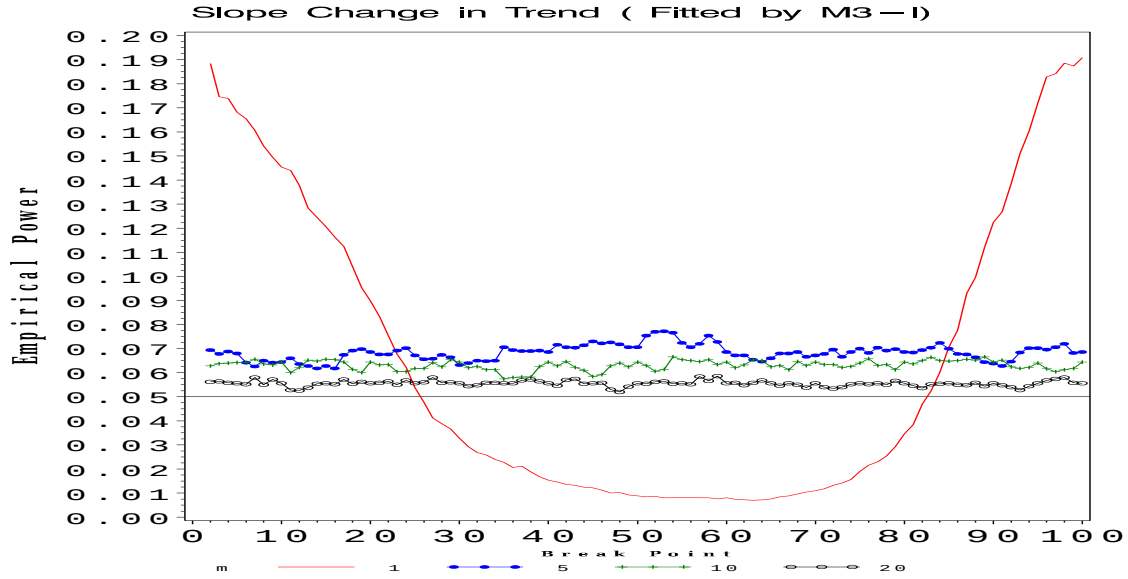


Figure 7.42: Size Adjusted Power of $t_{\tau_i}(0.05)[T = 100, \rho = 0.9, \beta_1 = 1, \beta_2 = .5]$

7.5 Empirical Power in the Presence of Multiple Breaks

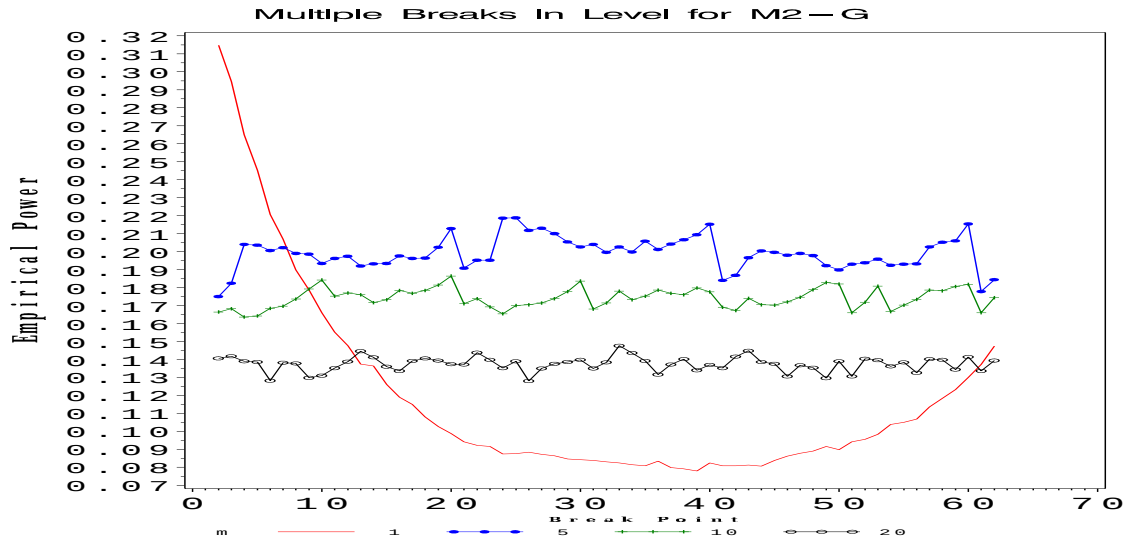


Figure 7.43: Empirical Power of $t_\mu(0.05)[T = 100, \rho = 0.9, \alpha_1 = 2.5, \alpha_2 = 2.5]$

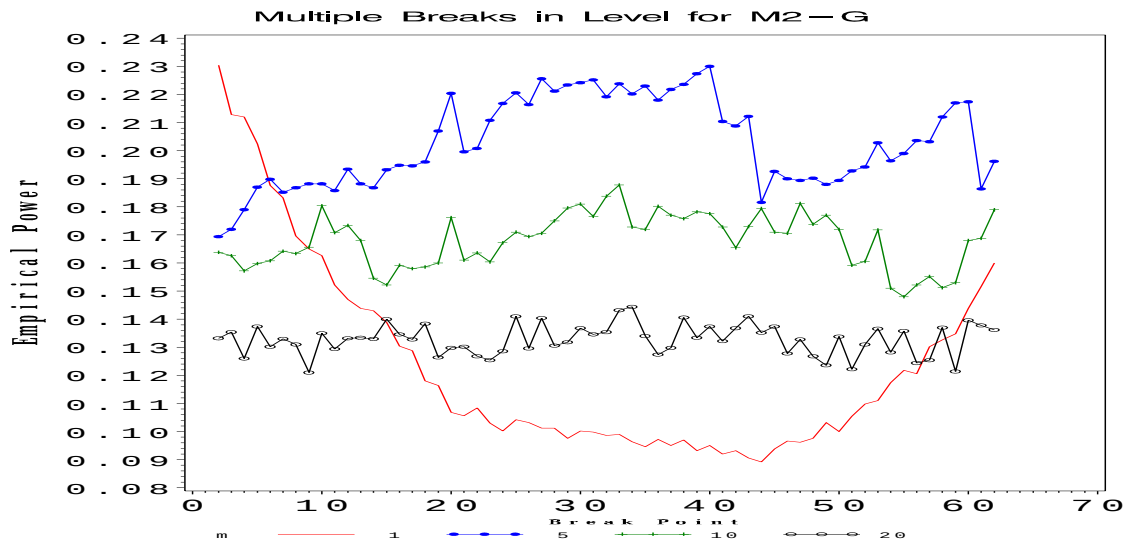


Figure 7.44: Size Adjusted Power of $t_\mu(0.05)[T = 100, \rho = 0.9, \alpha_1 = 2.5, \alpha_2 = 2.5]$

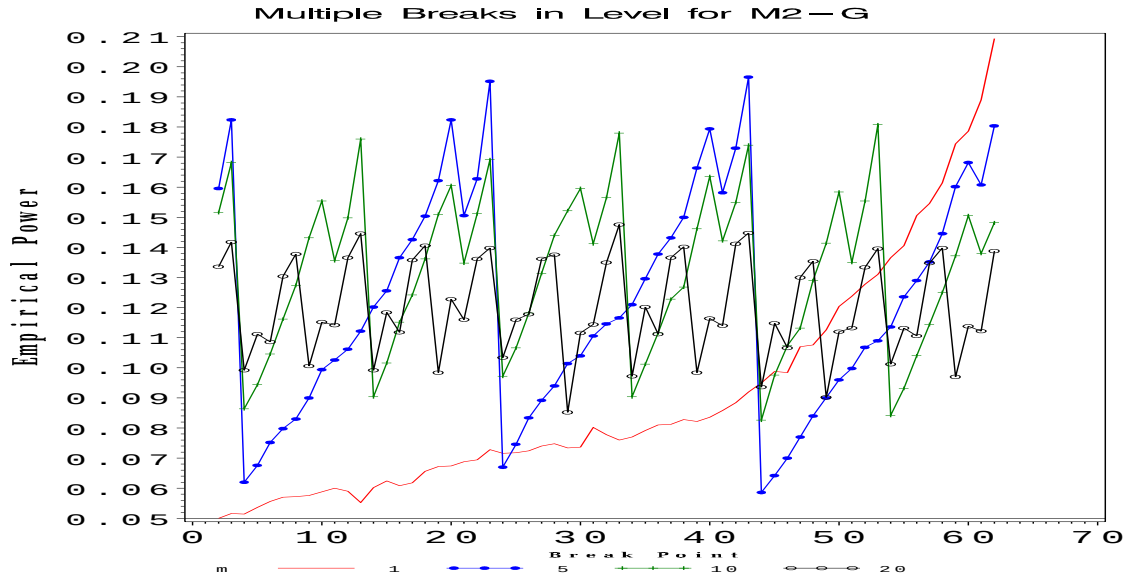


Figure 7.45: Empirical Power of $t_\mu(0.05)[T = 100, \rho = 0.9, \alpha_1 = 2.5, \alpha_2 = -5]$

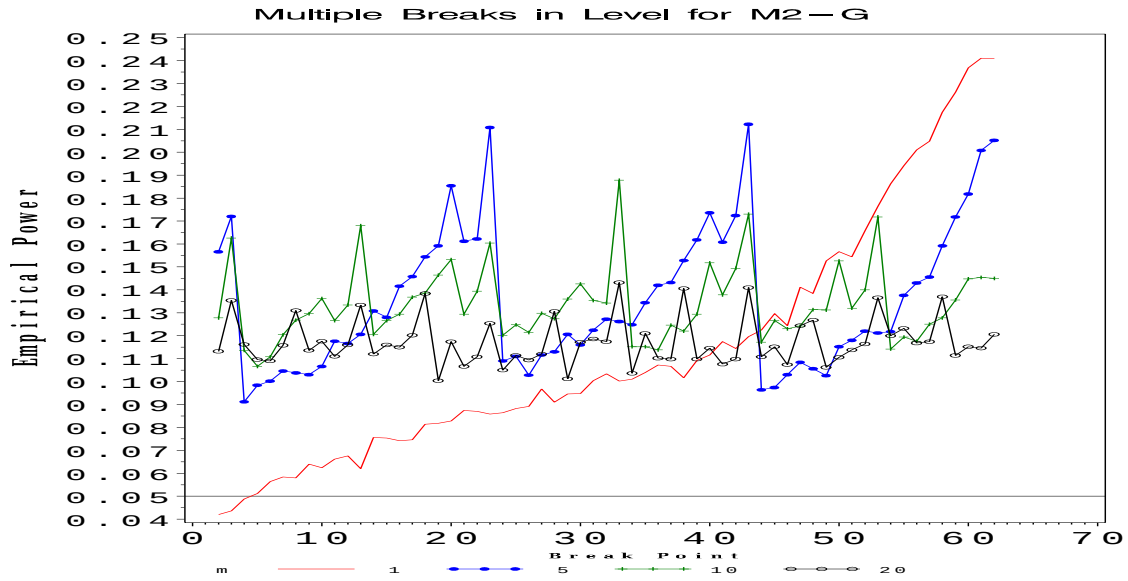


Figure 7.46: Size Adjusted Power of $t_\mu(0.05)[T = 100, \rho = 0.9, \alpha_1 = 2.5, \alpha_2 = -5]$