# การทดสอบความคงที่สำหรับตัวแบบการถดถอยในตัวอันดับหนึ่งที่ขึ้นกับความคลาดเคลื่อนจากการเลือกตัวอย่าง

นายวีรภัทร รัตนชัดเจน



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาคณิตศาสตร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2562

STATIONARY TEST FOR FIRST ORDER AUTOREGRESSIVE MODEL SUBJECT TO SAMPLING ERRORS

Mister Weerapat Rattanachadjane



A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Mathematics
Department of Mathematics and Computer Science
Faculty of Science
Chulalongkorn University
Academic Year 2019

| | |
|---|---|
| Thesis Title | STATIONARY TEST FOR FIRST ORDER AUTO-REGRESSIVE MODEL SUBJECT TO SAMPLING ERRORS |
| By | Mister Weerapat Rattanachadjane |
| Field of Study | Mathematics |
| Thesis Advisor | Assistant Professor Dr. Jiraphan Suntornchost |
| Thesis Co-Advisor | Professor Dr. Partha Lahiri |

---

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

.................................... Dean of the Faculty of Science

(Professor Polkit Sangvanich, Ph.D.)

THESIS COMMITTEE

.................................... Chairman

(Professor Kritsana Neammanee, Ph.D.)

.................................... Thesis Advisor

(Assistant Professor Jiraphan Suntornchost, Ph.D.)

.................................... Thesis Co-Advisor

(Professor Partha Lahiri, Ph.D.)

.................................... Examiner

(Assistant Professor Krung Sinapiromsaran, Ph.D.)

.................................... External Examiner

(Assistant Professor Thidaporn Supapakorn, Ph.D.)

วีรภัทร รัตนชัดเจน : การทดสอบความคงที่สำหรับตัวแบบการถดถอยในตัวอันดับหนึ่งที่ขึ้นกับความคลาดเคลื่อนจากการเลือกตัวอย่าง. (STATIONARY TEST FOR FIRST ORDER AUTOREGRESSIVE MODEL SUBJECT TO SAMPLING ERRORS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : ผศ.ดร.จิราพรรณ สุนทรโชติ, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม : ศ.ดร.พาร์ธา ลาฮีรี,  123 หน้า.

ในปี 1979 ดิกกี้และฟูลเลอร์ ได้เสนอการทดสอบความนิ่งบนตัวแบบถดถอยในตัวอันดับหนึ่งและลิมิตการกระจายของตัวประมาณค่าของสัมประสิทธิ์ถดถอยและสถิติทดสอบ กระบวนการดังกล่าวได้ถูกนำไปประยุกต์ใช้ในการตรวจสอบความนิ่งของอนุกรมเวลาบนตัวแบบถดถอยเฉลี่ยเคลื่อนที่ อย่างไรก็ตาม กระบวนการดังกล่าวได้ถูกนำไปประยุกต์ใช้โดยไม่คำนึงถึงความผิดพลาดจากการเลือกตัวอย่างซึ่งโดยปกติเกิดขึ้นจากการเก็บข้อมูล ในวิทยานิพนธ์นี้เราขยายงานศึกษาของดิกกี้และฟูลเลอร์ไปยังการประมาณค่าพารามิเตอร์และการทดสอบความนิ่งของข้อมูลบนตัวแบบถดถอยอันดับหนึ่งที่ขึ้นกับความคลาดเคลื่อนจากการเลือกตัวอย่าง



| | | | |
|---|---|---|---|
| ภาควิชา | คณิตศาสตร์และวิทยาการคอมพิวเตอร์ | ลายมือชื่อนิสิต | |
| สาขาวิชา | คณิตศาสตร์ | ลายมือชื่อ อ.ที่ปรึกษาหลัก | |
| ปีการศึกษา | 2562 | ลายมือชื่อ อ.ที่ปรึกษาร่วม | |

# # 6072101223 : MAJOR MATHEMATICS
KEYWORDS : AUTOREGRESSIVE COEFFICIENT/ PARAMETER ESTIMATION/ SAMPLING ERRORS / STATIONARY / TIME SERIES

WEERAPAT RATTANACHADJANE : STATIONARY TEST FOR FIRST ORDER AUTOREGRESSIVE MODEL SUBJECT TO SAMPLING ERRORS. ADVISOR : ASST. PROF. JIRAPHAN SUNTORNCHOST, Ph.D. CO-ADVISOR : PROF. PARTHA LAHIRI, Ph.D., 123 pp.

In 1979, Dickey and Fuller introduced a stationary test on the first order autoregressive model, AR(1), and limitting distribution of the estimator of autoregressive coefficient and the test statistics. The method has been applied to test the stationarity of the first order autoregressive time series model. However, the method has been applied regardless of sampling errors which usually occurs in data collection. In this thesis, we extend the study of Dickey and Fuller to introduce a parameter estimator of the autoregressive coefficient and a ststionary test for the first order autoregressive model subject to sampling errors.



| | |
|---|---|
| Department : Mathematics and Computer Science | Student's Signature ............ |
| Field of Study : Mathematics | Advisor's Signature ................ |
| Academic Year : 2019 | Co-Advisor's Signature ................ |

# ACKNOWLEDGEMENTS

First, I would like to express my deep gratitude to my thesis advisor, Assistant Professor Dr. Jiraphan Suntornchost, for her insightful suggestions and constant support through the thesis process. This thesis would not have been completed without her. Next, I would like to thank my co-advisor, Professor Dr. Partha Lahiri, for my inspiration and starting point of my thesis. I am most grateful for his advice. I also would like to thank my thesis committees, Professor Dr. Kritsana Neammanee, Assistant Professor Dr. Krung Sinapiromsaran, and Assistant Professor Dr. Thidaporn Supapakorn for their suggestions and comments. Moreover, I would like to thank all teachers who have educated and improved me for valuable knowledge.

Finally, I wish to thank the Development and Promotion of Science and Technology Talents Project (DPST) for financial support throughout my graduate study.

# CONTENTS

# LIST OF TABLES

page

# LIST OF FIGURES

# CHAPTER I
# INTRODUCTION

A time series is a sequence of observations indexed by times. Time series data can be found in many applications, for instance, forecast of daily closing prices in stock markets, explanation of the weekly speed of wind or daily air temperature in a specific location, and description of the annual unemployment rate. Among different types of time series models, one common family widely used in applications is the family of the autoregressive moving average (ARMA) models. The ARMA models have been widely applied in many forecasting problems. However, the models have a mandatory condition of being stationary, the condition such that mean, variance, and autocovariance do not depend on time. Therefore, in order to apply such models, an effective tool for the stationary test is important.

The stationary test was first introduced in Dickey and Fuller in 1979 [7]. They introduced a stationary test for the first order autoregressive model defined as



$$y_t = \rho y_{t-1} + e_t, \quad t = 1, 2, \ldots,$$



where $\{y_t\}_{t\geq 1}$ is the sequence of observed data, $\rho$ is the autoregressive coefficient, and $\{e_t\}$ is the sequence of Gaussian white noises with zero mean and variance $\sigma^2$. The absolute value of the autoregressive coefficient $\rho$ has an effect on the stationary condition. In particular, a time series $\{y_t\}$ is stationary if $|\rho| < 1$, and nonstationary if $|\rho| \geq 1$. Dickey and Fuller were interested in an estimator of $\rho$ and the test with the null hypothesis that the model is nonstationary, or $|\rho| = 1$. For $n$ observations of $y_t$, they used the maximum likelihood method for the estimator of $\rho$, denoted by $\hat{\rho}$. Rubin [19] showed that $\hat{\rho}$ is a consistent estimator of $\rho$. For $|\rho| < 1$, Mann and Wald [16] and Anderson [1] obtained the asymptotic distribution of $\sqrt{n}(\hat{\rho} - \rho)$, which is the normal distribution with zero mean and variance $1 - \rho^2$. White [25, 26] showed that when $|\rho| > 1$, the asymptotic distribution of $|\rho|^n(\rho^2 - 1)^{-1}(\hat{\rho} - \rho)$ is the Cauchy distribution. In addition, the asymptotic distribution of $\hat{\rho}$ and the corresponding test statistic of the stationary test when $|\rho| = 1$ is obtained by Dickey and Fuller [7]. Their results can be extended to higher orders of ARMA models. For example, Dickey et al. [6] extended the unit root test to autoregressive models of unknown order. Said and Dickey [21] developed a unit root test on ARMA models of unknown order and ARIMA models with order $(p, 1, q)$.

However, many stationary tests are applied to observed data regardless of sampling errors. When the studied data have sampling errors, such as the data obtained by survey sampling, the results are possibly different form the results obtained from the data with no sampling errors. The effects of measurement errors could cause bias results in many works of literature. For example, in the linear regression model, Lahiri and Suntornchost [12] showed that sampling errors could cause bias in variable selection methods. Marhuenda et al. [15] studied a residual likelihood ratio test for the variance component in Fay–Herriot model, the model is composed of a sampling model and a linking model. Angkunsit and Suntornchost [2] suggested ways to adjust variable selection statistics to reduce the biases of parameter estimates in the multivariate linear regression models subject to sampling errors.

In this thesis, we extend the concept of including sampling errors into parameter estimation and stationary test for the first order autoregressive models. We study the situation where the population means are not observed but are estimated by their sample means. We are interested in the estimator of autoregressive coefficients and the stationary test for the first order autoregressive models subject to sampling errors.

In our study, we first investigate the affect of sampling errors to the existing methods, called naive methods. We then propose some adjustments to reduce biases of the naive methods. Our study is divided into two main parts which are (1) the study of the autoregressive coefficient and (2) the study of the stationary test.

The organization of this thesis is as follows. In chapter 2, we give basic knowledge in probability, statistics, and time series analysis used in our study. In chapter 3, we introduce time series models subject to sampling errors. Moreover, we obtain a parameter estimate of the autoregressive coefficient, a test statistic for the stationary test, asymptotic properties, and simulations. Finally, we give conclusions of this thesis in Chapter 4.

# CHAPTER II

# PRELIMINARIES

In this chapter, we give definitions and theorems related to matrix algebra, calculus, some properties of probability and statistics, and time series models which will be used in this thesis.

## 2.1 Basic Knowledge in Matrix Algebra and Calculus

In this section, we introduce some notations of matrix algebra, which will be used in Chapter 3. For example, the transpose and the inverse of matrices. Furthermore, we discuss concepts of Taylor's formula for a function $f$ from $\mathbb{R}^n$ into $\mathbb{R}$.

### 2.1.1 Matrix Algebra

**Definition 2.1.** An $m \times n$ matrix $\mathbf{A}$ over a field $\mathbb{F}$ is a rectangular array of $mn$ elements $a_{ij}$, in $\mathbb{F}$, arranged in $m$ rows and $n$ columns as follows:



$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}. \tag{2.1}$$

**Definition 2.2.** The matrix obtained from a matrix $\mathbf{A}$ by interchanging the rows and columns of $\mathbf{A}$ is called the transpose of $\mathbf{A}$ and denoted by $\mathbf{A}'$.

**Definition 2.3.** Let $\mathbf{A}$ be an $n \times n$ matrix. $\mathbf{A}$ is invertible if and only if there exists a matrix $\mathbf{B}$ such that

$$\mathbf{AB} = I_n = \mathbf{BA},$$

where $I_n$ is $n \times n$ identity matrix. The matrix $\mathbf{B}$ is called the inverse of $\mathbf{A}$ and is denoted by $\mathbf{A}^{-1}$.

**Definition 2.4.** The determinant of an $n \times n$ matrix $\mathbf{A}$, denoted by $|\mathbf{A}|$ or $\det(\mathbf{A})$, is a scalar function of $\mathbf{A}$ defined as the sum of all $n!$ possible products of $n$ elements such that

1. each product contains one element from every row and every column of $\mathbf{A}$,

2. the factors in each product are the written so that the column subscripts appear in order of magnitude and each product is then preceded by a plus or minus sign according to whether the number of inversions in the row subscripts is even or odd.

**Theorem 2.5.** *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be invertable* $n \times n$ *matrices. Then*

1. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.

2. $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$.

3. $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

### 2.1.2 Taylor's Formula in Several Variables

**Definition 2.6.** [22] Let $a_1, a_2, \ldots, a_m, b_1, b_2, \ldots, b_m \in \mathbb{R}$. The set $(a_1, b_1) \times (a_2, b_2) \times \ldots \times (a_m, b_m) \subset \mathbb{R}^m$ is called an open rectangle of $\mathbb{R}^m$.

**Definition 2.7.** [22] A set $U \subset \mathbb{R}^m$ is called open if for each $\mathbf{x} \in U$ there is an open rectangle $A$ such that $\mathbf{x} \in A \subset U$.

**Definition 2.8.** [22] A subset $C$ of $\mathbb{R}^m$ is closed if $\mathbb{R}^m - C$ is open.

**Definition 2.9.** [22] Let $A \subset \mathbb{R}^n$. A function $f : A \to \mathbb{R}^m$ is called continuous at $\mathbf{a} \in A$ if $\lim\limits_{\mathbf{x} \to \mathbf{a}} f(\mathbf{x}) = f(\mathbf{a})$, and $f$ is simply called continuous if it is continuous at each $a \in A$.

**Theorem 2.10.** [22] *If* $A \subset \mathbb{R}^n$, *a function* $f : A \to \mathbb{R}^m$ *is continuous if and only if every open set* $U \subset \mathbb{R}^m$, *there is an open subset* $V$ *of* $\mathbb{R}^n$ *such that* $f^{-1}(U) = V \cap A$.

**Definition 2.11.** [22] A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{a} \in \mathbb{R}^n$ if there is a linear transformation $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\mathbf{h}\to 0} \frac{\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \lambda(\mathbf{h})\|}{\|\mathbf{h}\|} = 0. \tag{2.2}$$

The linear transformation $\lambda$ is denoted $Df(\mathbf{a})$ and called the derivative of $f$ at $\mathbf{a}$.

**Theorem 2.12.** [22] *If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{a} \in \mathbb{R}^n$, there is a unique linear transformation $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$\lim_{\mathbf{h}\to 0} \frac{\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \lambda(\mathbf{h})\|}{\|\mathbf{h}\|} = 0. \tag{2.3}$$

**Definition 2.13.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function and $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$. We say that $f$ has the $i$th partial derivative at $\mathbf{a}$ if the limit

$$\lim_{h\to 0} \frac{f(a_1, a_2, \ldots, a_i + h, \ldots, a_n) - f(a_1, a_2, \ldots, a_i, \ldots, a_n)}{h} \tag{2.4}$$

exists and is finite. The value is denoted by $D_i f(\mathbf{a})$. The $i$th partial derivative of $f$ is denoted by $D_i f$ or $\partial_i f$.

**Definition 2.14.** [10] A function $f$ is said to be of class $C^k$ on an open set $U$ if all of its partial derivatives of orders less than or equal to $k$ exist and are continuous on $U$.



If the partial derivatives of $f$ of all orders exist and are continuous on $U$, $f$ is said to be of class $C^\infty$ on $U$.

**Definition 2.15.** [10] A multi-index is an $n$-tuple of nonnegative integers. Multi-indices are generally denoted by the Greek letters $\alpha$ or $\beta$:

$$\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n), \quad \beta = (\beta_1, \beta_2, \ldots, \beta_n) \qquad (\alpha_j, \beta_j \in \{0, 1, 2, \ldots\}).$$

If $\alpha$ is a multi-index, we define

(1) $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_n$,

(2) $\alpha! = \alpha_1!\alpha_2! \cdots \alpha_n!$,

(3) $\mathbf{x}^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$, where $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$,

(4) $\partial^\alpha f = \partial_1^{\alpha_1} \partial_2^{\alpha_2} \cdots \partial_n^{\alpha_n} f$.

**Theorem 2.16** (Taylor's Theorem in Several Variables)**.** [10] *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is of class $C^k$ on an open convex set $S$. If $\mathbf{a} \in S$ and $\mathbf{a} + \mathbf{h} \in S$, then*

$$f(\mathbf{h} + \mathbf{a}) = \sum_{|\alpha| \le k} \frac{\partial^\alpha f(\mathbf{a})}{\alpha!} \mathbf{h}^\alpha + R_{\mathbf{a},k}(\mathbf{h}), \tag{2.5}$$

*where*

$$R_{\mathbf{a},k}(\mathbf{h}) = (k+1) \sum_{|\alpha| = k+1} \frac{\mathbf{h}^\alpha}{\alpha!} \int_0^1 (1-t)^k \partial^\alpha f(\mathbf{a} + t\mathbf{h}) dt. \tag{2.6}$$

**Corollary 2.17.** [10] *If $f$ is of class $C^k$ on $S$, then $\frac{R_{\mathbf{a},k}(\mathbf{h})}{|\mathbf{h}|^k} \to 0$ as $\mathbf{h} \to 0$. If $f$ is of class $C^{k+1}$ on $S$ and $|\partial^\alpha f(\mathbf{x})| \le M$ for $\mathbf{x} \in S$ and $|\alpha| = k+1$, then*

$$|R_{\mathbf{a},k}(\mathbf{h})| \le \frac{M}{(k+1)!} \|\mathbf{h}\|_1^{k+1}, \tag{2.7}$$

*where $\|\mathbf{h}\|_1 = \sum_{i=1}^n |h_i|$.*

**Lemma 2.18.** [10] *Let $P(\mathbf{h})$ be a polynomial of degree less than or equal to $k$. If $\lim_{\mathbf{h} \to 0} \frac{P(\mathbf{h})}{|\mathbf{h}|^k} = 0$, then $P \equiv 0$.*



**Theorem 2.19.** [10] *Suppose $f$ is of class $C^k$ near $\mathbf{a}$. If $f(\mathbf{a}+\mathbf{h}) = Q(\mathbf{h}) + E(\mathbf{h})$, where $Q$ is a polynomial with deg$(Q) \le k$ and $\lim_{\mathbf{h} \to 0} \frac{E(\mathbf{h})}{|\mathbf{h}|^k} = 0$, then $Q$ is the Taylor polynomial $P_{\mathbf{a},k}(\mathbf{h}) = \sum_{|\alpha| \le k} \frac{\partial^\alpha f(\mathbf{a})}{\alpha!} \mathbf{h}^\alpha$.*

**Definition 2.20.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a $C^k$ map an open convex set $S$ and $\mathbf{a} \in \mathbb{R}^n$. Let $\alpha = (\alpha_1, \ldots, \alpha_n)$ be a multi-index. If $k \ge 1$ or $\infty$, the first order Taylor approximation of $f$ around $\mathbf{a} \in S$ is

$$f(\mathbf{x}) \approx \sum_{|\alpha| \le 1} \frac{\partial^\alpha f(\mathbf{a})}{\alpha!} (\mathbf{x} - \mathbf{a})^\alpha = f(\mathbf{a}) + \sum_{i=1}^n \partial_i f(\mathbf{a})(x_i - a_i). \tag{2.8}$$

If $k \geq 2$ or $\infty$, the second order Taylor approximation of $f$ around $\mathbf{a} \in S$ is

$$\begin{aligned} f(\mathbf{x}) &\approx \sum_{|\alpha| \leq 2} \frac{\partial^\alpha f(\mathbf{a})}{\alpha!} (\mathbf{x} - \mathbf{a})^\alpha \\ &= f(\mathbf{a}) + \sum_{i=1}^{n} \partial_i f(\mathbf{a})(x_i - a_i) + \frac{1}{2} \sum_{i=1}^{n} \partial_i^2 f(\mathbf{a})(x_i - a_i)^2 \\ &\quad + \sum_{1 \leq i < j \leq n} \partial_{ij} f(\mathbf{a})(x_i - a_i)(x_j - a_j). \end{aligned} \tag{2.9}$$

## 2.2 Basic Knowledge in Probability Theory

In this section, we review some probability concepts used in our study. For example, important inequalities and convergence concepts of random variables. Furthermore, we provide expectation approximations of random variables by using the Taylor series.

### 2.2.1 Distribution and Moments

**Definition 2.21.** [3] The probability mass function (pmf) of a discrete random variable $X$ is given by

$$f_X(x) = P(X = x)$$

for all $x$.



**Definition 2.22.** [3] The probability density function (pdf) of a continuous random variable $X$ is the function $f_X(\cdot)$ that satisfies

$$P\left(X \leq x\right) = \int_{-\infty}^{x} f_X(t)dt,$$

for all $x$.

**Definition 2.23.** [3] The expected value or mean of a random variable $g(X)$, denoted by $\mathbb{E}\left(g(X)\right)$, is

$$\mathbb{E}\left(g(X)\right) = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x) dx, & \text{if } X \text{ is continuous,} \\ \sum_{-\infty}^{\infty} g(x) f_X(x), & \text{if } X \text{ is discrete,} \end{cases}$$

provided that the integral or the sum exists.

**Definition 2.24.** [3] For each integer $n$, the $n$th moment of a random variable $X$ is $\mathbb{E}(X^n)$, and the $n$th central moment of a random variable $X$ is $\mathbb{E}(X - \mathbb{E}(X))^n$.

**Definition 2.25.** [3] The variance of a random variable $X$ is its second order central moment, $\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}(X))^2$. The positive square root of $\mathrm{Var}(X)$ is the standard deviation of $X$.

**Definition 2.26.** [3] The covariance of random variables $X$ and $Y$ is defined by $\mathrm{Cov}(X, Y) = \mathbb{E}\left(\left(X - \mathbb{E}(X)\right)\left(Y - \mathbb{E}(Y)\right)\right)$.

**Theorem 2.27.** [3] *If $X$ and $Y$ are any two random variables and $a$ and $b$ are any two constants, then $\mathbb{E}(aX + bY) = a\,\mathbb{E}(X) + b\,\mathbb{E}(Y)$.*

**Theorem 2.28.** [3] *If $X$ and $Y$ are any two random variables and $a$ and $b$ are any two constants, then*



$$\mathrm{Var}(aX + bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X, Y).$$

**Theorem 2.29.** [3] *For any random variables $X$ and $Y$, $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y)$.*

## 2.2.2 Conditional Distributions

**Definition 2.30.** [3] Let $(X, Y)$ be a discrete bivariate random vector. Then the function $f(\cdot, \cdot)$ from $\mathbb{R}^2$ into $\mathbb{R}$ defined by $f(x, y) = P\left(X = x, Y = y\right)$ is called the joint probability mass function of $X$ and $Y$.

**Definition 2.31.** [3] Let $(X, Y)$ be a continuous bivariate random vector. A function $f$ from $\mathbb{R}^2$ into $\mathbb{R}$ is called a joint probability density function of $(X, Y)$ if, for any $A \subset \mathbb{R}^2$,

$$P((X,Y) \in A) = \iint_A f(x,y)\, dx\, dy.$$

**Definition 2.32.** [3] Let $X$ and $Y$ be discrete random variables. For any $x$ such that $P(X = x) > 0$, the conditional probability mass function of $Y$ given that $X = x$, denoted as $f_{Y|X}$, is the function of $y$ defined as

$$f_{Y|X}(y|x) = P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}.$$

**Definition 2.33.** [3] Let $X$ and $Y$ be continuous random variables. For any $x$ such that $f_X(x) > 0$, the conditional probability density function of $Y$ given that $X = x$, denoted as $f_{Y|X}$, is the function of $y$ defined as

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_Y(y)}.$$

**Theorem 2.34.** [3] *If $X$ and $Y$ are any two random variables, then*

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y)),$$



*provided that the expectations exist.*

**Theorem 2.35** (Conditional variance identity)**.** [3] *If $X$ and $Y$ are any two random variables, then*

$$\mathrm{Var}(X) = \mathbb{E}(\mathrm{Var}(X|Y)) + \mathrm{Var}(\mathbb{E}(X|Y)),$$

*provided that the expectations exist.*

### 2.2.3 Independence and Identical Distribution

**Definition 2.36.** [17] A pair $\{X, Y\}$ of random variables is (stochastically) independent if and only if for each pair of sets $\{M, N\}$, the following product rule holds

$$P(X \in M, Y \in N) = P(X \in M)P(Y \in N).$$

**Definition 2.37.** [17] A class $\{X_t \,|\, t \in T\}$ of random variables is said to be iid, an acronym for "independent and identically distributed", if and only if the class is independent and all members have the same distribution.

**Theorem 2.38.** [3] *If* $X$ *and* $Y$ *are independent random variables, then* $\mathrm{Cov}(X, Y) = 0$*.*

**Theorem 2.39.** [3] *A pair* $\{X, Y\}$ *of random variables is uncorrelated if and only if* $\mathrm{Cov}(X, Y) = 0$*.*

### 2.2.4 Convergence in Probability

Let $\{X_n\}$ and $\{Y_n\}$ be sequences of random variables for $n \geq 1$, and let $X$, $Y$ be random variables.

**Definition 2.40.** [18] The sequence of random variables $\{X_n\}$ converges in probability to $X$, written as $X_n \xrightarrow{p} X$, if for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(|X_n - X| > \varepsilon\right) = 0.$$

**Theorem 2.41.** [18] *Let* $g : \mathbb{R} \to \mathbb{R}$ *be a continuous function. If* $X_n \xrightarrow{p} X$*, then* $g\left(X_n\right) \xrightarrow{p} g\left(X\right)$*.*

**Theorem 2.42.** [18] *Suppose that* $X_n \xrightarrow{p} X$ *and* $Y_n \xrightarrow{p} Y$*. Then*

1. $X_n + Y_n \xrightarrow{p} X + Y$.

2. $X_n Y_n \xrightarrow{p} XY$.

**Theorem 2.43.** [11] *Suppose that* $|X_n - Y_n|$ *converges in probability to zero. If there exists a random variable* $X$ *such that* $X_n \xrightarrow{p} X$*, then* $Y_n \xrightarrow{p} X$.

**Definition 2.44.** [18] The sequence of random variables $\{X_n\}$ converges in $L_p$ to $X$, written as $X_n \xrightarrow{L_p} X$, if

$$\lim_{n\to\infty} \mathbb{E}\left(|X_n - X|^p\right) = 0.$$

**Theorem 2.45.** [18] *For* $p > 0$*, if* $X_n \xrightarrow{L_p} X$*, then* $X_n \xrightarrow{p} X$.

**Theorem 2.46.** [18] *For* $p > 0$*, if* $X_n \xrightarrow{L_p} X$*, then* $\mathbb{E}\left(|X_n|^p\right) \longrightarrow \mathbb{E}\left(|X|^p\right)$.

**Theorem 2.47.** *Let* $\{X_n\}$ *be a sequence of random variables with* $\mathbb{E}(X_n) = \mu$ *and* $\mathrm{Var}(X_n) \leq C < \infty$ *for some* $C$*. If* $\{X_n\}$ *are pairwise uncorrelated random variables, then*

*(1)* $\mathbb{E}\left(\left|\frac{1}{n}\sum_{t=1}^{n} X_t - \mu\right|^2\right) \longrightarrow 0,$

*(2)* $\frac{1}{n}\sum_{t=1}^{n} X_t \xrightarrow{p} \mu.$

**Theorem 2.48.** [11] *Let* $\{X_n\}$ *be a sequence of random variables. If* $\mu$ *is a constant such that* $\mathbb{E}(X_n) \to \mu$ *and* $\mathrm{Var}(X_n) \to 0$*, then* $X_n \xrightarrow{p} \mu$.

**Theorem 2.49.** [23] *Ler* $X_n$ *be a sequence of independent random variables with finite second moments. If*

$$\sum_{n=1}^{\infty} \mathrm{Var}(X_n) < \infty, \tag{2.10}$$

*then* $\sum_{n=1}^{\infty} (X_n - \mathbb{E}(X_n))$ *converges almost surely.*

**Definition 2.50.** [3] The sequence of random variables $\{X_n\}$ converges in distribution to $X$, written as $X_n \xrightarrow{d} X$, if

$$\lim_{n\to\infty} P\left(X_n \leq x\right) = P\left(X \leq x\right),$$

for every point $x$ and $F_X(x) = P\left(X \leq x\right)$ is continuous.

**Theorem 2.51.** [11] *If* $X_n \xrightarrow{p} X$, *then* $X_n \xrightarrow{d} X$.

**Theorem 2.52.** [11] *Let* $a$ *be a constant. If* $X_n \xrightarrow{d} a$, *then* $X_n \xrightarrow{p} a$.

**Theorem 2.53.** [11] *Let* $\{X_n\}$ *and* $\{Y_n\}$ *be sequences of random variables such that* $X_n - Y_n$ *converges in probability to zero. If there exists a random variable* $X$ *such that* $X_n \xrightarrow{d} X$, *then* $Y_n \xrightarrow{d} X$.

**Theorem 2.54.** *Let* $X_n, Y_n$, *and* $Z_n$ *be sequences of positive random variables for* $n = 1, 2, \ldots$. *Suppose that* $Y_n$ *and* $Z_n$ *both converge in probability to a random variable* $X$. *If* $Y_n \leq X_n \leq Z_n$, *then* $X_n$ *converges in probability to the random variable* $X$.

**Theorem 2.55** (Slutsky's Theorem,[3]). *Let* $\{X_n\}$ *and* $\{Y_n\}$ *be sequences of random variables. If* $X_n$ *converges in distribution to a random variable* $X$ *and* $Y_n$ *converges in probability to a constant* $a$, *then*

*(1)* $Y_n X_n \xrightarrow{d} aX$, *and*

*(2)* $X_n + Y_n \xrightarrow{d} X + a$.

### 2.2.5 Order in Probability



**Definition 2.56.** [11] Let $\{a_n\}$ be a sequence of real numbers and $\{g_n\}$ be a sequence of positive real numbers. Then, we say $a_n$ is of smaller order than $g_n$ and write $a_n = o(g_n)$ if $\lim_{n\to\infty} g_n^{-1} a_n = 0$.

**Definition 2.57.** [11] Let $\{a_n\}$ be a sequence of real numbers and $\{g_n\}$ be a sequence of positive real numbers. Then, we say thaty $a_n$ is at most of order $g_n$ and write $a_n = O(g_n)$ if there exist a real numbers $M$ and $N \in \mathbb{N}$ such that $g_n^{-1}|a_n| \leq M$ for all $n \geq N$.

**Lemma 2.58.** [11] *Let $\{a_n\}$ and $\{b_n\}$ be sequences of real numbers, and let $\{f_n\}$ and $\{g_n\}$ be sequences of positive real numbers.*

*(1) If $a_n = o(f_n)$ and $b_n = o(g_n)$, then*

$$\begin{aligned} a_n b_n &= o(f_n g_n), \\ |a_n|^s &= o(f_n^s) \quad \text{for } s > 0, \\ a_n + b_n &= o(\max\{f_n, g_n\}). \end{aligned}$$

*(2) If $a_n = O(f_n)$ and $b_n = O(g_n)$, then*

$$\begin{aligned} a_n b_n &= O(f_n g_n), \\ |a_n|^s &= O(f_n^s) \quad \text{for } s \geq 0, \\ a_n + b_n &= O(\max\{f_n, g_n\}). \end{aligned}$$

*(3) If $a_n = o(f_n)$ and $b_n = O(g_n)$, then $a_n b_n = o(f_n g_n)$.*

**Definition 2.59.** [11] Let $\{X_n\}$ be a sequence of random variables and $\{g_n\}$ be a sequence of positive real numbers. Then, we say that $X_n$ is of smaller order in probability than $g_n$ and write $X_n = o_p(g_n)$ if $g_n^{-1} X_n$ converges in probability to zero.



**Definition 2.60.** [11] Let $\{X_n\}$ be a sequence of random variables and $\{g_n\}$ be a sequence of positive real numbers. Then, we say that $X_n$ is at most of order in probability $g_n$ and write $X_n = O_p(g_n)$ if, for every $\varepsilon > 0$, there exists a positive real number $M$ such that $P\left(|g_n^{-1} X_n| \geq M\right) < \varepsilon$ for all $n$.

If $X_n = O_p(g_n)$, we sometimes say that $X_n$ is bounded in probability by $g_n$.

**Lemma 2.61.** [11] *Let $\{X_n\}$ and $\{Y_n\}$ be sequences of random variables, and let $\{f_n\}$ and $\{g_n\}$ be sequences of positive real numbers.*

*(1) If $X_n = o_p(f_n)$ and $Y_n = o_p(g_n)$, then*

$$\begin{aligned} X_nY_n &= o_p(f_ng_n), \\ |X_n|^s &= o_p(f_n^s) \quad \text{for } s > 0, \\ X_n + Y_n &= o_p(\max\{f_n, g_n\}). \end{aligned}$$

*(2) If $X_n = O_p(f_n)$ and $Y_n = O_p(g_n)$, then*

$$\begin{aligned} X_nY_n &= O_p(f_ng_n), \\ |X_n|^s &= O_p(f_n^s) \quad \text{for } s \geq 0, \\ X_n + Y_n &= O_p(\max\{f_n, g_n\}). \end{aligned}$$

*(3) If $X_n = o_p(f_n)$ and $Y_n = O_p(g_n)$, then $X_nY_n = o_p(f_ng_n)$.*

**Corollary 2.62.** [11] *Let $\{X_n\}$ be a sequence of random variables and $\{a_n\}$ be a sequence of positive real numbers. If $\mathbb{E}\,(X_n^2) = O(a_n^2)$, then $X_n = O_p(a_n)$.*

**Corollary 2.63.** [11] *Let $\{X_n\}$ be a sequence of random variables and $\{a_n\}$ be a sequence of positive real numbers. If $\mathrm{Var}\,(X_n) = O(a_n^2)$ and $\mathbb{E}\,(X_n) = O(a_n)$, then $X_n = O_p(a_n)$.*



**Corollary 2.64.** [11] *Let $\{X_n\}$ be a sequence of scalar random variables such that*

$$X_n = a + O_p(r_n),$$

*where $r_n \to 0$ as $n \to \infty$. If $g$ is a function with $s$ continuous derivatives at $x = a$, then*

$$g(X_n) = g(a) + g^{(1)}(a)(X_n - a) + \cdots + g^{(s)}(a)(X_n - a)^s + O_p(r_n^s),$$

*where $g^{(j)}(a)$ is the $j$th derivative of $g(x)$ evaluated at $x = a$.*

### 2.2.6 Inequalities

**Theorem 2.65.** [18] *Let $X$ and $Y$ be random variables. If $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ both exist and $X \leq Y$, then $\mathbb{E}(X) \leq \mathbb{E}(Y)$.*

**Theorem 2.66** (Modulus Inequality, [18]). *For any random variable $X$, if $\mathbb{E}(|X|) < \infty$, then $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.*

**Theorem 2.67** (Markov's Inequality, [18]). *For any random variable $X$ and $\lambda > 0$, if $\mathbb{E}(|X|) < \infty$, then*

$$P(|X| \geq \lambda) \leq \frac{\mathbb{E}(|X|)}{\lambda}.$$

**Theorem 2.68** (Chebychev's Inequality, [18]). *For any random variable $X$ and $\varepsilon > 0$, if $\mathbb{E}(|X|) < \infty$ and $\mathrm{Var}(X) \leq \infty$, then*

$$P(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon}.$$

**Theorem 2.69** (Chebychev's Inequality, [11]). *Let $r > 0$, and let $X$ be a random variable such that $\mathbb{E}(|X|^r) < \infty$. Then, for any $\varepsilon > 0$ and finite $A$,*

$$P(|X - A| \geq \varepsilon) \leq \frac{\mathbb{E}(|X - A|^r)}{\varepsilon^r}.$$



**Theorem 2.70** (Cauchy-Schwarz Inequality, [3]). *For any two random variables $X$ and $Y$, if $\mathbb{E}(X^2) < \infty$ and $\mathbb{E}(Y^2) < \infty$, then*



$$|\mathbb{E}(XY)| \leq \mathbb{E}(|XY|) \leq \sqrt{\mathbb{E}(X^2)\,\mathbb{E}(Y^2)}.$$

**Corollary 2.71.** [3] *For any two random variables $X$ and $Y$, if $\mathbb{E}(X^2) < \infty$ and $\mathbb{E}(Y^2) < \infty$, then*

$$|\mathrm{Cov}(X, Y)| \leq \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}.$$

**Theorem 2.72** (Hölder's Inequality, [3]). *Let $X$ and $Y$ be any two random variables, and let $p$ and $q$ be any positive numbers satisfying $p > 1$, $q > 1$, and $\frac{1}{p} + \frac{1}{q} = 1$. If $\mathbb{E}(|X|^p) < \infty$ and $\mathbb{E}(|Y|^q) < \infty$, then*

$$|\mathbb{E}(XY)| \leq \mathbb{E}(|XY|) \leq \mathbb{E}^{\frac{1}{p}}(|X|^p)\,\mathbb{E}^{\frac{1}{q}}(|Y|^q).$$

**Theorem 2.73** (Minkowski's Inequality, [3]). *Let $X$ and $Y$ be any two random variables. If $\mathbb{E}(|X|^p) < \infty$ and $\mathbb{E}(|Y|^p) < \infty$ for $1 \leq p < \infty$, then*

$$\mathbb{E}^{\frac{1}{p}}(|X+Y|^p) \leq \mathbb{E}^{\frac{1}{p}}(|X|^p) + \mathbb{E}^{\frac{1}{p}}(|Y|^p).$$

**Corollary 2.74.** *Let $X$ and $Y$ be any two random variables. If $\mathbb{E}(X^2) < \infty$ and $\mathbb{E}(Y^2) < \infty$, then*

$$\operatorname{Var}(X+Y) \leq \left(\sqrt{\operatorname{Var}(X)} + \sqrt{\operatorname{Var}(Y)}\right)^2 \leq 2\operatorname{Var}(X) + 2\operatorname{Var}(Y).$$

*Proof.* From Minkowski's inequality with $p = 2$, we have

$$\begin{aligned}
\sqrt{\operatorname{Var}(X+Y)} &= \sqrt{\mathbb{E}\left(X+Y-\mathbb{E}(X+Y)\right)^2} \\
&= \sqrt{\mathbb{E}\left((X-\mathbb{E}(X)) + (Y-\mathbb{E}(Y))\right)^2} \\
&\leq \sqrt{\mathbb{E}\left(X-\mathbb{E}(X)\right)^2} + \sqrt{\mathbb{E}\left(Y-\mathbb{E}(Y)\right)^2} \\
&= \sqrt{\operatorname{Var}(X)} + \sqrt{\operatorname{Var}(Y)}.
\end{aligned}$$

Hence, $\operatorname{Var}(X+Y) \leq \left(\sqrt{\operatorname{Var}(X)} + \sqrt{\operatorname{Var}(Y)}\right)^2$.

By using the fact that $(a+b)^2 \leq 2a^2 + 2b^2$, we have $\left(\sqrt{\operatorname{Var}(X)} + \sqrt{\operatorname{Var}(Y)}\right)^2 \leq 2\operatorname{Var}(X) + 2\operatorname{Var}(Y)$. □

**Theorem 2.75** (Jensen's Inequality, [3]). *For any random variable $X$, if $g(\cdot)$ is a convex function, then*

$$g(E(X)) \leq \mathbb{E}(g(X)).$$

*If $g(\cdot)$ is concave, the inequality reverses.*

### 2.2.7 Expectation Approximations of Random Variables

In many situations, we have to work with functions of random variables where the closed-forms of expectation and variance are difficult to obtain. In the situation where we are interested in obtaining the variance of the estimate of $\frac{p}{1-p}$, where $p$ is the success probability of the Bernoulli experiment. The odds $\frac{p}{1-p}$ can be estimated by $\frac{\hat{p}}{1-\hat{p}}$, where $\hat{p}$ is the estimator of the success probability $p$. To consider the properties of this estimator $\frac{\hat{p}}{1-\hat{p}}$, the variance of the estimator is necessary. However, the closed-form of its variance is impossible to obtain.

One method for approximation of the mean and the variance of a function of a random variable is to use the Taylor series. This approximation method allows us to approximate the mean and the variance of random variables.

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a vector of random variables with mean $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)$. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. From (2.8), the first-order Taylor series expansion of $f(\mathbf{X})$ about $\boldsymbol{\mu}$ is

$$f(\mathbf{X}) = (\boldsymbol{\mu}) + \sum_{i=1}^{n} f_i(\boldsymbol{\mu})(X_i - \mu_i) + R, \tag{2.11}$$

where $f_i = \dfrac{\partial f}{\partial x_i}$ and $R$ is the remainder of order smaller than the term in the equation. For approximating the mean and the variance of $f(\mathbf{X})$, we ignore the remainder $R$ and write



$$f(\mathbf{X}) \approx f(\boldsymbol{\mu}) + \sum_{i=1}^{n} f_i(\boldsymbol{\mu})(X_i - \mu_i). \tag{2.12}$$

We apply the expectation to all terms in (2.12) to obtain

$$\mathbb{E}\left(f(\mathbf{X})\right) \approx f(\boldsymbol{\mu}). \tag{2.13}$$

In the same way by applying the variance function to all terms in (2.12), the approximation of the variance of $f(\mathbf{X})$ is

$$\operatorname{Var}\left(f(\mathbf{X})\right) \approx \sum_{i=1}^{n} f_i^2(\boldsymbol{\mu}) \operatorname{Var}(X_i) + 2 \sum_{1 \le i < j \le n} f_i(\boldsymbol{\mu}) f_j(\boldsymbol{\mu}) \operatorname{Cov}(X_i, X_j). \tag{2.14}$$

The approximation (2.12) can be extended to the higher order to reduce approximation errors, such as the second-order Taylor series approximation. From (2.9), the second-order Taylor series approximation of $f(\mathbf{X})$ about $\boldsymbol{\mu}$ is

$$\begin{aligned} f(\mathbf{X}) \approx f(\boldsymbol{\mu}) &+ \sum_{i=1}^{n} f_i(\boldsymbol{\mu})(X_i - \mu_i) + \frac{1}{2} \sum_{i=1}^{n} f_{ii}^2(\boldsymbol{\mu})(X_i - \mu_i)^2 \\ &+ \sum_{1 \le i < j \le n} f_{ij}^2(\boldsymbol{\mu})(X_i - \mu_i)(X_j - \mu_j), \end{aligned} \tag{2.15}$$

where $f_i = \dfrac{\partial f}{\partial x_i}$.

Applying the expectation to the terms in (2.15), the approximation of the expectation of $f(\mathbf{X})$ is

$$\mathbb{E}\left(f(\mathbf{X})\right) \approx f(\boldsymbol{\mu}) + \frac{1}{2} \sum_{i=1}^{n} f_{ii}^2(\boldsymbol{\mu}) \operatorname{Var}(X_i) + \sum_{1 \le i < j \le n} f_{ij}^2(\boldsymbol{\mu}) \operatorname{Cov}\left(X_i, X_j\right). \tag{2.16}$$

## 2.3 Basic Knowledge in Statistics



In this section, we introduce some concepts in statistics used in this thesis such as the parameter estimation and the hypothesis testing.

### 2.3.1 Parameter Estimation

**Definition 2.76.** [3] A point estimator is any function $\mathbf{W}(X_1, X_2, \ldots, X_n)$ of samples $X_1, X_2, \ldots, X_n$.

**Definition 2.77.** [3] Let $f(\mathbf{x}|\boldsymbol{\theta})$ denote the joint pdf or pmf of the sample $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ and $\boldsymbol{\theta}$ denote the parameter. Then, given that $\mathbf{X} = \mathbf{x}$ is observed, the function of $\boldsymbol{\theta}$ defined by

$$L(\boldsymbol{\theta}|\mathbf{x}) = f(\mathbf{x}|\boldsymbol{\theta})$$

is called the likelihood function of $\boldsymbol{\theta}$.

**Definition 2.78.** [3] For each sample point $\mathbf{x}$, let $\hat{\boldsymbol{\theta}}(\mathbf{x})$ be a parameter value at which $L(\boldsymbol{\theta}|\mathbf{x})$ attains its maximum as a function of $\boldsymbol{\theta}$, with $\mathbf{x}$ held fixed. A maximum likelihood estimator (MLE) of the parameter $\boldsymbol{\theta}$ based on a sample $\mathbf{X}$ is $\hat{\boldsymbol{\theta}}(\mathbf{X})$.

**Definition 2.79.** [3] The mean squared error (MSE) of an estimator $\mathbf{W}$ of a parameter $\boldsymbol{\theta}$ is the function of $\boldsymbol{\theta}$ defined by $\mathbb{E}_{\boldsymbol{\theta}}\left(\mathbf{W}-\boldsymbol{\theta}\right)^2$.

**Definition 2.80.** [3] The bias of a point estimator $\mathbf{W}$ of a parameter $\boldsymbol{\theta}$ is is the difference between the expected value of $\mathbf{W}$ and $\boldsymbol{\theta}$; that is, $\text{Bias}_{\boldsymbol{\theta}}(\mathbf{W}) = \mathbb{E}_{\boldsymbol{\theta}}(\mathbf{W}) - \boldsymbol{\theta}$.

**Definition 2.81.** [3] An estimator $\mathbf{W}$ whose bias is identically equal to 0 is called unbiased and satisfies $\text{Bias}_{\boldsymbol{\theta}}(\mathbf{W}) = 0$ for all $\boldsymbol{\theta}$.

**Definition 2.82.** [3] A sequence of estimators $\mathbf{W}_n = \mathbf{W}_n(X_1, X_2, \ldots, X_n)$ is a consistent sequence of estimators of the parameter $\boldsymbol{\theta}$ if $\mathbf{W}_n$ converges to $\boldsymbol{\theta}$ in probability.



### 2.3.2 Hypothesis Testing

**Definition 2.83.** [3] A hypothesis is a statement about a population parameter.

**Definition 2.84.** [3] The two complementary hypotheses in a hypothesis testing problem are called the null hypothesis and the alternative hypothesis. They are denoted by $H_0$ and $H_1$, respectively.

**Definition 2.85.** [3] A hypothesis testing procedure or hypothesis test is a rule that specifies:

1. For which sample value the decision is made to accept $H_0$ as true.

2. For which sample value $H_0$ is rejected and $H_1$ is accepted as true.

The subset of the sample space for which $H_0$ will be rejected is called the rejection region or the critical region. The complement of the rejection region is called the acceptance region.

**Definition 2.86** (Likelihood Ratio Tests,[3])**.** The likelihood ratio test statistic for testing $H_0 : \boldsymbol{\theta} \in \Theta_0$ versus $H_1 : \boldsymbol{\theta} \notin \Theta_0$ is

$$\lambda(\mathbf{x}) = \frac{\sup\limits_{\boldsymbol{\theta} \in \Theta_0} L(\boldsymbol{\theta} \,|\, \mathbf{x})}{\sup\limits_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta} \,|\, \mathbf{x})}.$$

A likelihood ratio test (LRT) is any test that has a rejection region of the form $\{\mathbf{x} \,|\, \lambda(\mathbf{x}) \leq c\}$, where $c$ is any number satisfying $0 \leq c \leq 1$.

## 2.4 Time Series and Stationarity

In this section, we introduce some concepts in time series and their properties used in this thesis such as stationarity. Furthermore, we discuss some time series models, such as an autoregressive model and a moving average model.

**Definition 2.87.** [11] Let $T$ be an index set. A real valued time series, or stochastic process, is a real valued random variable $X_t$, where $t$ belongs to a time index set $T$.

**Definition 2.88.** [11] A time series $\{X_t\}_{t \in T}$ is called strictly stationary if for any $n, h > 0$ and any sets of indices $t_1, t_2, \ldots, t_n$ and $t_1 + h, t_2 + h, \ldots, t_n + h$ in the index set $T$,

$$\begin{aligned} &P(X_{t_1} \leq x_1, X_{t_2} \leq x_2, \ldots, X_{t_n} \leq x_n) \\ &\qquad = P(X_{t_1+h} \leq x_1, X_{t_2+h} \leq x_2, \ldots, X_{t_n+h} \leq x_n), \end{aligned} \tag{2.17}$$

where $(x_1, x_2, \ldots, x_n)$ is in the range of the random variable $X_t$.

**Definition 2.89.** [24] For a given real valued time series $\{X_t\}_{t\in T}$,

1. the mean function of the process defined by

$$\mu_t = \mathbb{E}(X_t),$$

2. the variance function of the process defined by

$$\sigma_t^2 = \mathrm{Var}(X_t),$$

3. the covariance function between $X_{t_1}$ and $X_{t_2}$ the process defined by

$$\gamma(t_1, t_2) = \mathrm{Cov}(X_{t_1}, X_{t_2}),$$

4. the correlation function between $X_{t_1}$ and $X_{t_2}$ the process defined by

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\sigma_{t_1}^2}\sqrt{\sigma_{t_2}^2}}.$$

**Definition 2.90.** [11] A time series $\{X_t\}$ is weakly stationary if

1. $E[X_t]$ is a constant for all $t$, and

2. For each $h$, $\gamma(t, t+h)$ is independent of $t$.



**Definition 2.91.** [24] A process $\{e_t\}$ is called a white noise process if it is a sequence of uncorrelated random variables from a fixed distribution with constant mean $\mathbb{E}(e_t) = \mu_e$, usually assumed to be 0, constant variance $\mathrm{Var}(e_t) = \sigma_e^2$, and $\gamma_k = \mathrm{Cov}(e_t, e_{t+k}) = 0$ for all $k \neq 0$.

**Definition 2.92.** The sequence $\{X_t\}$ defined by

$$X_t = \sum_{j=0}^{q} \beta_j e_{t-j},$$

where $\beta_0, \beta_q \neq 0$ and $\{e_t\}_{t\geq 1}$ are uncorrelated $(0, \sigma^2)$ random variables, is called a moving average model of order $q$, denoted by MA($q$).

**Definition 2.93.** The sequence $\{X_t\}$ defined by

$$\sum_{i=0}^{p} \alpha_i X_{t-i} = e_t,$$

where $\alpha_0, \alpha_q \neq 0$ and $\{e_t\}_{t\geq 1}$ are uncorrelated $(0, \sigma^2)$ random variables, is called an autoregressive model of order $p$, denoted by AR($p$). A sequence $\{e_t\}_{t\geq 1}$ is sometimes called white noise process.

**Definition 2.94.** The sequence $\{X_t\}$ defined by

$$X_t + a_1 X_{t-1} + \cdots + a_p X_{t-p} = e_t + b_1 e_{t-1} + \cdots + b_q e_{t-q},$$

where $a_p, b_q \neq 0$ and $\{e_t\}_{t\geq 1}$ are uncorrelated $(0, \sigma^2)$ random variables, is called an autoregressive moving average model of orders $(p, q)$, denoted by ARMA($p, q$).

# CHAPTER III

# FIRST ORDER AUTOREGRESSIVE MODEL SUBJECT TO SAMPLING ERRORS

In this chapter, we propose a new parameter estimation and a stationary test for the first order autoregressive model subject to sampling errors, called AR(1) model subject to sampling errors. The organization of this chapter is as follows. In Section 3.1, we give introduction of the stationary test of the AR(1) model and introduce the AR(1) model subject to sampling errors. In Section 3.2, we study the parameter estimation of the autoregressive coefficient of the AR(1) model subject to sampling errors. The stationary test for the AR(1) model subject to sampling errors is provided in Section 3.3. The positive adjustments to the estimators are discussed in Section 3.4. The simulation study is discussed in Section 3.5. The structure of this chapter is as follows.

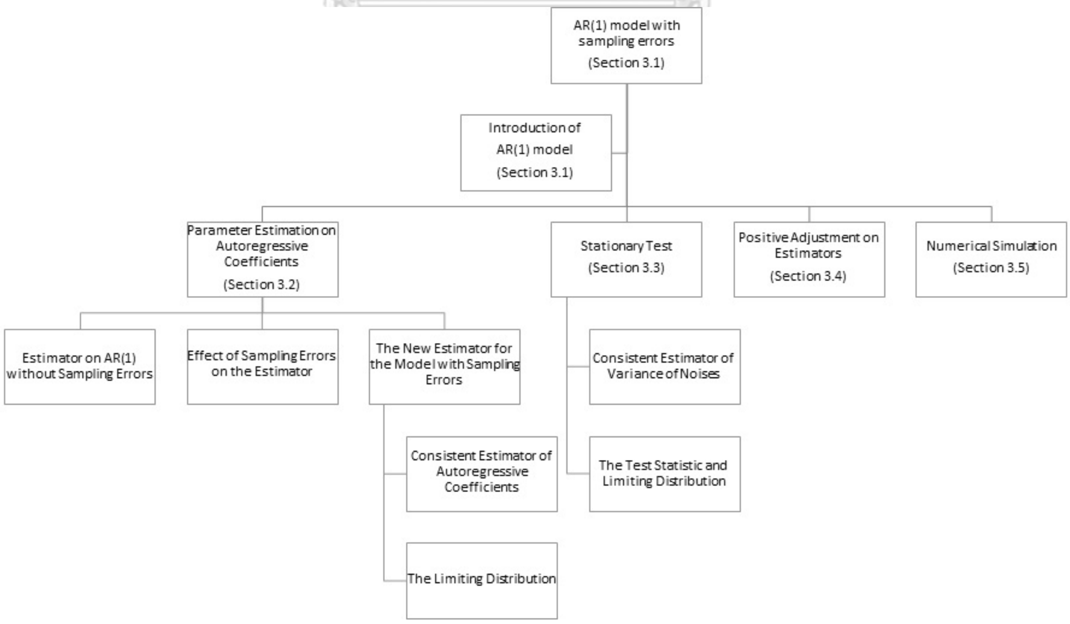



Figure 3.1: The structure of Chapter 3: First order autoregressive model subject to sampling errors.

## 3.1 The First Order Autoregressive Model

In this section, we first introduce some Dickey and Fuller results ([5, 7]) for the stationary test and the asymptotic distribution of the test statistic for the first order autoregressive model in Section 3.1.1. In Section 3.1.2, we consider the first order autoregressive model subject to sampling errors and introduce the concept of stationary for the model with sampling errors.

### 3.1.1 Introduction to Stationary Test in AR(1)

We will consider the first order autoregressive process

$$\theta_t = \rho\theta_{t-1} + \eta_t, \quad t = 1, \ldots, T, \tag{3.1}$$

where $\{\theta_t\}_{t\geq 1}$ is the sequence of the variable of interest with $\theta_0 = 0$ and $\{\eta_t\}_{t\geq 1}$ is a sequence of independent and identically distributed (iid) $N(0, \sigma^2)$ distributed random variables. Notice from (3.1) that the AR(1) model can be written to linear combination of $\eta_t$ as $\theta_t = \sum_{i=1}^{t} \rho^{t-i}\eta_i$.

If the model (3.1) is stationary, the mean and variance of $\theta_t$ are constants $\mu$ and $\gamma$. Then, by substituting constant variance function on the model, we know that the AR(1) model is stationary if and only if the absolute value of autoregressive coefficient $\rho$ is less than 1. When $|\rho| = 1$, the time series $\theta_t$ is not stationary and the variance of $\theta_t$ depends on $t$, which is $t\sigma^2$. If $|\rho| > 1$, then variance of $\theta_t$ increases as $t$ increases without bound as sufficiently large $t$, this process is called explosive.

Dickey and Fuller were interested in the estimator of the autoregressive coefficient $\rho$ and the stationary test. The stationary test is to test the null hypothesis that $|\rho| = 1$ or nonstationary against the alternative hypothesis $|\rho| \neq 1$ or stationary. Dickey-Fuller stationary test is to perform the likelihood ratio test defined as

$$\lambda(\boldsymbol{\theta}) = \frac{\sup\limits_{\rho=1} L(\rho, \sigma^2 \,|\, \boldsymbol{\theta})}{\sup\limits_{\rho\in\Theta} L(\rho, \sigma^2 \,|\, \boldsymbol{\theta})}, \tag{3.2}$$

where $\Theta$ is the parameter space and $L(\rho, \sigma^2 \,|\, \boldsymbol{\theta})$ is the likelihood function of parameter $\rho$, and $\sigma^2$ given observations $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_T)'$, with the rejection region $\{\boldsymbol{\theta} \,|\, \lambda(\boldsymbol{\theta}) \leq c\}$, where $0 \leq c \leq 1$. To evaluate the test statistic $\lambda(\boldsymbol{\theta})$, it is sufficient to compute the maximum likelihood estimators of $\rho$ and $\sigma^2$. Notices that the computation of the estimators is greatly simplified if $\theta_1$ is fixed and consider conditional likelihood. It is reasonable in some situations when we know the initial value at time 1 with $\theta_1$. Hence, the conditional log-likelihood with $\theta_1$ as fixed is

$$\log L(\rho, \sigma^2 \,|\, \boldsymbol{\theta}) = -(\frac{T-1}{2}) \log 2\pi - (\frac{T-1}{2}) \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=2}^{T} (\theta_t - \rho\theta_{t-1})^2. \tag{3.3}$$

By setting the partial derivatives of $\log L(\rho, \sigma^2 \,|\, \boldsymbol{\theta})$ with respect $\rho$ and $\sigma^2$ equal to zero:

$$0 = \frac{\partial l}{\partial \rho} = \frac{1}{\sigma^2} \sum_{t=1}^{T} (\theta_{t-1})(\theta_t - \rho\theta_{t-1}),$$

and

$$0 = \frac{\partial l}{\partial \sigma^2} = -\frac{T-1}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=2}^{T} (\theta_t - \rho\theta_{t-1})^2,$$



the maximum likelihood estimators of $\rho$ and $\sigma^2$ are

$$\hat{\rho}_{ML} = \frac{\sum\limits_{t=2}^{T} \theta_t \theta_{t-1}}{\sum\limits_{t=2}^{T} \theta_{t-1}^2}, \tag{3.4}$$

and

$$\hat{\sigma}^2_{ML} = \frac{1}{T-1} \sum_{t=2}^{T} (\theta_t - \hat{\rho}_{ML}\theta_{t-1})^2, \tag{3.5}$$

respectively. By substituting the estimators into (3.2), the likelihood ratio test rejects the null hypothesis when likelihood ratio test statistic

$$\lambda(\boldsymbol{\theta}) = \left( \frac{\sum_{t=2}^{T} (\theta_t - \theta_{t-1})^2}{\sum_{t=2}^{T} (\theta_t - \hat{\rho}_{ML}\theta_{t-1})^2} \right)^{-\frac{T-1}{2}} \tag{3.6}$$

is less than or equal to a constant $c$ when $c \in [0, 1]$. The constant $c$ is determined by the distribution of the statistic $\lambda(\boldsymbol{\theta})$. Notice that

$$\sum_{t=2}^{T} (\theta_t - \theta_{t-1})^2 = \sum_{t=2}^{T} (\theta_t - \hat{\rho}_{ML}\theta_{t-1})^2 + (1 - \hat{\rho}_{ML})^2 \sum_{t=2}^{T} \theta_{t-1}^2.$$

Therefore, the likelihood ratio test rejects null hypothesis, $\rho = 1$, when

$$|\tau| = \left| \frac{(\hat{\rho}_{ML} - 1)\sqrt{\sum_{t=2}^{T} \theta_{t-1}^2}}{\sqrt{\frac{1}{T-2} \sum_{t=2}^{T} (\theta_t - \hat{\rho}_{ML}\theta_{t-1})^2}} \right| \geq \sqrt{(a-1)(T-2)}, \tag{3.7}$$

where $a = c^{-\frac{2}{T-1}}$ is a constant greater than 1.

Dickey and Fuller considered the asymptotic distribution of the test statistic $\tau$ and the estimate of autoregressive coefficient $\hat{\rho}_{ML}$. Mann and Wald [16] showed that $\hat{\rho}_{ML}$ is a consistent estimator of $\rho$ for $|\rho| < 1$. Rubin [19] showed that under certain regularity conditions $\hat{\rho}_{ML}$ is a consistent estimator of $\rho$ for $|\rho| \geq 1$. Moreover, Dickey [5] showed that the estimator of $\sigma^2$ is a consistent estimator of $\sigma^2$ for each $\rho$. The asymptotic distribution of $\hat{\rho}_{ML}$ obtained in White [25, 26] for the case $|\rho| > 1$, and Anderson [1] for case $|\rho| < 1$, and Dickey [5] for case $|\rho| = 1$, respectively. The following theorems show some results of asymptotic distributions for the case $|\rho| \leq 1$, obtained in [1] and [5].

**Theorem 3.1** ([1]). *Assume the model (3.1) and $|\rho| < 1$. Define $\hat{\rho}_{ML} - \rho = \frac{A_T}{B_T}$, where $A_T = \sum_{t=2}^{T} \theta_t\theta_{t-1} - \rho \sum_{t=2}^{T} \theta_{t-1}^2$ and $B_n = \sum_{t=2}^{T} \theta_{t-1}^2$. Then,* *(1)* $\dfrac{A_T}{\sqrt{T}} \xrightarrow{d} N\left(0, \dfrac{\sigma^4}{1-\rho^2}\right)$,

*(2)* $\dfrac{B_T}{T} \xrightarrow{p} \dfrac{\sigma^2}{1-\rho^2}$,

*(3)* $\sqrt{T}(\hat{\rho}_{ML} - \rho) \xrightarrow{d} N(0, 1-\rho^2)$.

**Theorem 3.2** ([5])**.** *Assume the model (3.1) and* $\rho = 1$. *Define* $\hat{\rho}_{ML} - \rho = \dfrac{A_T}{B_T}$, *where* $A_T = \sum_{t=2}^{T} \theta_t\theta_{t-1} - \sum_{t=2}^{T} \theta_{t-1}^2$ *and* $B_T = \sum_{t=2}^{T} \theta_{t-1}^2$. *Let* $\gamma_i = (-1)^{i+1}\dfrac{2}{(2i-1)\pi}$. *Then,*

*(1)* $\dfrac{A_T}{T} \xrightarrow{p} \dfrac{1}{2}\left(\sum_{i=1}^{\infty} \sqrt{2}\gamma_i Z_i'\right)^2 - \dfrac{\sigma^2}{2}$, *where* $Z_i' \overset{iid}{\sim} N(0, \sigma^2)$,

*(2)* $\frac{B_T}{T^2} \xrightarrow{p} \sum_{i=1}^{\infty} \gamma_i^2 Z_i'^2$, *where* $Z_i' \overset{iid}{\sim} N(0, \sigma^2)$,

*(3)* $\hat{\rho}_{ML} - 1 = O_p(T^{-1})$ *and* $T(\hat{\rho}_{ML} - 1) \xrightarrow{d} \dfrac{\left(\sum_{i=1}^{\infty} \sqrt{2}\gamma_i Z_i\right)^2 - 1}{2\sum_{i=1}^{\infty} \gamma_i^2 Z_i^2}$,

*where* $Z_i \overset{iid}{\sim} N(0,1)$,

*(4)* $\tau \xrightarrow{d} \dfrac{\left(\sum_{i=1}^{\infty} \sqrt{2}\gamma_i Z_i\right)^2 - 1}{2\sqrt{\sum_{i=1}^{\infty} \gamma_i^2 Z_i^2}}$, *where* $Z_i \overset{iid}{\sim} N(0,1)$.



From Theorem 3.2, the terms of $\sum_{i=1}^{\infty} \sqrt{2}\gamma_i Z_i$ and $\sum_{i=1}^{\infty} \gamma_i^2 Z_i^2$ are both infinite sums of random variables. However, Dickey [5] showed that all of infinite sums in Theorem 3.2 are well-defined.

### 3.1.2 AR(1) with Sampling Errors

In this section, we introduce the first order autoregressive models subject to sampling errors. In particular, we consider the autoregressive of order 1, defined as

$$\theta_t = \rho\theta_{t-1} + \eta_t, \tag{3.8}$$

subject to the sampling model

$$y_t = \theta_t + e_t, \tag{3.9}$$

where $\{y_t\}_{t\geq 1}$ are observations with $y_0 = 0$, $\{\theta_t\}_{t\geq 1}$ are studied variables, $\{e_t\}_{t\geq 1}$ is a sequence of independent $N(0, D_t^2)$ distributed random variables, and $\{\eta_t\}_{t\geq 1}$ is a sequence of iid $N(0, \sigma^2)$ distributed random variables. The model (3.8) is called the linking model. We assume that the variables $\{\theta_t\}$ are unobserved directly, but we can observe their direct estimators $\{y_t\}$, under the model (3.9). This situation can be found in small area estimation where the true population means are not observed. The true variables are estimated by the corresponding sample mean obtained from a sample. In general, the variance of sampling errors can be estimated by many methods, such as empirical variance modeling from Fay and Herriot [9]. Hence, many researchers studied the sampling errors under the assumption of known sampling variances, such as Angkunsit and Suntornchost [2], Chatterjee and Lahiri [4], and Lahiri and Suntornchost [12]. In this study, we assume that the sampling variance $D_t^2$ is known and bounded, and $e_t$, and $\eta_t$ are independent. The model (3.8)-(3.9) can be represented as

$$y_t = \sum_{i=1}^{t} \rho^{t-i}\eta_i + e_t, \tag{3.10}$$



or,

$$y_t = \rho y_{t-1} + a_t, \tag{3.11}$$

where $a_t = \eta_t + e_t - \rho e_{t-1}$ are dependent random variables with $\mathbb{E}(a_t) = 0$, $\mathrm{Var}(a_t) = (1+\rho^2)D_t^2 + \sigma^2$, $\mathrm{Cov}(a_t, a_{t+1}) = -\rho D_t^2$, and $\mathrm{Cov}(a_t, a_{t+h}) = 0$ for all $h \geq 2$. Notice that $\mathbb{E}(y_t) = 0$, $\mathrm{Var}(y_t) = \frac{\sigma^2}{1-\rho^2} + D_t^2$, and $\mathrm{Cov}(y_t, y_{t+h}) = \mathrm{Cov}(\theta_t, \theta_{t+h}) = \rho^h \frac{\sigma^2}{1-\rho^2}$ for each $h > 0$.

Under the models (3.8) and (3.9), the stationary property refers to the stationary property on time series process $\theta_t$, not the observed variables $y_t$, in spite of the fact that the sampling variance $D_t^2$ depends on time $t$. Hence, we will consider the estimator of autoregressive coefficient $\rho$ and the test of the null hypothesis that $|\rho| = 1$. In particular, the test of hypothesis

$$\begin{aligned} &H_0 : |\rho| = 1 \quad \text{or nonstationary AR(1)},\\ \text{against} \\ &H_1 : |\rho| < 1 \quad \text{or stationary AR(1)}, \end{aligned} \tag{3.12}$$

under models (3.8)-(3.9).

## 3.2 Parameter Estimation of Autoregressive Coefficient for the Model with Sampling Errors

In pervious section, we have introduced the first order autoregressive model subject to sampling errors and hypothesis for the stationary test. To analyze the stationary test, an estimator of the autoregressive coefficient is necessary. In this section, we study parameter estimation of the autoregressive coefficient in the model. We consider the effect of sampling errors on the estimator of the autoregressive coefficient, and investigate a good estimator to reduce sampling errors. Furthermore, we investigate the asymptotic distribution of the obtained estimator when $\rho = 1$.



### 3.2.1 Likelihood Function

Consider the model

$$\begin{aligned} y_t &= \theta_t + e_t,\\ \theta_t &= \rho\theta_{t-1} + \eta_t, \end{aligned} \tag{3.13}$$

where $y_0 = \theta_0 = 0$, $e_t$'s are independent $N(0, D_t^2)$ distributed random variables, $\eta_t$'s are iid $N(0, \sigma^2)$ distributed random variables, $e_t$ and $\eta_t$ are independent. We assume $T$ variables $y_1, y_2, \ldots, y_T$ are avaliable, but variables $\theta_1, \theta_2, \ldots, \theta_T$ are unknown. The variables $\{y_t\}$ are direct estimators of $\{\theta_t\}$ with known and bounded sampling variances $\{D_t^2\}$. One of many methods to investigate estimators of unknown parameters $\rho$ and $\sigma^2$ is the maximum likelihood method. The log-likelihood function for the model (3.13) is

$$l\left(\rho, \sigma^2 \,|\, \mathbf{y}, \mathbf{D}\right) = -\frac{1}{2}T\log 2\pi - \frac{1}{2}\log|\mathbf{\Gamma}| - \frac{1}{2}\mathbf{y}'\mathbf{\Gamma}^{-1}\mathbf{y}, \qquad (3.14)$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_T)'$ and $\mathbf{\Gamma}$ is the $T \times T$ covariance matrix of $\mathbf{y}$, and $\mathbf{D}$ is covariance matrix of $\mathbf{e} = (e_1, e_2, \ldots, e_T)$. Notice that $\mathbf{\Gamma} = \sigma^2\mathbf{P} + \mathbf{D}$ where

$$\mathbf{P} = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{T-1} \\ \rho & 1 & \rho & \ldots & \rho^{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \ldots & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{D} = \begin{bmatrix} D_1^2 & 0 & \ldots & 0 \\ 0 & D_2^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & D_T^2 \end{bmatrix}.$$

The maximum likelihood estimators of $\rho$ and $\sigma^2$ can be found by solving the system of equations from setting the partial derivatives of $l(\rho, \sigma^2 \,|\, \mathbf{y})$ with respect $\rho$ and $\sigma^2$ equal to zero. In particular,

$$0 = \frac{\partial l}{\partial \rho} = -\frac{1}{2}\operatorname{Tr}\left(\sigma^2\mathbf{\Gamma}^{-1}\frac{\partial \boldsymbol{P}}{\partial \rho}\right) + \frac{\sigma^2}{2}\mathbf{y}'\mathbf{\Gamma}^{-1}\frac{\partial \boldsymbol{P}}{\partial \rho}\mathbf{\Gamma}^{-1}\mathbf{y}, \qquad (3.15)$$

and

$$0 = \frac{\partial l}{\partial \sigma^2} = -\frac{1}{2}\operatorname{Tr}\left(\mathbf{\Gamma}^{-1}\mathbf{P}\right) + \frac{1}{2}\mathbf{y}'\mathbf{\Gamma}^{-1}\mathbf{P}\mathbf{\Gamma}^{-1}\mathbf{y}, \qquad (3.16)$$

where $\operatorname{Tr}(\mathbf{A})$ denotes the trace of a matrix $\mathbf{A}$. It is usually difficult to obtain closed-form solutions from (3.15)-(3.16). However, the solutions can be obtained by any available numerical methods.



### 3.2.2 Effect of Sampling Errors to the Estimator

If sampling variances $(D_t^2)_{t\geq 1}$ in model (3.13) are all zero, the observed variable $\{y_t\}$ are the same as the unknown variables $\{\theta_t\}$. Hence, estimator of $\rho$, denoted by $\rho_\theta$, is

$$\rho_\theta = \frac{\sum_{t=2}^{T}\theta_t\theta_{t-1}}{\sum_{t=2}^{T}\theta_{t-1}^2}. \qquad (3.17)$$

However, sampling variances $\{D_t^2\}_{t\geq 1}$ are nonzero. Therefore, ignoring the sampling variances will cause errors in parameter estimation and the stationary test. In this section, we are interested in examining errors in the true estimators $\rho_\theta$ caused by the naive estimators $\rho_y$, where

$$\rho_y = \frac{\sum\limits_{t=2}^{T} y_t y_{t-1}}{\sum\limits_{t=2}^{T} y_{t-1}^2}, \tag{3.18}$$

in the existence of the sampling errors. In addition, we consider the effect of naive estimator by the expectation of $\rho_y - \rho_\theta$. However, the closed forms of the mean and the variance are difficult to obtain. Therefore, the approximation by Taylor series expansion will be applied in our study. To simplify notations, we define the following terms.

$$\begin{aligned}
U_{1,\theta} &= \sum_{t=2}^{T} \theta_t \theta_{t-1}, & U_{1,y} &= \sum_{t=2}^{T} y_t y_{t-1}, \\
U_{2,\theta} &= \sum_{t=2}^{T} \theta_{t-1}^2, & U_{2,y} &= \sum_{t=2}^{T} y_{t-1}^2, \\
U_{3,\theta} &= \rho_\theta = \frac{U_{1,\theta}}{U_{2,\theta}}, & U_{3,y} &= \rho_y = \frac{U_{1,y}}{U_{2,y}}, \\
U_{4,\theta} &= \frac{1}{T-2} \sum_{t=2}^{T} (\theta_t - \rho_\theta \theta_{t-1})^2, & U_{4,y} &= \frac{1}{T-2} \sum_{t=2}^{T} (y_t - \rho_y y_{t-1})^2.
\end{aligned} \tag{3.19}$$

Notice that $\rho_\theta$ and $\rho_y$ are functions of $(U_{1,\theta}, U_{2,\theta})$ and $(U_{1,y}, U_{2,y})$, respectively. We first study some properties of the estimator $\rho_\theta$, and the effect of naive estimator $\rho_y$ later.

### 1) Some properties of the Estimator of Autoregressive Coefficient on AR(1)

In this section, we will consider some properties of the estimator $\rho_\theta$ of $\rho$ when $\theta_t$'s are observed. Under the AR(1) model, Mann and Wald [16] have shown that $\rho_\theta$, obtained by the maximum likelihood method, is a consistent estimator for $|\rho| < 1$. In addition, Rubin [19] has shown that $\rho_\theta$ is a consistent estimator for the case $|\rho| \geq 1$. However, we prove consistency $\rho_\theta$ for the case $|\rho| \leq 1$ alternatively.

**Lemma 3.3.** *The asymptotic properties of $U_{2,\theta}$ defined in (3.19) are as follows.*

*(1) If $|\rho| < 1$, then*

$$\begin{aligned} \lim_{T\to\infty} \frac{\mathbb{E}(U_{2,\theta})}{T} &= \frac{\sigma^2}{1-\rho^2}, \\ \lim_{T\to\infty} \frac{\mathrm{Var}(U_{2,\theta})}{T} &= \frac{2\sigma^4(1+\rho^2)}{(1-\rho^2)^3}. \end{aligned} \tag{3.20}$$

*(2) If $|\rho| = 1$, then*

$$\begin{aligned} \lim_{T\to\infty} \frac{\mathbb{E}(U_{2,\theta})}{T^2} &= \frac{\sigma^2}{2}, \\ \lim_{T\to\infty} \frac{\mathrm{Var}(U_{2,\theta})}{T^4} &= \frac{\sigma^4}{3}. \end{aligned} \tag{3.21}$$

*(3) For any positive integer $k \geq 1$,*

$$\mathbb{E}(U_{2,\theta}^{-k}) = \begin{cases} O(T^{-k}), & \text{if } |\rho| < 1, \\ O(T^{-2k}), & \text{if } |\rho| = 1. \end{cases} \tag{3.22}$$

*Proof.* (1) Notice that $(\theta_t)_{t\geq 1}$ satisfies AR(1) model with $\theta_0 = 0$. Then $\theta_t = \sum_{i=1}^{t} \rho^{t-i}\eta_i$ where $(\eta_i)_{t\geq 1}$ are independent normal distribution with mean 0 and variance $\sigma^2$. Consider the case $|\rho| < 1$, we can transform $U_{2,\theta}$ as

$$\begin{aligned} U_{2,\theta} &= \sum_{t=2}^{T} \theta_{t-1}^2 \\ &= \sum_{t=1}^{T-1} \theta_t^2 \\ &= \sum_{t=1}^{T-1} \left( \sum_{i=1}^{t} \rho^{t-i}\eta_i \right)^2 \\ &= \eta_1^2 + (\eta_2 + \rho\eta_1)^2 + (\eta_3 + \rho\eta_2 + \rho^2\eta_1)^2 + \cdots + (\eta_{T-1} + \cdots + \rho^{T-2}\eta_1)^2 \\ &= \big((1 + \rho^2 + \cdots + \rho^{2(T-2)})\eta_1^2 + (1 + \rho^2 + \cdots + \rho^{2(T-3)})\eta_2^2 + \cdots + \eta_{T-1}^2\big) \end{aligned}$$

$$+\,2\left(\rho\eta_2\eta_1+\sum_{1\le i<j\le 3}\rho^{2(3)-(i+j)}\eta_i\eta_j+\cdots+\sum_{1\le i<j\le T-1}\rho^{2(T-1)-(i+j)}\eta_i\eta_j\right)$$

$$=\big((1+\rho^2+\cdots+\rho^{2(T-2)})\eta_1^2+(1+\rho^2+\cdots+\rho^{2(T-3)})\eta_2^2+\cdots+\eta_{T-1}^2\big)$$

$$+\,2\sum_{j=2}^{T-1}\sum_{i=1}^{j-1}\rho^{j-i}(1+\rho^2+\cdots+\rho^{2(T-j-1)})\eta_i\eta_j$$

$$=\sum_{i=1}^{T-1}\frac{1-\rho^{2(T-i)}}{1-\rho^2}\eta_i^2+2\sum_{j=2}^{T-1}\sum_{i=1}^{j-1}\rho^{j-i}\frac{1-\rho^{2(T-j)}}{1-\rho^2}\eta_i\eta_j. \tag{3.23}$$

Let $\xi_1$ and $\xi_2$ be the first and the second term of (3.23), respectively. Since $\mathbb{E}(\eta_i^2)=\sigma^2$ and $\mathrm{Var}(\eta_i^2)=2\sigma^4$, the asymptotic mean of $\frac{\xi_1}{T}$ can be computed as follows.

$$\begin{aligned}\lim_{T\to\infty}\frac{\mathbb{E}(\xi_1)}{T}&=\lim_{T\to\infty}\frac{1}{T(1-\rho^2)}\sum_{i=1}^{T-1}\left(1-\rho^{2(T-i)}\right)\mathbb{E}\left(\eta_i^2\right)\\&=\lim_{T\to\infty}\frac{\sigma^2}{T(1-\rho^2)}\sum_{j=1}^{T-1}(1-\rho^{2j})\\&=\lim_{T\to\infty}\frac{\sigma^2}{T(1-\rho^2)}\left((T-1)-\rho^2\frac{1-\rho^{2(T-1)}}{1-\rho^2}\right)\\&=\frac{\sigma^2}{1-\rho^2}.\end{aligned}\tag{3.24}$$

Since $(\eta_i)_{i\ge1}$ are independent and normally distributed with variance $\sigma^2$,

$$\begin{aligned}\lim_{T\to\infty}\frac{\mathrm{Var}(\xi_1)}{T}&=\lim_{T\to\infty}\frac{2\sigma^4}{T(1-\rho^2)^2}\sum_{i=1}^{T-1}(1-\rho^{2i})^2\\&=\lim_{T\to\infty}\frac{2\sigma^4}{T(1-\rho^2)^2}\sum_{i=1}^{T-1}(1-2\rho^{2i}+\rho^{4i})\\&=\lim_{T\to\infty}\frac{2\sigma^4}{T(1-\rho^2)^2}\left((T-1)-2\rho^2\frac{1-\rho^{2(T-1)}}{1-\rho^2}+\rho^4\frac{1-\rho^{4(T-1)}}{1-\rho^4}\right)\\&=\frac{2\sigma^4}{(1-\rho^2)^2}.\end{aligned}\tag{3.25}$$

For the mean and variane of $\xi_2$, since $(\eta_i)_{i\ge1}$ are independent random variables with zero mean,

$$\mathbb{E}(\xi_2)=0. \tag{3.26}$$

To find the variance of $\xi_2$, note that $\mathrm{Cov}(\eta_i\eta_j, \eta_i\eta_k) = 0$ for any $i, j, k$ such that $i \neq j \neq k$. Then

$$\begin{aligned}
\mathrm{Var}(\xi_2) &= \frac{4\sigma^4}{(1-\rho^2)^2}\sum_{j=2}^{T-1}\sum_{i=1}^{j-1}\rho^{2(j-i)}(1-\rho^{2(T-j)})^2 \\
&= \frac{4\sigma^4}{(1-\rho^2)^2}\sum_{j=2}^{T-1}(1-\rho^{2(T-j)})^2(\rho^2+\rho^4+\ldots+\rho^{2(j-1)}) \\
&= \frac{4\sigma^4}{(1-\rho^2)^2}\sum_{j=2}^{T-1}(1-\rho^{2(T-j)})^2\frac{\rho^2(1-\rho^{2(j-1)})}{1-\rho^2} \\
&= \frac{4\rho^2\sigma^4}{(1-\rho^2)^3}\sum_{j=2}^{T-1}(1-2\rho^{2(T-j)}+\rho^{4(T-j)})(1-\rho^{2(j-1)}) \\
&= \frac{4\rho^2\sigma^4}{(1-\rho^2)^3}\sum_{j=2}^{T-1}(1-2\rho^{2(T-j)}+\rho^{4(T-j)}-\rho^{2(j-1)}+2\rho^{2(T-1)}-\rho^{4T-2j-2}) \\
&= \frac{4\rho^2\sigma^4}{(1-\rho^2)^3}\Bigg((T-1)-2\frac{\rho^2(1-\rho^{2(T-2)})}{1-\rho^2}+\frac{\rho^4(1-\rho^{4(T-2)})}{1-\rho^4}-\frac{\rho^2(1-\rho^{2(T-2)})}{1-\rho^2} \\
&\quad +2(T-2)\rho^{2(T-1)}-\rho^{2(T-1)}\frac{\rho^2(1-\rho^{2(T-2)})}{1-\rho^2}\Bigg).
\end{aligned}$$

Therefore,

$$\lim_{T\to\infty}\frac{\mathrm{Var}(\xi_2)}{T} = \frac{4\rho^2\sigma^4}{(1-\rho^2)^3}. \tag{3.27}$$

Notice that $\mathbb{E}(U_{2,\theta}) = \mathbb{E}(\xi_1) + \mathbb{E}(\xi_2)$. From (3.23), (3.24), and (3.26),

$$\begin{aligned}
\lim_{T\to\infty}\frac{\mathbb{E}(U_{2,\theta})}{T} &= \lim_{T\to\infty}\frac{\mathbb{E}(\xi_1)}{T} + \lim_{T\to\infty}\frac{\mathbb{E}(\xi_2)}{T} \\
&= \frac{\sigma^2}{1-\rho^2}.
\end{aligned}$$

Since $\mathrm{Cov}(\eta_i^2, \eta_i\eta_j) = \mathbb{E}(\eta_i^3\eta_j) = 0$ for any $i, j$ such that $i \neq j$, we have $\mathrm{Var}(U_{2,\theta}) = \mathrm{Var}(\xi_1) + \mathrm{Var}(\xi_2)$. From (3.23), (3.25), and (3.27),

$$\lim_{T\to\infty}\frac{\mathrm{Var}(U_{2,\theta})}{T} = \lim_{T\to\infty}\frac{\mathrm{Var}(\xi_1)}{T} + \lim_{T\to\infty}\frac{\mathrm{Var}(\xi_2)}{T}$$

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

Moreover, since the limits of $\frac{\mathbb{E}(U_{2,\theta})}{T}$ and $\frac{\mathrm{Var}(U_{2,\theta})}{T}$ both exist, for any $k \in \mathbb{R}$,

$$\lim_{T\to\infty} \frac{\mathbb{E}^k(U_{2,\theta})}{T^k} = \frac{\sigma^{2k}}{(1-\rho^2)^k}, \tag{3.28}$$

and
$$\lim_{T\to\infty} \frac{\mathrm{Var}^k(U_{2,\theta})}{T^k} = \frac{2^k\sigma^{4k}(1+\rho^2)^k}{(1-\rho^2)^{3k}}. \tag{3.29}$$

Therefore, we can see that

$$\mathbb{E}^k(U_{2,\theta}) = O(T^k), \tag{3.30}$$

and
$$\mathrm{Var}^k(U_{2,\theta}) = O(T^k). \tag{3.31}$$

(2) Consider the case $|\rho| = 1$. WLOG, we will consider only the case $\rho = 1$. Notice that

$$\begin{aligned} U_{2,\theta} =& \eta_1^2 + (\eta_2+\rho\eta_1)^2 + (\eta_3+\rho\eta_2+\rho^2\eta_1)^2 + \ldots + (\eta_{T-1}+\ldots+\rho^{T-2}\eta_1)^2 \\ =& \big((T-1)\eta_1^2 + (T-2)\eta_2^2 + \ldots + \eta_{T-1}^2\big) \\ &+ 2\Big((T-2)\eta_1\eta_2 + (T-3)(\eta_1+\eta_2)\eta_3 + \ldots (\eta_1+\eta_2+\ldots\eta_{T-2})\eta_{T-1}\Big). \end{aligned}$$



Then,

$$\begin{aligned} \lim_{T\to\infty} \frac{\mathbb{E}(U_{2,\theta})}{T^2} &= \lim_{T\to\infty} \frac{1}{T^2}\sum_{i=1}^{T-1} i\sigma^2 \\ &= \lim_{T\to\infty} \frac{\sigma^2}{T^2}\frac{(T-1)T}{2} \\ &= \frac{\sigma^2}{2}. \end{aligned} \tag{3.32}$$

Notice that $\mathrm{Var}(\eta_i^2) = 2\sigma^4$ and $\mathrm{Var}(\eta_i\eta_j) = \sigma^4$ for $i \neq j$. By using the fact that

$\mathrm{Cov}(\eta_i^2, \eta_i\eta_j) = 0$ for $i \neq j$ and $\mathrm{Cov}(\eta_i\eta_j, \eta_i\eta_k) = 0$ for any $i \neq j \neq k$, we have

$$\begin{aligned}
\lim_{T\to\infty} \frac{\mathrm{Var}(U_{2,\theta})}{T^4} &= \lim_{T\to\infty} \left( \frac{2\sigma^4}{T^4} \sum_{i=1}^{T-1} i^2 + \frac{4\sigma^4}{T^4} \sum_{i=1}^{T-2} i^2 (T-1-i) \right) \\
&= \lim_{T\to\infty} \frac{\sigma^4}{T^4} \left( \frac{2(T-1)T(2T-1)}{6} + 4(T-1)\frac{(T-2)(T-1)(2T-3)}{6} \right. \\
&\quad \left. - \frac{4(T-2)^2(T-1)^2}{4} \right) \\
&= \lim_{T\to\infty} \frac{\sigma^4}{T^4} \left( \frac{(T-1)T(2T-1)}{3} + \frac{(T-2)(T-1)^2 T}{3} \right) \\
&= \lim_{T\to\infty} \frac{\sigma^4}{3T^4} (T-1)(T)(T^2-T+1) \\
&= \frac{\sigma^4}{3}. \qquad (3.33)
\end{aligned}$$

Hence, we can conclude that, for the case $|\rho| = 1$, the limit of $T^{-2}\,\mathbb{E}(U_{2,\theta})$ and $T^{-4}\,\mathrm{Var}(U_{2,\theta})$ both exist and converge to some positive values. Moreover, for any $k \in \mathbb{R}$,

$$\lim_{T\to\infty} \frac{\mathbb{E}^k(U_{2,\theta})}{T^k} = \frac{\sigma^{2k}}{2^k}, \qquad (3.34)$$

and

$$\lim_{T\to\infty} \frac{\mathrm{Var}^k(U_{2,\theta})}{T^k} = \frac{\sigma^{4k}}{3^k}. \qquad (3.35)$$

Therefore, we can see that

$$\mathbb{E}^k(U_{2,\theta}) = O(T^{2k}), \qquad (3.36)$$

and

$$\mathrm{Var}^k(U_{2,\theta}) = O(T^{4k}). \qquad (3.37)$$

(3) Let $k$ be any positive integer. To compute the mean of $U_{2,\theta}^{-k}$, we will apply the second order Taylor approximation to the function $f_k(x) = x^{-k}$ about $\mu = \mathbb{E}(X)$. From (2.16), we have

$$\mathbb{E}(X^{-k}) \approx \frac{1}{\mathbb{E}^k(X)} + \frac{k(k+1)}{2} \frac{1}{\mathbb{E}^{k+2}(X)} \mathrm{Var}(X). \qquad (3.38)$$

Substituting $X$ by $U_{2,\theta}$, we consider two cases: $|\rho| < 1$, and $|\rho| = 1$.

If $|\rho| < 1$, from (3.30)-(3.31), we have

$$\begin{aligned}\mathbb{E}(U_{2,\theta}^{-k}) &\approx \mathbb{E}^{-k}(U_{2,\theta}) + \frac{k(k+1)}{2}\,\mathbb{E}^{-(k+2)}(U_{2,\theta})\,\mathrm{Var}(U_{2,\theta}) \\ &= O(T^{-k}) + O(T^{-(k+2)})O(T) \\ &= O(T^{-k}). \end{aligned} \tag{3.39}$$

If $|\rho| = 1$, from (3.36)-(3.37), we have

$$\begin{aligned}\mathbb{E}(U_{2,\theta}^{-k}) &\approx \mathbb{E}^{-k}(U_{2,\theta}) + \frac{k(k+1)}{2}\,\mathbb{E}^{-(k+2)}(U_{2,\theta})\,\mathrm{Var}(U_{2,\theta}) \\ &= O(T^{-2k}) + O(T^{-2(k+2)})O(T^{4}) \\ &= O(T^{-2k}). \end{aligned} \tag{3.40}$$

□

**Lemma 3.4.** *Let*

$$\rho_\theta = \frac{U_{1,\theta}}{U_{2,\theta}} = \frac{\sum_{t=2}^{T} \theta_t \theta_{t-1}}{\sum_{t=2}^{T} \theta_{t-1}^2}.$$

*Then, for* $|\rho| \leq 1$,

*(1)* $\mathbb{E}(|\rho_\theta - \rho|)$ *converges to zero.*

*(2)* $\rho_\theta$ *is a consistent estimator of* $\rho$.

*(3)* $\mathbb{E}(|\rho_\theta|)$ *is bounded by* $|\rho| + C_1 T^{-1}$ *for some constant* $C_1$.

*(4)* $\mathbb{E}(\rho_\theta^2)$ *is bounded with some constant* $C_2$.

*Proof.* (1) Notice that

$$\begin{aligned}
\rho_\theta - \rho &= \frac{\sum\limits_{t=2}^{T} \theta_t \theta_{t-1}}{\sum\limits_{t=2}^{T} \theta_{t-1}^2} - \rho \\
&= \frac{\sum\limits_{t=2}^{T} (\rho\theta_{t-1} + \eta_t)\theta_{t-1}}{\sum\limits_{t=2}^{T} \theta_{t-1}^2} - \rho \\
&= \frac{\sum\limits_{t=2}^{T} \eta_t \theta_{t-1}}{\sum\limits_{t=2}^{T} \theta_{t-1}^2}. \qquad (3.41)
\end{aligned}$$

For the numerator, since $\eta_t$ and $\theta_{t-1}$ are independent of all $t$, we have

$$\mathbb{E}\left(\sum_{t=2}^{T} \eta_t \theta_{t-1}\right) = 0, \qquad (3.42)$$

and, consequently,

$$\mathrm{Var}\left(\sum_{t=2}^{T} \eta_t \theta_{t-1}\right) = \mathbb{E}\left(\sum_{t=2}^{T} \eta_t \theta_{t-1}\right)^2, \qquad (3.43)$$

for all $\rho$.

To compute $\mathbb{E}(|\rho_\theta - \rho|)$, we will consider two cases: $|\rho| < 1$ and $|\rho| = 1$. Consider the case $|\rho| < 1$. To compute variance of the numerator term, we using the fact that $\mathrm{Cov}(\eta_i\eta_j, \eta_i\eta_k) = 0$ for all $i, j, k$ such that $i \neq j \neq k$, and $\mathrm{Var}(\eta_i\eta_j) = \sigma^4$. Hence,

$$\begin{aligned}
\mathbb{E}\left(\sum_{t=2}^{T} \eta_t \theta_{t-1}\right)^2 &= \mathrm{Var}\left(\sum_{t=2}^{T} \eta_t \theta_{t-1}\right) \\
&= \mathrm{Var}\left(\sum_{t=2}^{T} \eta_t \left(\sum_{i=1}^{t-1} \rho^{t-i} \eta_i\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&= \operatorname{Var}\left(\sum_{t=2}^{T}\sum_{i=1}^{t-1}\rho^{t-i}\eta_t\eta_i\right)\\
&= \sigma^4\sum_{t=2}^{T}\sum_{i=1}^{t-1}\rho^{2(t-i)}\\
&= \rho^2\sigma^4\sum_{t=2}^{T}\frac{1-\rho^{2(t-1)}}{1-\rho^2}\\
&= \frac{\rho^2\sigma^4}{1-\rho^2}\left(\sum_{t=2}^{T}(1-\rho^{2(t-1)})\right)\\
&= \frac{\rho^2\sigma^4}{1-\rho^2}\left((T-1)-\frac{\rho^4(1-\rho^{2(T-2)})}{1-\rho^2}\right)\\
&= \frac{\rho^2\sigma^4}{1-\rho^2}(T-1)-\frac{\rho^6\sigma^4}{(1-\rho^2)^2}(1-\rho^{2(T-2)}). \qquad (3.44)
\end{aligned}
$$

By using Holder inequality, (3.22) and (3.44), we have

$$
\begin{aligned}
\mathbb{E}(|\rho_\theta-\rho|) &= \mathbb{E}\left(\left|\frac{\sum_{t=2}^{T}\eta_t\theta_{t-1}}{\sum_{t=2}^{T}\theta_{t-1}^2}\right|\right)\\
&\le \sqrt{\mathbb{E}\left(\sum_{t=2}^{T}\eta_t\theta_{t-1}\right)^2}\sqrt{\mathbb{E}\left(\sum_{t=2}^{T}\theta_{t-1}^2\right)^{-2}}\\
&= \sqrt{\left(\frac{\rho^2\sigma^4}{1-\rho^2}(T-1)-\frac{\rho^6\sigma^4}{(1-\rho^2)^2}(1-\rho^{2(T-2)})\right)}\sqrt{O(T^{-2})}\\
&= O(T^{\frac{1}{2}})O(T^{-1})\\
&= O(T^{-\frac{1}{2}}). \qquad (3.45)
\end{aligned}
$$

In the same way, for the case $|\rho| = 1$, variance of the numerator term can be compute as:

$$
\begin{aligned}
\operatorname{Var}\left(\sum_{t=2}^{T}\eta_t\theta_{t-1}\right) &= \operatorname{Var}\left(\sum_{t=2}^{T}\eta_t\left(\sum_{i=1}^{t-1}\eta_i\right)\right)\\
&= \operatorname{Var}\left(\sum_{t=2}^{T}\sum_{i=1}^{t-1}\eta_t\eta_i\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \mathrm{Var}\left(\sum_{1\le i<j\le T} \eta_j\eta_i\right) \\
&= \sum_{1\le i<j\le T} \sigma^4 \\
&= \frac{(T)(T-1)\sigma^4}{2}. \qquad (3.46)
\end{aligned}
$$

Consequently, by using Holder inequality, (3.22) and (3.46), we have

$$
\begin{aligned}
\mathbb{E}(|\rho_\theta - \rho|) &= \mathbb{E}\left(\left|\frac{\sum_{t=2}^{T} \eta_t\theta_{t-1}}{\sum_{t=2}^{T} \theta_{t-1}^2}\right|\right) \\
&\le \sqrt{\mathbb{E}\left(\sum_{t=2}^{T} \eta_t\theta_{t-1}\right)^2}\sqrt{\mathbb{E}\left(\sum_{t=2}^{T} \theta_{t-1}^2\right)^{-2}} \\
&= \sqrt{\frac{(T)(T-1)\sigma^4}{2}}\sqrt{O(T^{-4})} \\
&= O(T)O(T^{-2}) \\
&= O(T^{-1}). \qquad (3.47)
\end{aligned}
$$

Hence, we can conclude that $\mathbb{E}(|\rho_\theta - \rho|)$ converges to zero. In particular, $\mathbb{E}(|\rho_\theta - \rho|)$ is of order $T^{-\frac{1}{2}}$ for $|\rho| < 1$, and order $T^{-1}$ for $|\rho| = 1$.

(2) Since $\mathbb{E}(|\rho_\theta - \rho|)$ converges to zero, $\rho_\theta - \rho$ converges to zero in probability. Hence, $\rho_\theta$ converges in probability to $\rho$.

(3) Consider the case $|\rho| < 1$. By using the fact that $|x| \le |x-y| + |y|$ and (3.45), we have

$$
\mathbb{E}(|\rho_\theta|) \le \mathbb{E}(|\rho_\theta - \rho|) + |\rho| = |\rho| + O(T^{-\frac{1}{2}}) \le |\rho| + \frac{C_1}{\sqrt{T}}, \qquad (3.48)
$$

for some positive constant $C_1$. Similarly for the case $|\rho| = 1$, from (3.47), we have

$$
\mathbb{E}(|\rho_\theta|) \le \mathbb{E}(|\rho_\theta - \rho|) + |\rho| = |\rho| + O(T^{-1}) \le |\rho| + \frac{C_2}{T^1}, \qquad (3.49)
$$

for some positive constant $C_2$. Hence, $\mathbb{E}(|\rho_\theta|)$ is bounded.

(4) Notice that

$$\mathbb{E}(\rho_\theta^2) = \mathrm{Var}(\rho_\theta) + \mathbb{E}^2(\rho_\theta) = \mathrm{Var}(\rho_\theta - \rho) + \mathbb{E}^2(\rho_\theta). \tag{3.50}$$

First, we will consider $\mathrm{Var}(\rho_\theta - \rho)$. We apply the first order Taylor series approximation on $f(X, Y) = \dfrac{X}{Y}$, where $X = \sum_{t=2}^{T} \eta_t \theta_{t-1}$ and $Y = U_{2,\theta} = \sum_{t=2}^{T} \theta_{t-1}^2$. From (2.14), we have

$$\begin{aligned}
\mathrm{Var}(\rho_\theta - \rho) &= \mathrm{Var}\left(\frac{X}{Y}\right) \\
&\approx \mathbb{E}^{-2}(Y)\,\mathrm{Var}(X) + \mathbb{E}^2(X)\,\mathbb{E}^{-4}(Y)\,\mathrm{Var}(Y) - 2\,\mathbb{E}(X)\,\mathbb{E}^{-3}(Y)\,\mathrm{Cov}(X, Y) \\
&= \mathbb{E}^{-2}(Y)\,\mathrm{Var}(X),
\end{aligned} \tag{3.51}$$

where the last two terms vanish since $\mathbb{E}(X) = \mathbb{E}\left(\sum_{t=2}^{T} \eta_t \theta_{t-1}\right) = 0$.

Consider the case $|\rho| < 1$. From (3.30), (3.44) and (3.51), we have

$$\begin{aligned}
\mathrm{Var}(\rho_\theta - \rho) &\le \mathbb{E}^{-2}\left(U_{2,\theta}\right) \mathrm{Var}\left(\sum_{t=2}^{T} \eta_t \theta_{t-1}\right) \\
&= O(T^{-2}) \left(\frac{\sigma^4}{1-\rho^2}(T-1) - \frac{\sigma^4 \rho^4}{(1-\rho^2)^2}(1 - \rho^{2(T-1)})\right) \\
&= O(T^{-1}).
\end{aligned} \tag{3.52}$$

By using the fact that $\mathbb{E}^2(\rho_\theta) \le \mathbb{E}^2(|\rho_\theta|)$, (3.50), (3.48), and (3.52), we have

$$\begin{aligned}
\mathbb{E}(\rho_\theta^2) &= \mathrm{Var}(\rho_\theta) + \mathbb{E}^2(\rho_\theta) \\
&= \mathrm{Var}(\rho_\theta - \rho) + \mathbb{E}^2(\rho_\theta) \\
&\le O(T^{-1}) + \left(|\rho| + O(T^{-\frac{1}{2}})\right)^2 \\
&= O(T^{-1}) + \rho^2 + |\rho| O(T^{-\frac{1}{2}}) + O(T^{-1}) \\
&= \rho^2 + O(T^{-\frac{1}{2}}) \\
&= O(1).
\end{aligned} \tag{3.53}$$

Consider the case $|\rho| = 1$. From (3.36), (3.46) and (3.51), we have

$$\begin{aligned}\mathrm{Var}(\rho_\theta - \rho) &\le \mathbb{E}^{-2}\left(U_{2,\theta}\right)\mathrm{Var}\left(\sum_{t=2}^{T}\eta_t\theta_{t-1}\right)\\ &= O(T^{-4})\left(\frac{(T)(T-1)\sigma^4}{2}\right)\\ &= O(T^{-2}). \end{aligned} \tag{3.54}$$

Similarly, from (3.50), (3.49), and (3.54), we have

$$\begin{aligned}\mathbb{E}(\rho_\theta^2) &= \mathrm{Var}(\rho_\theta) + \mathbb{E}^2(\rho_\theta)\\ &= \mathrm{Var}(\rho_\theta - \rho) + \mathbb{E}^2(\rho_\theta)\\ &\le O(T^{-2}) + \left(|\rho| + O(T^{-1})\right)^2\\ &= O(T^{-2}) + \rho^2 + |\rho|O(T^{-1}) + O(T^{-2})\\ &= \rho^2 + O(T^{-1})\\ &= O(1). \end{aligned} \tag{3.55}$$

From (3.53) and (3.55), we can conclude that $\mathbb{E}(\rho_\theta^2)$ is bounded. □



**2) The Effect of the Naive Estimator**

From Lemma 3.4, the estimator $\rho_\theta$ is a consistent estimator of $\rho$. However, under the model (3.13), the studied variables $\theta_t$'s are unobserved. To consider the effect of sampling errors, we first show some conditional expectation of the naive statistics $U_{1,y}$ and $U_{2,y}$ given $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_T)'$.

**Proposition 3.5.** *The conditional expectation, variation, and covariation of* $U_{1,y}$ *and* $U_{2,y}$ *given* $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_T)'$ *are respectively defined as:*

*(1)* $\mathbb{E}\left(U_{1,y} \,|\, \boldsymbol{\theta}\right) = U_{1,\theta}$.

*(2)* $\mathbb{E}\left(U_{2,y} \,|\, \boldsymbol{\theta}\right) = U_{2,\theta} + \sum_{t=2}^{T} D_{t-1}^2$.

*(3)* $\mathrm{Var}\left(U_{1,y} \,|\, \boldsymbol{\theta}\right) = \sum_{t=2}^{T}\left(\theta_t^2 D_{t-1}^2 + \theta_{t-1}^2 D_t^2 + D_t^2 D_{t-1}^2 + 2\theta_t\theta_{t-2}D_{t-1}^2\right)$.

*(4)* $\mathrm{Var}\left(U_{2,y}\,|\,\boldsymbol{\theta}\right) = \sum_{t=2}^{T}\left(2D_{t-1}^{4} + 4\theta_{t-1}^{2}D_{t-1}^{2}\right).$

*(5)* $\mathrm{Cov}\left(U_{1,y}, U_{2,y}\,|\,\boldsymbol{\theta}\right) = 2\sum_{t=2}^{T}\left(\theta_t\theta_{t-1} + \theta_{t-1}\theta_{t-2}\right)D_{t-1}^{2}.$

*Proof.* (1) The conditional expectation of $U_{1,y}$ given $\boldsymbol{\theta}$ can be computed as

$$\begin{aligned}
\mathbb{E}\left(U_{1,y}\,|\,\boldsymbol{\theta}\right) &= \mathbb{E}\left(\sum_{t=2}^{T} y_t y_{t-1}\,\middle|\,\boldsymbol{\theta}\right) \\
&= \sum_{t=2}^{T}\mathbb{E}\left((\theta_t + e_t)(\theta_{t-1} + e_{t-1})\,|\,\boldsymbol{\theta}\right) \\
&= \sum_{t=2}^{T}\mathbb{E}\left(\theta_t\theta_{t-1} + \theta_t e_{t-1} + \theta_{t-1}e_t + e_t e_{t-1}|\,\boldsymbol{\theta}\right) \\
&= \sum_{t=2}^{T}\theta_t\theta_{t-1} + \theta_t\,\mathbb{E}\left(e_{t-1}\right) + \theta_{t-1}\,\mathbb{E}\left(e_t\right) + \mathbb{E}\left(e_t e_{t-1}\right) \\
&= \sum_{t=2}^{T}\theta_t\theta_{t-1},
\end{aligned}$$

where we use the fact that $e_t's$ are independent random variables with zero mean to obtain the last equation.

(2) The conditional expectation of $U_{2,y}$ given $\boldsymbol{\theta}$ can be computed as

$$\begin{aligned}
\mathbb{E}\left(U_{2,y}\,|\,\boldsymbol{\theta}\right) &= \mathbb{E}\left(\sum_{t=2}^{T} y_{t-1}^{2}\,\middle|\,\boldsymbol{\theta}\right) \\
&= \sum_{t=2}^{T}\mathbb{E}\left((\theta_{t-1} + e_{t-1})^2\,\middle|\,\boldsymbol{\theta}\right) \\
&= \sum_{t=2}^{T}\mathbb{E}\left(\theta_{t-1}^{2} + e_{t-1}^{2} + 2\theta_{t-1}e_{t-1}\,\middle|\,\boldsymbol{\theta}\right) \\
&= \sum_{t=2}^{T}\theta_{t-1}^{2} + \mathbb{E}\left(e_{t-1}^{2}\right) + 2\theta_{t-1}\,\mathbb{E}\left(e_{t-1}\right) \\
&= \sum_{t=2}^{T}\theta_{t-1}^{2} + D_{t-1}^{2},
\end{aligned}$$

where we use the fact that $e_t's$ are independent random variables with zero mean and variance $D_t^2$ to obtain the last equation.

(3) The conditional variance of $U_{1,y}$ given $\boldsymbol{\theta}$ can be computed as

$$\begin{aligned}
\mathrm{Var}\left(U_{1,y} \,|\, \boldsymbol{\theta}\right) &= \mathrm{Var}\left(\sum_{t=2}^{T} y_t y_{t-1} \,\middle|\, \boldsymbol{\theta}\right) \\
&= \mathrm{Var}\left(\sum_{t=2}^{T} \theta_t\theta_{t-1} + \theta_t e_{t-1} + \theta_{t-1}e_t + e_t e_{t-1} \,\middle|\, \boldsymbol{\theta}\right) \\
&= \mathrm{Var}\left(\sum_{t=2}^{T} \theta_t e_{t-1} \,\middle|\, \boldsymbol{\theta}\right) + \mathrm{Var}\left(\sum_{t=2}^{T} \theta_{t-1} e_t \,\middle|\, \boldsymbol{\theta}\right) + \mathrm{Var}\left(\sum_{t=2}^{T} e_t e_{t-1} \,\middle|\, \boldsymbol{\theta}\right) \\
&\quad + 2\,\mathrm{Cov}\left(\sum_{t=2}^{T} \theta_t e_{t-1}, \sum_{t=2}^{T} \theta_{t-1} e_t \,\middle|\, \boldsymbol{\theta}\right) + 2\,\mathrm{Cov}\left(\sum_{t=2}^{T} \theta_{t-1} e_t, \sum_{t=2}^{T} e_t e_{t-1} \,\middle|\, \boldsymbol{\theta}\right) \\
&\quad + 2\,\mathrm{Cov}\left(\sum_{t=2}^{T} e_t e_{t-1}, \sum_{t=2}^{T} \theta_t e_{t-1} \,\middle|\, \boldsymbol{\theta}\right) \\
&= \sum_{t=2}^{T} \theta_t^2 D_{t-1}^2 + \sum_{t=2}^{T} \theta_{t-1}^2 D_t^2 + \sum_{t=2}^{T} D_t^2 D_{t-1}^2 + 2\sum_{t=2}^{T} \mathrm{Cov}\left(\theta_t e_{t-1}, \theta_{t-2} e_{t-1} \middle| \boldsymbol{\theta}\right) \\
&\quad + 2\sum_{t=2}^{T} \mathrm{Cov}\left(\theta_{t-1} e_t, e_{t+1}e_t + e_t e_{t-1} \middle| \boldsymbol{\theta}\right) \\
&\quad + 2\sum_{t=2}^{T} \mathrm{Cov}\left(\theta_t e_{t-1}, e_t e_{t-1} + e_{t-1} e_{t-2} \middle| \boldsymbol{\theta}\right) \\
&= \sum_{t=2}^{T} \theta_t^2 D_{t-1}^2 + \sum_{t=2}^{T} \theta_{t-1}^2 D_t^2 + \sum_{t=2}^{T} D_t^2 D_{t-1}^2 + 2\sum_{t=2}^{T} \theta_t \theta_{t-2} D_{t-1}^2,
\end{aligned}$$

where $\theta_0 = 0$ and we use the fact that $\mathrm{Cov}(e_t, e_t e_{t+h}) = \mathbb{E}(e_{t+h} e_t^2) = 0$ for $h > 0$.

(4) The conditional variance of $U_{2,y}$ given $\boldsymbol{\theta}$ can be computed as

$$\begin{aligned}
\mathrm{Var}\left(U_{2,y} \,|\, \boldsymbol{\theta}\right) &= \mathrm{Var}\left(\sum_{t=2}^{T} y_{t-1}^2 \,\middle|\, \boldsymbol{\theta}\right) \\
&= \mathrm{Var}\left(\sum_{t=2}^{T} (\theta_{t-1}^2 + e_{t-1}^2 + 2\theta_{t-1}e_{t-1}) \,\middle|\, \boldsymbol{\theta}\right) \\
&= \sum_{t=2}^{T} \mathrm{Var}\left(e_{t-1}^2 + 2\theta_{t-1}e_{t-1} \,\middle|\, \boldsymbol{\theta}\right) \\
&= \sum_{t=2}^{T} \mathbb{E}\left((e_{t-1}^2 + 2\theta_{t-1}e_{t-1})^2 \,\middle|\, \boldsymbol{\theta}\right) - \mathbb{E}\left(e_{t-1}^2 + 2\theta_{t-1}e_{t-1} \,\middle|\, \boldsymbol{\theta}\right)^2
\end{aligned}$$

$$= \sum_{t=2}^{T} \mathbb{E}\left(e_{t-1}^4 + 4\theta_{t-1}e_{t-1}^3 + 4\theta_{t-1}^2 e_{t-1}^2 \,\middle|\, \boldsymbol{\theta}\right) - D_{t-1}^4$$
$$= \sum_{t=2}^{T} 2D_{t-1}^4 + 4\theta_{t-1}^2 D_{t-1}^2.$$

(5) The conditional covariance of $U_{1,y}$ and $U_{2,y}$ given $\boldsymbol{\theta}$ can be computed as

$$\begin{aligned}
\mathrm{Cov}\left(U_{1,y}, U_{2,y} \,|\, \boldsymbol{\theta}\right) &= \mathrm{Cov}\left(\sum_{t=2}^{T} y_t y_{t-1}, \sum_{t=2}^{T} y_{t-1}^2 \,\middle|\, \boldsymbol{\theta}\right) \\
&= \mathrm{Cov}\left(\sum_{t=2}^{T} \theta_t e_{t-1} + \theta_{t-1}e_t + e_t e_{t-1}, \sum_{s=2}^{T} 2\theta_{s-1}e_{s-1} + e_{s-1}^2 \,\middle|\, \boldsymbol{\theta}\right) \\
&= \sum_{t=2}^{T} \mathrm{Cov}\left(2\theta_{t-1}e_{t-1} + e_{t-1}^2, \sum_{s=2}^{T} \theta_s e_{s-1} + \theta_{s-1}e_s + e_s e_{s-1} \,\middle|\, \boldsymbol{\theta}\right) \\
&= \sum_{t=2}^{T} \mathrm{Cov}\left(2\theta_{t-1}e_{t-1} + e_{t-1}^2, \theta_t e_{t-1} + \theta_{t-2}e_{t-1} + e_t e_{t-1} + e_{t-1}e_{t-2} \,\middle|\, \boldsymbol{\theta}\right) \\
&= \sum_{t=2}^{T} 2\theta_t\theta_{t-1}D_{t-1}^2 + 2\theta_{t-1}\theta_{t-2}D_{t-1}^2 \\
&= 2\sum_{t=2}^{T} \left(\theta_t\theta_{t-1} + \theta_{t-1}\theta_{t-2}\right) D_{t-1}^2.
\end{aligned}$$

□

Next, we will apply the results to investigate the effect of naive estimator $\rho_y$ by using Taylor's expansion.

**Proposition 3.6.** *Let $\rho_y$ be the naive estimator of $\rho$ defined in (3.19). Under the condition that the sampling variance $\inf_t D_t^2 > 0$, $\mathbb{E}(\rho_y - \rho)$ does not converge to zero when $0 < |\rho| < 1$. In addition, the estimator $\rho_y$ is underestimated when $\rho > 0$, and overestimated when $\rho < 0$.*

*Proof.* Let $f(x, y) = \dfrac{x}{y}$. From (2.13), the first order Taylor approximations of $\mathbb{E}\left(f(X,Y)|\boldsymbol{\theta}\right)$ about $(\mu_X, \mu_Y) = \left(\mathbb{E}\left(X|\boldsymbol{\theta}\right), \mathbb{E}\left(Y|\boldsymbol{\theta}\right)\right)$ is

$$\mathbb{E}\left(\frac{X}{Y}\middle|\boldsymbol{\theta}\right) \approx \frac{\mathbb{E}\left(X|\boldsymbol{\theta}\right)}{\mathbb{E}\left(Y|\boldsymbol{\theta}\right)}, \tag{3.56}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_n)'$, and $X$ and $Y$ are random variables. Substituting $(X, Y) = (U_{1,y}, U_{2,y})$, from Proposition 3.5 (1)-(2), we have

$$\mathbb{E}\left(\rho_y|\boldsymbol{\theta}\right) \approx \frac{\mathbb{E}\left(\sum\limits_{t=2}^{T} y_t y_{t-1} \,\middle|\, \boldsymbol{\theta}\right)}{\mathbb{E}\left(\sum\limits_{t=2}^{T} y_{t-1}^2 \,\middle|\, \boldsymbol{\theta}\right)} = \frac{\sum\limits_{t=2}^{T} \theta_t \theta_{t-1}}{\sum\limits_{t=2}^{T} \theta_{t-1}^2 + D_{t-1}^2}.$$

It follows that

$$\begin{aligned}
\mathbb{E}(\rho_y - \rho) &= \mathbb{E}(\mathbb{E}(\rho_y - \rho) \,|\, \boldsymbol{\theta}) \\
&\approx \mathbb{E}\left(\frac{\sum\limits_{t=2}^{T} \theta_t \theta_{t-1}}{\sum\limits_{t=2}^{T} \theta_{t-1}^2 + D_{t-1}^2} - \rho\right) \\
&= \mathbb{E}\left(\frac{\sum\limits_{t=2}^{T} (\rho\theta_{t-1} + \eta_t)\theta_{t-1}}{\sum\limits_{t=2}^{T} \theta_{t-1}^2 + D_{t-1}^2} - \rho\right) \\
&= \mathbb{E}\left(\frac{\sum\limits_{t=2}^{T} \eta_t \theta_{t-1} - \rho \sum\limits_{t=2}^{T} D_{t-1}^2}{\sum\limits_{t=2}^{T} \theta_{t-1}^2 + \sum\limits_{t=2}^{T} D_{t-1}^2}\right) \\
&= \mathbb{E}\left(\frac{\sum\limits_{t=2}^{T} \eta_t \theta_{t-1}}{\sum\limits_{t=2}^{T} \theta_{t-1}^2 + \sum\limits_{t=2}^{T} D_{t-1}^2}\right) - \rho \frac{1}{T} \sum_{t=2}^{T} D_{t-1}^2 \, \mathbb{E}\left(\frac{1}{\frac{1}{T}\sum\limits_{t=2}^{T} \theta_{t-1}^2 + \frac{1}{T}\sum\limits_{t=2}^{T} D_{t-1}^2}\right).
\end{aligned} \tag{3.57}$$



For the first term of (3.57), from (3.41) and Lemma 3.4 (1), we have

$$\left|\mathbb{E}\left(\frac{\sum\limits_{t=2}^{T} \eta_t \theta_{t-1}}{\sum\limits_{t=2}^{T} \theta_{t-1}^2 + \sum\limits_{t=2}^{T} D_{t-1}^2}\right)\right| \leq \mathbb{E}\left(\left|\frac{\sum\limits_{t=2}^{T} \eta_t \theta_{t-1}}{\sum\limits_{t=2}^{T} \theta_{t-1}^2}\right|\right) = \mathbb{E}(|\rho_\theta - \rho|), \tag{3.58}$$

converges to zero as $T$ goes to infinity.

For the second term of (3.57), from Lemma 3.3 (1),

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} \mathbb{E}\left(\theta_{t-1}^2\right) = \frac{\sigma^2}{1-\rho^2}.$$

Let $D_M^2 = \sup_t D_t^2$ and $D_m^2 = \inf_t D_t^2$. By applying Jensen's inequality and $f(x) = \frac{1}{x}$ is convex function, and using the fact that $D_M^2 \geq D_t^2$ for all $t$, we have

$$\begin{aligned}
\lim_{T\to\infty} \mathbb{E}\left(\frac{1}{\frac{1}{T}\sum_{t=2}^{T}\theta_{t-1}^2 + \frac{1}{T}\sum_{t=2}^{T} D_{t-1}^2}\right) &\geq \lim_{T\to\infty} \frac{1}{\mathbb{E}(\frac{1}{T}\sum_{t=2}^{T}\theta_{t-1}^2) + \frac{1}{T}\sum_{t=2}^{T} D_{t-1}^2} \\
&= \frac{1}{\lim_{T\to\infty}\mathbb{E}(\frac{1}{T}\sum_{t=2}^{T}\theta_{t-1}^2) + \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T} D_{t-1}^2} \\
&= \frac{1}{\frac{\sigma^2}{1-\rho^2} + \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T} D_{t-1}^2} \\
&\geq \frac{1}{\frac{\sigma^2}{1-\rho^2} + \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T} D_{M}^2} \\
&= \frac{1}{\frac{\sigma^2}{1-\rho^2} + D_M^2}. \qquad (3.59)
\end{aligned}$$

Hence, $\mathbb{E}\left(\frac{1}{\frac{1}{T}\sum_{t=2}^{T}\theta_{t-1}^2 + \frac{1}{T}\sum_{t=2}^{T} D_{t-1}^2}\right)$ is bounded below by the positive value $\left(\frac{\sigma^2}{1-\rho^2} + D_M^2\right)^{-1}$ as $T$ goes to infinity.



To show that the estimator $\rho_y$ is underestimated when $0 < \rho < 1$, from (3.57)-(3.59) and $D_m^2 \leq D_t^2$ for all $t$, we have

$$\begin{aligned}
\lim_{T\to\infty} \mathbb{E}(\rho_y - \rho) &= \lim_{T\to\infty} \sum_{t=2}^{T} D_{t-1}^2 \rho \frac{1}{T}\, \mathbb{E}\left(\frac{1}{\frac{1}{T}\sum_{t=2}^{T}\theta_{t-1}^2 + \frac{1}{T} D_t^2 \sum_{t=2}^{T} D_{t-1}^2}\right) \\
&\leq -\rho D_m^2 \lim_{T\to\infty} \mathbb{E}\left(\frac{1}{\frac{1}{T}\sum_{t=2}^{T}\theta_{t-1}^2 + \frac{1}{T}\sum_{t=2}^{T} D_{t-1}^2}\right)
\end{aligned}$$

$$\leq -\rho D_m^2 \left( \frac{1}{\frac{\sigma^2}{1-\rho^2} + D_{M}^2} \right)$$
$$< 0. \tag{3.60}$$

Hence, $\rho_y$ is underestimate of $\rho$ for $0 < \rho < 1$.

Similarly for $-1 < \rho < 0$, from (3.58)-(3.59), we have

$$\lim_{T\to\infty} \mathbb{E}(\rho_y - \rho) = \lim_{T\to\infty} \sum_{t=2}^{T} D_{t-1}^2 \rho \frac{1}{T} \mathbb{E} \left( \frac{1}{\frac{1}{T}\sum\limits_{t=2}^{T} \theta_{t-1}^2 + \frac{1}{T} D_t^2 \sum\limits_{t=2}^{T} D_{t-1}^2} \right)$$
$$\geq -\rho D_m^2 \lim_{T\to\infty} \mathbb{E} \left( \frac{1}{\frac{1}{T}\sum\limits_{t=2}^{T} \theta_{t-1}^2 + \frac{1}{T}\sum\limits_{t=2}^{T} D_{t-1}^2} \right)$$
$$\geq -\rho D_m^2 \left( \frac{1}{\frac{\sigma^2}{1-\rho^2} + D_M^2} \right)$$
$$> 0. \tag{3.61}$$

Hence, $\rho_y$ is overestimate of $\rho$ for $-1 < \rho < 0$. □

### 3.2.3 Estimator for Autoregressive Coefficient



In the pervious section, we have shown the effect of sampling errors on the naive estimator. When we ignore sampling errors, there is a bias term causes the naive estimator to be underestimate when $0 < \rho < 1$. In this section, we propose a simple adjustment of the parameter estimation of $\rho$ alternative to the naive estimator $\rho_y$. This adjustment can reduce bias caused from sampling errors, and depends only on the observed variables $y_t$'s.

**Corollary 3.7.** *Let* $U_{1,\theta} = \sum\limits_{t=2}^{T} \theta_t \theta_{t-1}$ *be defined in (3.19). Then*

*(1)* $\mathbb{E}(U_{1,\theta}) = O(T)$ *and* $\mathrm{Var}(U_{1,\theta}) = O(T)$ *if* $|\rho| < 1$,

*(2)* $\mathbb{E}(U_{1,\theta}) = O(T^2)$ *and* $\mathrm{Var}(U_{1,\theta}) = O(T^4)$ *if* $|\rho| = 1$,

*Proof.* Notice that

$$U_{1,\theta} = \sum_{t=2}^{T} \theta_t \theta_{t-1} = \sum_{t=2}^{T} (\rho\theta_{t-1} + \eta_t)\theta_{t-1} = \rho U_{2,\theta} + \sum_{t=2}^{T} \eta_t \theta_{t-1}.$$

Hence, since $\mathbb{E}\left(\sum_{t=2}^{T} \eta_t\theta_{t-1}\right) = 0$, $\mathbb{E}(U_{1,\theta}) = \rho\,\mathbb{E}(U_{2,\theta})$ for all cases of $\rho$. This implies that the orders of $U_{1,\theta}$ and $U_{2,\theta}$ are the same.

Consider the case $|\rho| < 1$. By using the fact that $\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y)$, we have

$$\frac{\operatorname{Var}(U_{1,\theta})}{T} = \rho^2 \frac{\operatorname{Var}(U_{2,\theta})}{T} + \frac{\operatorname{Var}\left(\sum_{t=2}^{T} \eta_t\theta_{t-1}\right)}{T} + 2\rho \frac{\operatorname{Cov}\left(U_{2,\theta}, \sum_{t=2}^{T} \eta_t\theta_{t-1}\right)}{T}. \tag{3.62}$$

From Lemma 3.3 and (3.44), the sum of the first and the second terms converges to constant $C = \frac{2\sigma^4\rho^2(1+\rho^2)}{(1-\rho^2)^3} + \frac{\sigma^4}{1-\rho^2}$. Moreover, by using Chebyshev's inequality,

$$\lim_{T\to\infty} \left| \frac{\operatorname{Cov}\left(U_{2,\theta}, \sum_{t=2}^{T} \eta_t\theta_{t-1}\right)}{T} \right| \le \lim_{T\to\infty} \sqrt{\frac{\operatorname{Var}(U_{2,\theta})}{T} \frac{\operatorname{Var}\left(\sum_{t=2}^{T} \eta_t\theta_{t-1}\right)}{T}} = \frac{\sigma^4\sqrt{2(1+\rho^2)}}{(1-\rho^2)^2}. \tag{3.63}$$

Therefore, we can conclude that $\frac{\operatorname{Var}(U_{1,\theta})}{T}$ lies in $[C - 2\rho D, C + 2\rho D]$ where $C = \frac{2\sigma^4\rho^2(1+\rho^2)}{(1-\rho^2)^3} + \frac{\sigma^4}{1-\rho^2}$ and $D = \frac{\sigma^4\sqrt{2(1+\rho^2)}}{(1-\rho^2)^2}$, and implies that $\operatorname{Var}(U_{1,\theta}) = O(T)$. Similarly, for the cases $|\rho| = 1$, we have

$$\frac{\operatorname{Var}(U_{1,\theta})}{T^4} = \rho^2 \frac{\operatorname{Var}(U_{2,\theta})}{T^4} + \frac{\operatorname{Var}\left(\sum_{t=2}^{T} \eta_t\theta_{t-1}\right)}{T^4} + 2\rho \frac{\operatorname{Cov}\left(U_{2,\theta}, \sum_{t=2}^{T} \eta_t\theta_{t-1}\right)}{T^4}. \tag{3.64}$$

From (3.46), the second term of (3.64) converges to zero. By using Chebyshev's inequality, (3.21), and (3.46) on the third term, we have

$$\left| \frac{\mathrm{Cov}\left(U_{2,\theta}, \sum_{t=2}^{T} \eta_t \theta_{t-1}\right)}{T^4} \right| \leq \frac{1}{T} \sqrt{\frac{\mathrm{Var}\,(U_{2,\theta})}{T^4} \frac{\mathrm{Var}\left(\sum_{t=2}^{T} \eta_t \theta_{t-1}\right)}{2}}, \tag{3.65}$$

and this term converges to zero as $T$ goes to infinity.

Hence, by (3.21) and (3.64), $\frac{\mathrm{Var}(U_{1,\theta})}{T^4}$ converges to $\frac{\sigma^4 \rho^2}{3}$. Consequently, $\mathrm{Var}(U_{1,\theta}) = O(T^4)$. □

**Lemma 3.8.** *Let $U_{1,\theta}$ and $U_{2,\theta}$ be defined as in Lemma 3.3 and Corollary 3.7, respectively. For any positive integers $m$ and $n$ such that $n > m$,*

$$\mathbb{E}(U_{1,\theta}^m U_{2,\theta}^{-n}) \leq \begin{cases} O(T^{m-n}) & \text{if } |\rho| < 1, \\ O(T^{2(m-n)}) & \text{if } |\rho| = 1. \end{cases} \tag{3.66}$$

*Proof.* Let $f(X,Y) = X^m Y^{-n}$ where $X, Y$ are random variables. Let $\mu = (\mu_X, \mu_Y) = (\mathbb{E}(X), \mathbb{E}(Y))$. The second order Taylor approximation of $f$ about $\mu$ is

$$\begin{aligned} f(X,Y) \approx\ & \mu_X^m \mu_Y^{-n} + m\mu_X^{m-1}\mu_Y^{-n}(X-\mu_X) - n\mu_X^m \mu_Y^{-n-1}(Y-\mu_Y) \\ & + \frac{m(m-1)}{2}\mu_X^{m-2}\mu_Y^{-n}(X-\mu_X)^2 - 2mn\mu_X^{m-1}\mu_Y^{-n-1}(X-\mu_X)(Y-\mu_Y) \\ & + \frac{n(n+1)}{2}\mu_X^m \mu_Y^{-n-2}(Y-\mu_Y)^2 \end{aligned} \tag{3.67}$$



By substituting $X = U_{1,\theta}$ and $Y = U_{2,\theta}$ into (3.67) and taking expectation, we have

$$\begin{aligned} |\,\mathbb{E}(U_{1,\theta}^m U_{2,\theta}^{-n})| \approx \big|\, & \mathbb{E}^m(U_{1,\theta})\,\mathbb{E}^{-n}(U_{2,\theta}) + \frac{m(m-1)}{2}\,\mathbb{E}^{m-2}(U_{1,\theta})\,\mathbb{E}^{-n}(U_{2,\theta})\,\mathrm{Var}(U_{1,\theta}) \\ & + \frac{n(n+1)}{2}\,\mathbb{E}^m(U_{1,\theta})\,\mathbb{E}^{-n-2}(U_{2,\theta})\,\mathrm{Var}(U_{2,\theta}) \\ & - 2mn\,\mathbb{E}^{m-1}(U_{1,\theta})\,\mathbb{E}^{-n-1}(U_{2,\theta})\,\mathrm{Cov}(U_{1,\theta}, U_{2,\theta})\big| \end{aligned}$$

$$\leq |\,\mathbb{E}^m(U_{1,\theta})|\,\mathbb{E}^{-n}(U_{2,\theta}) + \frac{m(m-1)}{2}|\,\mathbb{E}^{m-2}(U_{1,\theta})|\,\mathbb{E}^{-n}(U_{2,\theta})\,\mathrm{Var}(U_{1,\theta})$$
$$+ \frac{n(n+1)}{2}|\,\mathbb{E}^m(U_{1,\theta})|\,\mathbb{E}^{-n-2}(U_{2,\theta})\,\mathrm{Var}(U_{2,\theta})$$
$$+ 2mn|\,\mathbb{E}^{m-1}(U_{1,\theta})|\,\mathbb{E}^{-n-1}(U_{2,\theta})\sqrt{\mathrm{Var}(U_{1,\theta})}\sqrt{\mathrm{Var}(U_{2,\theta})}, \tag{3.68}$$

where we use the triangle inequality and the Chebyshev's inequality to obtain the last inequality.

When $|\rho| < 1$, from Corollary 3.7 (1) and (3.30)-(3.31), we have

$$\begin{aligned}|\,\mathbb{E}(U_{1,\theta}^m U_{2,\theta}^{-n})| &\leq O(T)^{m-n} + \frac{m(m-1)}{2}O(T)^{m-2}O(T)^{-n}O(T)\\ &\quad + \frac{n(n+1)}{2}O(T)^m O(T)^{-n-2}O(T)\\ &\quad + 2mnO(T)^{m-1}O(T)^{-n-1}O(T)^{\frac{1}{2}}O(T)^{\frac{1}{2}}\\ &= O(T^{m-n}) + O(T^{m-n-1}) + O(T^{m-n-1}) + O(T^{m-n-1})\\ &= O(T^{m-n}).\end{aligned} \tag{3.69}$$

Similarly for the case $|\rho| = 1$, from Corollary 3.7 (2) and (3.36)-(3.37), we have

$$\begin{aligned}|\,\mathbb{E}(U_{1,\theta}^m U_{2,\theta}^{-n})| &\leq O(T)^{2m-2n} + \frac{m(m-1)}{2}O(T)^{2m-4}O(T)^{-2n}O(T^4)\\ &\quad + \frac{n(n-1)}{2}O(T)^{2m}O(T)^{-2n-4}O(T^4)\\ &\quad + 2mnO(T)^{2m-2}O(T)^{-2n-2}O(T^4)^{\frac{1}{2}}O(T^4)^{\frac{1}{2}}\\ &= O(T^{2(m-n)}) + O(T^{2(m-n)}) + O(T^{2(m-n)}) + O(T^{2(m-n)})\\ &= O(T^{2(m-n)}).\end{aligned} \tag{3.70}$$

□

**Corollary 3.9.** *Let $U_{1,\theta}$ and $U_{2,\theta}$ be defined as in Lemma 3.3 and Corollary 3.7, respectively. For any positive integers $m$ and $n$ such that $n > m$,*

$$\mathbb{E}\left(|U_{1,\theta}|^m\, U_{2,\theta}^{-n}\right) \leq \begin{cases} O(T^{m-n}) & \text{if } |\rho| < 1,\\ O(T^{2(m-n)}) & \text{if } |\rho| = 1.\end{cases} \tag{3.71}$$

*Proof.* From Hölder's inequality, we have

$$\mathbb{E}\left(|U_{1,\theta}|^m U_{2,\theta}^{-n}\right) \leq \mathbb{E}^{\frac{1}{2}}\left(U_{1,\theta}^{2m}U_{2,\theta}^{-2n}\right). \tag{3.72}$$

Consider the case $|\rho| < 1$. From Lemma 3.8, we have

$$\mathbb{E}\left(|U_{1,\theta}|^m U_{2,\theta}^{-n}\right) \leq \sqrt{O(T^{2m-2n})} = O(T^{m-n}). \tag{3.73}$$

Similarly, for the case $|\rho| < 1$. From Lemma 3.8, we have

$$\mathbb{E}\left(|U_{1,\theta}|^m U_{2,\theta}^{-n}\right) \leq \sqrt{O(T^{2(2m-2n)})} = O(T^{2(m-n)}). \tag{3.74}$$

□

**Theorem 3.10.** *(1) The statistic*

$$\hat{U}_{1,\theta} = \sum_{t=2}^{T} y_t y_{t-1}$$

*is an unbiased estimator of* $U_{1,\theta}$.

*(2) The statistic*

$$\hat{U}_{2,\theta} = \sum_{t=2}^{T}(y_{t-1}^2 - D_{t-1}^2)$$

*is an unbiased estimator of* $U_{2,\theta}$.

*(3) The statistic*

$$\hat{U}_{3,\theta} = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T}(y_{t-1}^2 - D_{t-1}^2)}$$

*is an asymptotically unbiased estimator of* $U_{3,\theta} = \rho_\theta$. *The statistic* $\hat{U}_{3,\theta}$ *can be denoted as* $\hat{\rho}_\theta$.

*Proof.* (1) From Proposition 3.5 (1), we can show that

$$\mathbb{E}(\hat{U}_{1,\theta} - U_{1,\theta}) = \mathbb{E}\left(\mathbb{E}\left(\hat{U}_{1,\theta} - U_{1,\theta}\middle|\boldsymbol{\theta}\right)\right) = \mathbb{E}(U_{1,\theta} - U_{1,\theta}) = 0. \tag{3.75}$$

(2) From Proposition 3.5 (2), we can show that

$$\mathbb{E}(\hat{U}_{2,\theta}-U_{2,\theta}) = \mathbb{E}\left(\mathbb{E}\left(\hat{U}_{2,\theta} - U_{2,\theta}\middle|\boldsymbol{\theta}\right)\right) = \mathbb{E}\left(U_{2,\theta} + \sum_{t=2}^{T} D_{t-1}^2 - \sum_{t=2}^{T} D_{t-1}^2 - U_{2,\theta}\right) = 0. \tag{3.76}$$

(3) To show that $\hat{U}_{3,\theta}$ is an asymptotically unbiased estimator of $U_{3,\theta}$, we will apply the second-order Taylor approximations (2.16) of $\mathbb{E}\left(\hat{U}_{3,\theta}\middle|\boldsymbol{\theta}\right) = \mathbb{E}\left(f(\hat{U}_{1,\theta}, \hat{U}_{2,\theta})\middle|\boldsymbol{\theta}\right)$ around $\boldsymbol{\mu} = (\mathbb{E}(\hat{U}_{1,\theta}|\boldsymbol{\theta}), \mathbb{E}(\hat{U}_{2,\theta}|\boldsymbol{\theta}))$, when $f(x,y) = \dfrac{x}{y}$ as follows.

$$\begin{aligned}
\mathbb{E}\left(f(\hat{U}_{1,\theta}, \hat{U}_{2,\theta})\middle|\boldsymbol{\theta}\right) &\approx f(\boldsymbol{\mu}) + \frac{1}{2}\left(f''_{xx}(\mu)\operatorname{Var}(\hat{U}_{1,\theta}|\boldsymbol{\theta}) + f''_{yy}(\mu)\operatorname{Var}(\hat{U}_{2,\theta}|\boldsymbol{\theta})\right) \\
&\quad + f''_{xy}(\mu)\operatorname{Cov}(\hat{U}_{1,\theta}, \hat{U}_{2,\theta}|\boldsymbol{\theta}) \\
&= f(\boldsymbol{\mu}) + \left(\frac{\mathbb{E}(\hat{U}_{1,\theta}|\boldsymbol{\theta})}{\mathbb{E}^2(\hat{U}_{2,\theta}|\boldsymbol{\theta})}\operatorname{Var}(\hat{U}_{2,\theta}|\boldsymbol{\theta}) - \frac{1}{\mathbb{E}^2(\hat{U}_{2,\theta}|\boldsymbol{\theta})}\operatorname{Cov}(\hat{U}_{1,\theta}, \hat{U}_{2,\theta}|\boldsymbol{\theta})\right).
\end{aligned}$$

Since $(\hat{U}_{1,\theta}, \hat{U}_{2,\theta}) = (U_{1,y}, U_{2,y} - \sum_{t=2}^{T} D_{t-1}^2)$, the conditional covariance matrix given $\boldsymbol{\theta}$ of $(\hat{U}_{1,\theta}, \hat{U}_{2,\theta})$ are the same as $(U_{1,y}, U_{2,y})$. That is,



$$\operatorname{Var}\left(\hat{U}_{1,\theta}\middle|\boldsymbol{\theta}\right) = \sum_{t=2}^{T}\left(\theta_t^2 D_{t-1}^2 + \theta_{t-1}^2 D_t^2 + D_t^2 D_{t-1}^2 + 2\theta_t\theta_{t-2}D_{t-1}^2\right), \tag{3.77}$$

$$\operatorname{Var}\left(\hat{U}_{2,\theta}\middle|\boldsymbol{\theta}\right) = \sum_{t=2}^{T}\left(2D_{t-1}^4 + 4\theta_{t-1}^2 D_{t-1}^2\right), \tag{3.78}$$

$$\operatorname{Cov}\left(\hat{U}_{1,\theta}, \hat{U}_{2,\theta}\middle|\boldsymbol{\theta}\right) = 2\sum_{t=2}^{T}\left(\theta_t\theta_{t-1} + \theta_{t-1}\theta_{t-2}\right) D_{t-1}^2. \tag{3.79}$$

Hence, by Proposition 3.5 (1)-(2) and (3.78)-(3.79), we have

$$\mathbb{E}\left(\hat{U}_{3,\theta}\middle|\boldsymbol{\theta}\right) \approx \frac{U_{1,\theta}}{U_{2,\theta}} - \frac{\operatorname{Cov}(\hat{U}_{1,\theta}, \hat{U}_{2,\theta}|\boldsymbol{\theta})}{U_{2,\theta}^2} + \frac{U_{1,\theta}}{U_{2,\theta}}\frac{\operatorname{Var}(\hat{U}_{2,\theta}|\boldsymbol{\theta})}{U_{2,\theta}^2}$$

$$= U_{3,\theta} - (2\sum_{t=2}^{T}(\theta_t\theta_{t-1} + \theta_{t-1}\theta_{t-2})\, D_{t-1}^2)(U_{2,\theta})^{-2}$$
$$+ U_{1,\theta}U_{2,\theta}^{-3}(2\sum_{t=2}^{T} D_{t-1}^4 + 4\sum_{t=2}^{T}\theta_{t-1}^2 D_{t-1}^2). \tag{3.80}$$

Therefore, by using the fact that $|2xy| \leq x^2 + y^2$ for all $x, y$ and $U_{2,\theta} > 0$, we have

$$\begin{aligned}
\left|\mathbb{E}\left(\hat{U}_{3,\theta} - U_{3,\theta}\middle|\boldsymbol{\theta}\right)\right| &= \Big|-(2\sum_{t=2}^{T}(\theta_t\theta_{t-1} + \theta_{t-1}\theta_{t-2})\, D_{t-1}^2)(U_{2,\theta})^{-2} \\
&\quad + U_{1,\theta}U_{2,\theta}^{-3}(2\sum_{t=2}^{T} D_{t-1}^4 + 4\sum_{t=2}^{T}\theta_{t-1}^2 D_{t-1}^2)\Big| \\
&\leq \Big|-(2\sum_{t=2}^{T}(\theta_t\theta_{t-1} + \theta_{t-1}\theta_{t-2})\, D_{t-1}^2)(U_{2,\theta})^{-2}\Big| \\
&\quad + \Big|U_{1,\theta}U_{2,\theta}^{-3}(2\sum_{t=2}^{T} D_{t-1}^4 + 4\sum_{t=2}^{T}\theta_{t-1}^2 D_{t-1}^2)\Big| \\
&\leq (\sum_{t=2}^{T}\left(\theta_t^2 + 2\theta_{t-1}^2 + \theta_{t-2}^2\right) D_{t-1}^2) \\
&\quad + |U_{1,\theta}|U_{2,\theta}^{-3}(2\sum_{t=2}^{T} D_{t-1}^4 + 4\sum_{t=2}^{T}\theta_{t-1}^2 D_{t-1}^2) \\
&\leq D_M^2 U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_t^2 + D_M^2 U_{2,\theta}^{-1} + D_M^2 U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_{t-2}^2 \\
&\quad + 2D_M^2(T-1)|U_{1,\theta}|U_{2,\theta}^{-3} + 4D_M^2|U_{1,\theta}|U_{2,\theta}^{-2}.
\end{aligned} \tag{3.81}$$



To compute the expectation of $\left(U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_t^2 + U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_{t-2}^2\right)$, we notice that

$$\begin{aligned}
U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_t^2 &= U_{2,\theta}^{-1} + (\theta_T^2 - \theta_1^2)U_{2,\theta}^{-2} \\
&\leq U_{2,\theta}^{-1} + \theta_T^2 U_{2,\theta}^{-2} \\
&= U_{2,\theta}^{-1} + (\rho\theta_{T-1} + \eta_T)^2 U_{2,\theta}^{-2} \\
&\leq U_{2,\theta}^{-1} + 2(\rho^2\theta_{T-1}^2 + \eta_T^2)U_{2,\theta}^{-2}
\end{aligned}$$

$$
\begin{aligned}
&\leq U_{2,\theta}^{-1} + 2\rho^2 U_{2,\theta} U_{2,\theta}^{-2} + 2\eta_T^2 \left( \sum_{t=2}^{T} \theta_{t-1}^2 \right)^{-2} \\
&= (2\rho^2 + 1) U_{2,\theta}^{-1} + 2\eta_T^2 U_{2,\theta}^{-2}. \qquad (3.82)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
U_{2,\theta}^{-2} \sum_{t=2}^{T} \theta_{t-2}^2 &= U_{2,\theta}^{-1} - \theta_{T-1}^2 \left( \sum_{t=2}^{T} \theta_{t-1}^2 \right)^{-2} \\
&\leq \left| U_{2,\theta}^{-1} \right| + \left| \theta_{T-1}^2 U_{2,\theta}^{-2} \right| \\
&\leq U_{2,\theta}^{-1} + U_{2,\theta} U_{2,\theta}^{-2} \\
&\leq 2U_{2,\theta}^{-1}. \qquad (3.83)
\end{aligned}
$$

Since $U_{2,\theta}$ is a function of $(\theta_1, \theta_2, \ldots, \theta_{T-1})$, $U_{2,\theta}$ and $\eta_T$ are independent. Let $M_1 = \max\{\mathbb{E}(U_{2,\theta}^{-1}), \mathbb{E}(U_{2,\theta}^{-2})\}$. From (3.82), we have

$$
\begin{aligned}
\mathbb{E}\left( U_{2,\theta}^{-2} \sum_{t=2}^{T} \theta_t^2 \right) &\leq (2\rho^2 + 1)\, \mathbb{E}(U_{2,\theta}^{-1}) + 2\, \mathbb{E}(\eta_T^2)\, \mathbb{E}(U_{2,\theta}^{-2}) \\
&\leq (2\rho^2 + 1 + 2\sigma^2) M_1. \qquad (3.84)
\end{aligned}
$$

Similarly, from (3.83), we have

$$
\mathbb{E}\left( U_{2,\theta}^{-2} \sum_{t=2}^{T} \theta_{t-2}^2 \right) \leq 2\, \mathbb{E}\left( U_{2,\theta}^{-1} \right) \leq 2M_1. \qquad (3.85)
$$

Therefore, from (3.84)-(3.85), we have

$$
\mathbb{E}\left( U_{2,\theta}^{-2} \sum_{t=2}^{T} \theta_t^2 + U_{2,\theta}^{-2} \sum_{t=2}^{T} \theta_{t-2}^2 \right) \leq (2\rho^2 + 3 + 2\sigma^2) M_1. \qquad (3.86)
$$

To investigate expectations of $|U_{1,\theta}|U_{2,\theta}^{-2}$, and $|U_{1,\theta}|U_{2,\theta}^{-3}$, we apply Corollary 3.9 to find the order of expectation of $|U_{1,\theta}|U_{2,\theta}^{-2}$, $|U_{1,\theta}|U_{2,\theta}^{-3}$ with $(m, n) = (1, 2)$, and

$(1,3)$, respectively. It follows that

$$\mathbb{E}(|U_{1,\theta}|U_{2,\theta}^{-2}) = \begin{cases} O(T^{-1}) & \text{if } |\rho| < 1, \\ O(T^{-2}) & \text{if } |\rho| = 1, \end{cases} \tag{3.87}$$

and
$$\mathbb{E}(|U_{1,\theta}|U_{2,\theta}^{-3}) = \begin{cases} O(T^{-2}) & \text{if } |\rho| < 1, \\ O(T^{-4}) & \text{if } |\rho| = 1. \end{cases} \tag{3.88}$$

Consider the case $|\rho| < 1$. Since $M_1 = \max\{\mathbb{E}(U_{2,\theta}^{-1}), \mathbb{E}(U_{2,\theta}^{-2})\} \leq \mathbb{E}(U_{2,\theta}^{-1}) + \mathbb{E}(U_{2,\theta}^{-2})$, from Corollary 2.58 (2) and (3.22) we have

$$M_1 \leq \mathbb{E}(U_{2,\theta}^{-1}) + \mathbb{E}(U_{2,\theta}^{-2}) = O\left(\max\{T^{-1}, T^{-2}\}\right) = O(T^{-1}). \tag{3.89}$$

From (3.22), (3.81), (3.86)-(3.88), and (3.89), we have

$$\begin{aligned} \left|\mathbb{E}\left(\hat{U}_{3,\theta} - U_{3,\theta}\right)\right| &= \left|\mathbb{E}\left(\mathbb{E}\left(\hat{U}_{3,\theta} - U_{3,\theta}\middle|\boldsymbol{\theta}\right)\right)\right| \\ &\leq \mathbb{E}\left(\left|\mathbb{E}\left(\hat{U}_{3,\theta} - U_{3,\theta}\middle|\boldsymbol{\theta}\right)\right|\right) \\ &\leq \mathbb{E}\left(D_M^2 U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_t^2 + D_M^2 U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_{t-2}^2\right) + \mathbb{E}\left(D_M^2 U_{2,\theta}^{-1}\right) \\ &\quad + \mathbb{E}\left(2D_M^2(T-1)|U_{1,\theta}|U_{2,\theta}^{-2}\right) + \mathbb{E}\left(4D_M^2|U_{1,\theta}|U_{2,\theta}^{-2}\right) \\ &\leq (2\rho^2 + 3 + 2\sigma^2)D_M^2 M_1 + D_M^2\,\mathbb{E}\left(U_{2,\theta}^{-1}\right) \\ &\quad + 2D_M^2\,\mathbb{E}\left((T-1)|U_{1,\theta}|U_{2,\theta}^{-3}\right) + 4D_M^2\,\mathbb{E}\left(|U_{1,\theta}|U_{2,\theta}^{-2}\right) \\ &= O(T^{-1}) + O(T^{-1}) + O(T)O(T^{-2}) + O(T^{-1}) \\ &= O(T^{-1}). \end{aligned} \tag{3.90}$$

Similarly for the case $|\rho| = 1$, from Corollary 2.58 (2) and (3.22) we have

$$M_1 \leq \mathbb{E}(U_{2,\theta}^{-1}) + \mathbb{E}(U_{2,\theta}^{-2}) = O\left(\max\{T^{-2}, T^{-4}\}\right) = O(T^{-2}). \tag{3.91}$$

From (3.22), (3.81), (3.86)-(3.88), and (3.91), we have

$$\begin{aligned}
\left|\mathbb{E}\left(\hat{U}_{3,\theta} - U_{3,\theta}\right)\right| &= \left|\mathbb{E}\left(\mathbb{E}\left(\hat{U}_{3,\theta} - U_{3,\theta}\middle|\boldsymbol{\theta}\right)\right)\right| \\
&\leq \mathbb{E}\left(\left|\mathbb{E}\left(\hat{U}_{3,\theta} - U_{3,\theta}\middle|\boldsymbol{\theta}\right)\right|\right) \\
&\leq \mathbb{E}\left(D_M^2 U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_t^2 + D_M^2 U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_{t-2}^2\right) + \mathbb{E}\left(D_M^2 U_{2,\theta}^{-1}\right) \\
&\quad + \mathbb{E}\left(2D_M^2(T-1)|U_{1,\theta}|U_{2,\theta}^{-2}\right) + \mathbb{E}\left(4D_M^2|U_{1,\theta}|U_{2,\theta}^{-2}\right) \\
&\leq (2\rho^2 + 3 + \sigma^2)D_M^2 M_1 + D_M^2\,\mathbb{E}\left(U_{2,\theta}^{-1}\right) \\
&\quad + 2D_M^2\,\mathbb{E}\left((T-1)|U_{1,\theta}|U_{2,\theta}^{-3}\right) + 4D_M^2\,\mathbb{E}\left(|U_{1,\theta}|U_{2,\theta}^{-2}\right) \\
&= O(T^{-2}) + O(T^{-2}) + O(T)O(T^{-4}) + O(T^{-2}) \\
&= O(T^{-2}). \qquad (3.92)
\end{aligned}$$

Hence, $\mathbb{E}\left(\hat{U}_{3,\theta} - U_{3,\theta}\right)$ converges to zero for each case of $\rho$ and implies that $\hat{U}_{3,\theta}$ is asymptotically unbiased estimator of $U_{3,\theta}$. In particular, $\mathbb{E}\left(\hat{U}_{3,\theta} - U_{3,\theta}\right)$ is bounded with $CT^{-1}$ for $|\rho| < 1$ and $CT^{-2}$ for $|\rho| = 1$ where $C$ is a constant. □

From Theorem 3.10, the statistic $\hat{U}_{3,\theta}$ can reduce the bias from naive estimator $\rho_y$ when the sample size $T$ goes to infinite. Hence, we denote the statistic $\hat{U}_{3,\theta}$ as $\hat{\rho}_\theta$, the estimator of autoregressive coefficient $\rho$. The following theorem will show some properties of the estimator $\hat{\rho}_\theta$.



**Theorem 3.11.** *Let $\hat{\rho}_\theta = \hat{U}_{3,\theta}$ be defined in Theorem 3.10. Then* $\mathrm{Var}(\hat{\rho}_\theta - \rho_\theta)$ *converges to zero as $T$ goes to infinite.*

*Proof.* Notice that, from Theorem 2.35,

$$\mathrm{Var}\left(\hat{\rho}_\theta - \rho_\theta\right) = \mathbb{E}\left(\mathrm{Var}(\hat{\rho}_\theta - \rho_\theta|\boldsymbol{\theta})\right) + \mathrm{Var}\left(\mathbb{E}(\hat{\rho}_\theta - \rho_\theta|\boldsymbol{\theta})\right). \qquad (3.93)$$

To compute the first term on the right hand side, we will consider $\mathrm{Var}(\hat{\rho}_\theta|\boldsymbol{\theta})$ by applying the first-order Taylor approximations of $\mathrm{Var}\left(\hat{\rho}_\theta|\boldsymbol{\theta}\right) = \mathrm{Var}\left(f(\hat{U}_{1,\theta}, \hat{U}_{2,\theta})\middle|\boldsymbol{\theta}\right)$ around the point $\boldsymbol{\mu} = (\mathbb{E}(\hat{U}_{1,\theta}|\boldsymbol{\theta}), \mathbb{E}(\hat{U}_{2,\theta}|\boldsymbol{\theta}))$, when $f(x,y) = \dfrac{x}{y}$. From (2.14), we have

$$\operatorname{Var}\left(f(\hat{U}_{1,\theta},\hat{U}_{2,\theta})\middle|\boldsymbol{\theta}\right) \approx f_x^2(\boldsymbol{\mu})\operatorname{Var}\left(\hat{U}_{1,\theta}\middle|\boldsymbol{\theta}\right) + f_y^2(\boldsymbol{\mu})\operatorname{Var}\left(\hat{U}_{2,\theta}\middle|\boldsymbol{\theta}\right) + 2f_x(\boldsymbol{\mu})f_y(\boldsymbol{\mu})\operatorname{Cov}\left(\hat{U}_{1,\theta},\hat{U}_{2,\theta}\middle|\boldsymbol{\theta}\right).$$

From (3.77)-(3.79), we can compute the conditional variance of $\hat{\rho}_\theta$ as follows.

$$\begin{aligned}
\operatorname{Var}\left(f(\hat{U}_{1,\theta},\hat{U}_{2,\theta})\middle|\boldsymbol{\theta}\right) &\approx U_{2,\theta}^{-2}\operatorname{Var}(\hat{U}_{1,\theta}|\boldsymbol{\theta}) + U_{1,\theta}^2U_{2,\theta}^{-4}\operatorname{Var}(\hat{U}_{2,\theta}|\boldsymbol{\theta}) \\
&\quad - 2U_{1,\theta}U_{2,\theta}^{-3}\operatorname{Cov}(\hat{U}_{1,\theta},\hat{U}_{2,\theta}|\boldsymbol{\theta}) \\
&= U_{2,\theta}^{-2}\sum_{t=2}^{T}\left(\theta_t^2D_{t-1}^2 + \theta_{t-1}^2D_t^2 + D_t^2D_{t-1}^2 + 2\theta_t\theta_{t-2}D_{t-1}^2\right) \\
&\quad + U_{1,\theta}^2U_{2,\theta}^{-4}(2\sum_{t=2}^{T}D_{t-1}^4 + 4\sum_{t=2}^{T}\theta_{t-1}^2D_{t-1}^2) \\
&\quad - U_{1,\theta}U_{2,\theta}^{-3}(2\sum_{t=2}^{T}\left(\theta_t\theta_{t-1} + \theta_{t-1}\theta_{t-2}\right)D_{t-1}^2) \\
&\leq U_{2,\theta}^{-2}\sum_{t=2}^{T}\left(\theta_t^2D_{t-1}^2 + \theta_{t-1}^2D_t^2 + D_t^2D_{t-1}^2 + (\theta_t^2 + \theta_{t-2}^2)D_{t-1}^2\right) \\
&\quad + U_{1,\theta}^2U_{2,\theta}^{-4}(2\sum_{t=2}^{T}D_{t-1}^4 + 4\sum_{t=2}^{T}\theta_{t-1}^2D_{t-1}^2) \\
&\quad + \left|U_{1,\theta}\right|U_{2,\theta}^{-3}(\sum_{t=2}^{T}\left(\theta_t^2 + 2\theta_{t-1}^2 + \theta_{t-2}^2\right)D_{t-1}^2) \\
&\leq D_M^2U_{2,\theta}^{-2}\sum_{t=2}^{T}\left(\theta_t^2 + \theta_{t-1}^2 + \theta_t^2 + \theta_{t-2}^2\right) + D_M^4U_{2,\theta}^{-2}(T-1) \\
&\quad + 2D_M^2U_{1,\theta}^2U_{2,\theta}^{-4}(T-1) + 4D_M^2U_{1,\theta}^2U_{2,\theta}^{-3} \\
&\quad + D_M^2\left|U_{1,\theta}\right|U_{2,\theta}^{-3}(\sum_{t=2}^{T}\left(\theta_t^2 + 2\theta_{t-1}^2 + \theta_{t-2}^2\right))
\end{aligned}$$

$$
\begin{aligned}
&= 2D_M^2 U_{2,\theta}^{-2} \sum_{t=2}^{T} \theta_t^2 + D_M^2 U_{2,\theta}^{-1} + D_M^2 U_{2,\theta}^{-2} \sum_{t=2}^{T} \theta_{t-2}^2 \\
&\quad + D_M^4 U_{2,\theta}^{-2}(T-1) + 2D_M^2 U_{1,\theta}^2 U_{2,\theta}^{-4}(T-1) + 4D_M^2 U_{1,\theta}^2 U_{2,\theta}^{-3} \\
&\quad + D_M^2 \left|U_{1,\theta}\right| U_{2,\theta}^{-3} \sum_{t=2}^{T} \theta_t^2 + 2D_M^2 \left|U_{1,\theta}\right| U_{2,\theta}^{-2} \\
&\quad + D_M^2 \left|U_{1,\theta}\right| U_{2,\theta}^{-3} \sum_{t=2}^{T} \theta_{t-2}^2, \qquad (3.94)
\end{aligned}
$$

where $D_M^2 = \sup_t D_t^2$. Therefore, by the fact that $\mathrm{Var}(\hat{\rho}_\theta - \rho_\theta|\boldsymbol{\theta}) = \mathrm{Var}(\hat{\rho}_\theta|\boldsymbol{\theta})$, we have

$$
\begin{aligned}
\mathbb{E}\left(\mathrm{Var}(\hat{\rho}_\theta - \rho_\theta|\boldsymbol{\theta})\right) &\leq 2D_M^2\, \mathbb{E}\left(U_{2,\theta}^{-2} \sum_{t=2}^{T} \theta_t^2\right) + D_M^2\, \mathbb{E}\left(U_{2,\theta}^{-1}\right) + D_M^2\, \mathbb{E}\left(U_{2,\theta}^{-2} \sum_{t=2}^{T} \theta_{t-2}^2\right) \\
&\quad + D_M^4(T-1)\, \mathbb{E}\left(U_{2,\theta}^{-2}\right) + 2D_M^2(T-1)\, \mathbb{E}\left(U_{1,\theta}^2 U_{2,\theta}^{-4}\right) \\
&\quad + 4D_M^2\, \mathbb{E}\left(U_{1,\theta}^2 U_{2,\theta}^{-3}\right) + D_M^2\, \mathbb{E}\left(\left|U_{1,\theta}\right| U_{2,\theta}^{-3} \sum_{t=2}^{T} \theta_t^2\right) \\
&\quad + 2D_M^2\, \mathbb{E}\left(\left|U_{1,\theta}\right| U_{2,\theta}^{-2}\right) + D_M^2\, \mathbb{E}\left(\left|U_{1,\theta}\right| U_{2,\theta}^{-3} \sum_{t=2}^{T} \theta_{t-2}^2\right) \\
&= 2D_M^2\, \mathbb{E}\left(U_{2,\theta}^{-2} \sum_{t=2}^{T} \theta_t^2\right) + D_M^2\, \mathbb{E}\left(U_{2,\theta}^{-2} \sum_{t=2}^{T} \theta_{t-2}^2\right) \\
&\quad + D_M^2\, \mathbb{E}\left(\left|U_{1,\theta}\right| U_{2,\theta}^{-3} \sum_{t=2}^{T} \theta_t^2\right) + D_M^2\, \mathbb{E}\left(\left|U_{1,\theta}\right| U_{2,\theta}^{-3} \sum_{t=2}^{T} \theta_{t-2}^2\right) \\
&\quad + D_M^2\, \mathbb{E}\left(U_{2,\theta}^{-1}\right) + D_M^4(T-1)\, \mathbb{E}\left(U_{2,\theta}^{-2}\right) \\
&\quad + 2D_M^2(T-1)\, \mathbb{E}\left(U_{1,\theta}^2 U_{2,\theta}^{-4}\right) + 4D_M^2\, \mathbb{E}\left(U_{1,\theta}^2 U_{2,\theta}^{-3}\right) \\
&\quad + 2D_M^2\, \mathbb{E}\left(\left|U_{1,\theta}\right| U_{2,\theta}^{-2}\right) \\
&\leq 2D_M^2\, \mathbb{E}\left(U_{2,\theta}^{-2} \sum_{t=2}^{T} \theta_t^2\right) + D_M^2\, \mathbb{E}\left(U_{2,\theta}^{-2} \sum_{t=2}^{T} \theta_{t-2}^2\right) \\
&\quad + D_M^2 \sqrt{\mathbb{E}\left(U_{1,\theta}^2 U_{2,\theta}^{-2}\right)} \sqrt{\mathbb{E}\left(U_{2,\theta}^{-2} \sum_{t=2}^{T} \theta_t^2\right)^2} \\
&\quad + D_M^2 \sqrt{\mathbb{E}\left(U_{1,\theta}^2 U_{2,\theta}^{-2}\right)} \sqrt{\mathbb{E}\left(U_{2,\theta}^{-2} \sum_{t=2}^{T} \theta_{t-2}^2\right)^2}
\end{aligned}
$$

$$
\begin{aligned}
&+ D_M^2\,\mathbb{E}\left(U_{2,\theta}^{-1}\right) + D_M^4(T-1)\,\mathbb{E}\left(U_{2,\theta}^{-2}\right) \\
&+ 2D_M^2(T-1)\,\mathbb{E}\left(U_{1,\theta}^{2}U_{2,\theta}^{-4}\right) + 4D_M^2\,\mathbb{E}\left(U_{1,\theta}^{2}U_{2,\theta}^{-3}\right) \\
&+ 2D_M^2\,\mathbb{E}\left(|U_{1,\theta}|\,U_{2,\theta}^{-2}\right) \qquad (3.95)\\
= {}& D_M^2\left(2\,\mathbb{E}\left(U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_t^2\right) + \mathbb{E}\left(U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_{t-2}^2\right)\right) \\
&+ D_M^2\sqrt{\mathbb{E}\left(\rho_\theta^2\right)}\left(\sqrt{\mathbb{E}\left(U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_t^2\right)^2} + \sqrt{\mathbb{E}\left(U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_{t-2}^2\right)^2}\right) \\
&+ D_M^2\,\mathbb{E}\left(U_{2,\theta}^{-1}\right) + D_M^4(T-1)\,\mathbb{E}\left(U_{2,\theta}^{-2}\right) \\
&+ 2D_M^2(T-1)\,\mathbb{E}\left(U_{1,\theta}^{2}U_{2,\theta}^{-4}\right) + 4D_M^2\,\mathbb{E}\left(U_{1,\theta}^{2}U_{2,\theta}^{-3}\right) \\
&+ 2D_M^2\,\mathbb{E}\left(|U_{1,\theta}|\,U_{2,\theta}^{-2}\right), \qquad (3.96)
\end{aligned}
$$

where we use Cauchy-Schwarz inequality to obtain (3.95).

To compute the right hand side of (3.96), we compute the first and the second moments of $U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_t^2$ and $U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_{t-2}^2$, respectively. From (3.84) and (3.85), we have

$$
\mathbb{E}\left(U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_t^2\right) \le (2\rho^2+1+2\sigma^2)M_1,
$$

and

$$
\mathbb{E}\left(U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_{t-2}^2\right) \le 2M_1,
$$



where $M_1 = \max\{\mathbb{E}(U_{2,\theta}^{-1}), \mathbb{E}(U_{2,\theta}^{-2})\}$. Therefore

$$
2D_M^2\,\mathbb{E}(U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_t^2) + D_M^2\,\mathbb{E}(U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_{t-2}^2) \le 2D_M^2(2\rho^2+2+2\sigma^2)M_1. \qquad (3.97)
$$

Next, we will find the second moment of $U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_t^2$ and $U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_{t-2}^2$. Since $U_{2,\theta}$ is a function of $(\theta_1, \theta_2, \ldots, \theta_{T-1})$, $U_{2,\theta}$ and $\eta_T$ are independent. From (3.82), the second moment of $U_{2,\theta}^{-2}\sum\limits_{t=2}^{T}\theta_t^2$ can be computed as follows.

$$
\begin{aligned}
\mathbb{E}\left(U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_t^2\right)^2 &\leq \mathbb{E}\left((2\rho^2+1)U_{2,\theta}^{-1}+2\eta_T^2U_{2,\theta}^{-2}\right)^2 \\
&\leq 2\,\mathbb{E}\left((2\rho^2+1)U_{2,\theta}^{-1}\right)^2+2\,\mathbb{E}\left(2\eta_T^2U_{2,\theta}^{-2}\right)^2 \\
&= 2(2\rho^2+1)^2\,\mathbb{E}(U_{2,\theta}^{-2})+12\sigma^4\,\mathbb{E}(U_{2,\theta}^{-4}) \\
&\leq \left(2(2\rho^2+1)^2+12\sigma^4\right)M_2, \qquad (3.98)
\end{aligned}
$$

where $M_2=\max\{\mathbb{E}(U_{2,\theta}^{-2}),\mathbb{E}(U_{2,\theta}^{-4})\}$. Similarly, from (3.83), we have

$$
\mathbb{E}\left(U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_{t-2}^2\right)^2 \leq 4\,\mathbb{E}(U_{2,\theta}^{-2}). \qquad (3.99)
$$

Next, we compute the expectations of the $U_{1,\theta}^2U_{2,\theta}^{-4}$ and $U_{1,\theta}^2U_{2,\theta}^{-3}$. From Lemma 3.8 with $(m,n)=(2,4)$ and $(1,3)$, we have

$$
\mathbb{E}(U_{1,\theta}^2U_{2,\theta}^{-4}) \leq \begin{cases} O(T^{-2}) & \text{if } |\rho|<1, \\ O(T^{-4}) & \text{if } |\rho|=1, \end{cases} \qquad (3.100)
$$

and

$$
\mathbb{E}(U_{1,\theta}^2U_{2,\theta}^{-3}) \leq \begin{cases} O(T^{-1}) & \text{if } |\rho|<1, \\ O(T^{-2}) & \text{if } |\rho|=1, \end{cases} \qquad (3.101)
$$

respectively.

Consider the case $|\rho|<1$. Notice from Lemma 3.4 (3) that $\mathbb{E}(\rho_\theta^2)=O(1)$. Since $M_2=\max\{\mathbb{E}(U_{2,\theta}^{-2}),\mathbb{E}(U_{2,\theta}^{-4})\}$, from Corollary 2.58 (2) and (3.22) we have

$$
M_2 \leq \mathbb{E}(U_{2,\theta}^{-2})+\mathbb{E}(U_{2,\theta}^{-4}) = O\left(\max\{T^{-2},T^{-4}\}\right)=O(T^{-2}). \qquad (3.102)
$$

From (3.22), (3.87)-(3.89) (3.96)-(3.102), we have

$$
\begin{aligned}
\mathbb{E}\left(\operatorname{Var}(\hat{\rho}_\theta - \rho_\theta|\boldsymbol{\theta})\right) &\leq 2D_M^2(2\rho^2+2+2\sigma^2)M_1 \\
&\quad + D_M^2\sqrt{\mathbb{E}(\rho_\theta^2)}\left(\sqrt{(2(2\rho^2+1)^2+12\sigma^4)\,M_2} + \sqrt{4\,\mathbb{E}(U_{2,\theta}^{-2})}\right) \\
&\quad + D_M^2\,\mathbb{E}\left(U_{2,\theta}^{-1}\right) + D_M^4(T-1)\,\mathbb{E}\left(U_{2,\theta}^{-2}\right) \\
&\quad + 2D_M^2(T-1)\,\mathbb{E}\left(U_{1,\theta}^2U_{2,\theta}^{-4}\right) + 4D_M^2\,\mathbb{E}\left(U_{1,\theta}^2U_{2,\theta}^{-3}\right) \\
&\quad + 2D_M^2\,\mathbb{E}\left(|U_{1,\theta}|\,U_{2,\theta}^{-2}\right) \\
&= O(T^{-1}) + O(1)O(T^{-1}) + O(T^{-1}) + O(T)O(T^{-2}) \\
&\quad + O(T)O(T^{-2}) + O(T^{-1}) + O(T^{-1}) \\
&= O(T^{-1}). \qquad (3.103)
\end{aligned}
$$

Similary for the case $|\rho| = 1$. Since $M_2 = \max\{\mathbb{E}(U_{2,\theta}^{-2}), \mathbb{E}(U_{2,\theta}^{-4})\}$, from Corollary 2.58 (2) and (3.22) we have,

$$
M_2 \leq \mathbb{E}(U_{2,\theta}^{-2}) + \mathbb{E}(U_{2,\theta}^{-4}) = O\left(\max\{T^{-4}, T^{-8}\}\right) = O(T^{-4}). \qquad (3.104)
$$

From (3.22), (3.87)-(3.88), (3.91) (3.96)-(3.101), and (3.104), we have

$$
\begin{aligned}
\mathbb{E}\left(\operatorname{Var}(\hat{\rho}_\theta - \rho_\theta|\boldsymbol{\theta})\right) &\leq 2D_M^2(2\rho^2+2+2\sigma^2)M_1 \\
&\quad + D_M^2\sqrt{\mathbb{E}(\rho_\theta^2)}\left(\sqrt{(2(2\rho^2+1)^2+12\sigma^4)\,M_2} + \sqrt{4\,\mathbb{E}(U_{2,\theta}^{-2})}\right) \\
&\quad + D_M^2\,\mathbb{E}\left(U_{2,\theta}^{-1}\right) + D_M^4(T-1)\,\mathbb{E}\left(U_{2,\theta}^{-2}\right) \\
&\quad + 2D_M^2(T-1)\,\mathbb{E}\left(U_{1,\theta}^2U_{2,\theta}^{-4}\right) + 4D_M^2\,\mathbb{E}\left(U_{1,\theta}^2U_{2,\theta}^{-3}\right) \\
&\quad + 2D_M^2\,\mathbb{E}\left(|U_{1,\theta}|\,U_{2,\theta}^{-2}\right) \\
&= O(T^{-2}) + O(1)O(T^{-2}) + O(T^{-2}) + O(T)O(T^{-4}) \\
&\quad + O(T)O(T^{-4}) + O(T^{-2}) + O(T^{-2}) \\
&= O(T^{-2}). \qquad (3.105)
\end{aligned}
$$

It follows that $\mathbb{E}\left(\text{Var}(\hat{\rho}_\theta - \rho_\theta|\boldsymbol{\theta})\right)$ is of order $T^{-1}$ for $|\rho| < 1$, and of order $T^{-2}$ for $|\rho| = 1$. Hence, the expected value of $\text{Var}(\hat{\rho}_\theta|\boldsymbol{\theta})$ is bound as $C_1T^{-1}$ for each case of $\rho$. In particular,

$$\mathbb{E}\left(\text{Var}(\hat{\rho}_\theta - \rho_\theta|\boldsymbol{\theta})\right) = \begin{cases} O(T^{-1}) & \text{if } |\rho| < 1, \\ O(T^{-2}) & \text{if } |\rho| = 1. \end{cases} \tag{3.106}$$

Next, we will compute the variance of $\mathbb{E}(\hat{\rho}_\theta - \rho_\theta|\boldsymbol{\theta})$. By applying the Cauchy-Schwarz inequality $\left(\sum_{i=1}^n x_i\right)^2 \leq n\left(\sum_{i=1}^n x_i^2\right)$ to (3.81), we have

$$\begin{aligned} \text{Var}\left(|\mathbb{E}(\hat{\rho}_\theta - \rho_\theta|\boldsymbol{\theta})|\right) &\leq \mathbb{E}\left(\mathbb{E}(\hat{\rho}_\theta - \rho_\theta|\boldsymbol{\theta})^2\right) \\ &\leq \mathbb{E}\left(D_M^2 U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_t^2 + D_M^2 U_{2,\theta}^{-1} + D_M^2 U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_{t-2}^2\right. \\ &\qquad \left. + 2D_M^2(T-1)|U_{1,\theta}|U_{2,\theta}^{-3} + 4D_M^2|U_{1,\theta}|U_{2,\theta}^{-2}\right)^2 \\ &\leq 5D_M^2\,\mathbb{E}\left(U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_t^2\right)^2 + 5D_M^2\,\mathbb{E}\left(U_{2,\theta}^{-2}\right) \\ &\qquad + 5D_M^2\,\mathbb{E}\left(U_{2,\theta}^{-2}\sum_{t=2}^{T}\theta_{t-2}^2\right)^2 + 10D_M^2(T-1)^2\,\mathbb{E}\left(U_{1,\theta}^2U_{2,\theta}^{-6}\right) \\ &\qquad + 20D_M^2\,\mathbb{E}\left(U_{1,\theta}^2U_{2,\theta}^{-4}\right). \end{aligned} \tag{3.107}$$



From Lemma 3.8 with $(m, n) = (2, 6)$, the order of expectation of $U_{1,\theta}^2U_{2,\theta}^{-6}$ can be found as:

$$\mathbb{E}(U_{1,\theta}^2U_{2,\theta}^{-6}) \begin{cases} O(T^{-4}) & \text{if } |\rho| < 1, \\ O(T^{-8}) & \text{if } |\rho| = 1. \end{cases} \tag{3.108}$$

Similarly for $U_{1,\theta}^2U_{2,\theta}^{-4}$, from Lemma 3.8 with $(m, n) = (2, 4)$, the order of expectation of $U_{1,\theta}^2U_{2,\theta}^{-4}$ is

$$\mathbb{E}(U_{1,\theta}^2U_{2,\theta}^{-4}) \begin{cases} O(T^{-2}) & \text{if } |\rho| < 1, \\ O(T^{-4}) & \text{if } |\rho| = 1. \end{cases} \tag{3.109}$$

If $|\rho| < 1$, then $\mathbb{E}\left(U_{2,\theta}^{-2}\right) = O(T^{-2})$ by Lemma 3.3 (3). From (3.98)-(3.99), and (3.107)-(3.109), we have

$$\begin{aligned}\operatorname{Var}\left(\mathbb{E}(\hat{\rho}_\theta - \rho_\theta|\boldsymbol{\theta})\right) &= \left(5D_M^2(2(2\rho^2+1)^2 + 12\sigma^4\right) M_2 + 5D_M^2\,\mathbb{E}\left(U_{2,\theta}^{-2}\right) \\ &\quad + 20D_M^2\,\mathbb{E}\left(U_{2,\theta}^{-2}\right) + 10D_M^2(T-1)^2\,\mathbb{E}\left(U_{1,\theta}^2 U_{2,\theta}^{-6}\right) \\ &\quad + 20D_M^2\,\mathbb{E}\left(U_{1,\theta}^2 U_{2,\theta}^{-4}\right) \\ &= O(T^{-2}) + O(T^{-2}) + O(T^{-2}) + O(T^2)O(T^{-4}) + O(T^{-2}) \\ &= O(T^{-2}). \end{aligned} \tag{3.110}$$

Similarly for the case $|\rho| = 1$, we have $\mathbb{E}\left(U_{2,\theta}^{-2}\right) = O(T^{-4})$. From (3.98)-(3.99), and (3.107)-(3.109), we have

$$\begin{aligned}\operatorname{Var}\left(\mathbb{E}(\hat{\rho}_\theta - \rho_\theta|\boldsymbol{\theta})\right) &= \left(5D_M^2(2(2\rho^2+1)^2 + 12\sigma^4\right) M_2 + 5D_M^2\,\mathbb{E}\left(U_{2,\theta}^{-2}\right) \\ &\quad + 20D_M^2\,\mathbb{E}\left(U_{2,\theta}^{-2}\right) + 10D_M^2(T-1)^2\,\mathbb{E}\left(U_{1,\theta}^2 U_{2,\theta}^{-6}\right) \\ &\quad + 20D_M^2\,\mathbb{E}\left(U_{1,\theta}^2 U_{2,\theta}^{-4}\right) \\ &= O(T^{-4}) + O(T^{-4}) + O(T^{-4}) + O(T^2)O(T^{-8}) + O(T^{-4}) \\ &= O(T^{-4}). \end{aligned} \tag{3.111}$$

Therefore, by using (3.106) and (3.110)-(3.111) in (3.93), we have

$$\begin{aligned}\operatorname{Var}\left(\hat{\rho}_\theta - \rho_\theta\right) &= \mathbb{E}\left(\operatorname{Var}(\hat{\rho}_\theta - \rho_\theta|\boldsymbol{\theta})\right) + \operatorname{Var}\left(\mathbb{E}(\hat{\rho}_\theta - \rho_\theta|\boldsymbol{\theta})\right) \\ &= \begin{cases} O(T^{-1}) + O(T^{-2}) & \text{if } |\rho| < 1, \\ O(T^{-2}) + O(T^{-4}) & \text{if } |\rho| = 1. \end{cases} \\ &= \begin{cases} O(T^{-1}) & \text{if } |\rho| < 1, \\ O(T^{-2}) & \text{if } |\rho| = 1. \end{cases} \end{aligned} \tag{3.112}$$

Hence, $\operatorname{Var}\left(\hat{\rho}_\theta - \rho_\theta\right) < \frac{C_3}{T}$ for some constant $C_3$, and implies that it converges to zero for all $|\rho| \leq 1$. □

**Theorem 3.12.** *Let $\hat{\rho}_\theta = \hat{U}_{3,\theta}$ be defined in Theorem 3.10. Then, $\hat{\rho}_\theta$ is a consistent estimator of $\rho$.*

*Proof.* From Theorem 2.48, we can conclude that

$$\hat{\rho}_\theta - \rho_\theta \xrightarrow{p} 0. \tag{3.113}$$

Since $\rho_\theta$ is a consistent estimator of $\rho$, we have

$$\hat{\rho}_\theta - \rho = (\hat{\rho}_\theta - \rho_\theta) + (\rho_\theta - \rho) \xrightarrow{p} 0.$$

□

**Corollary 3.13.** *Let $\hat{\rho}_\theta = \hat{U}_{3,\theta}$ be defined in Theorem 3.10. Then,*

*(1) $\hat{\rho}_\theta - \rho_\theta = O_p(T^{-\frac{1}{2}})$ if $|\rho| < 1$,*

*(2) $\hat{\rho}_\theta - \rho_\theta = O_p(T^{-1})$ if $|\rho| = 1$.*

*Proof.* (1) Notice that, from (3.90) and (3.112) with $|\rho| < 1$,

$$\mathbb{E}\left(\hat{\rho}_\theta - \rho_\theta\right)^2 = \mathrm{Var}\left(\hat{\rho}_\theta - \rho_\theta\right) + \mathbb{E}^2\left(\hat{\rho}_\theta - \rho_\theta\right) = O(T^{-1}) + O(T^{-2}) = O(T^{-1}). \tag{3.114}$$

By Corollary 2.62, $\hat{\rho}_\theta - \rho_\theta = O_p(T^{-\frac{1}{2}})$.

(2) Similarly for the case $|\rho| = 1$, from (3.92) and (3.112), we have

$$\mathbb{E}\left(\hat{\rho}_\theta - \rho_\theta\right)^2 = \mathrm{Var}\left(\hat{\rho}_\theta - \rho_\theta\right) + \mathbb{E}^2\left(\hat{\rho}_\theta - \rho_\theta\right) = O(T^{-2}) + O(T^{-4}) = O(T^{-2}). \tag{3.115}$$

By Corollary 2.62, $\hat{\rho}_\theta - \rho_\theta = O_p(T^{-1})$. □

### 3.2.4 Asymptotic Distribution for Estimator of Autoregressive Coefficient

In this section, we consider the asymptotic distribution of the estimator $\hat{\rho}_\theta$ with $\rho = 1$. Moreover, we will show that $\hat{\rho}_\theta$ has the same asymptotic distribution as $\rho_\theta$.

**Lemma 3.14.** *Under the assumption on the model (3.13) with* $\rho = 1$ , $\dfrac{\hat{U}_{2,\theta}}{T^2} - \dfrac{U_{2,\theta}}{T^2}$ *converges to zero in probability. In particular,*

$$\frac{\hat{U}_{2,\theta}}{T^2} \xrightarrow{p} \sum_{i=1}^{\infty} \gamma_i^2 Z_i'^2, \tag{3.116}$$

*where* $\gamma_i = (-1)^{i+1} \frac{2}{(2i-1)\pi}$ *and* $Z_i' \overset{iid}{\sim} N(0, \sigma^2)$.

*Proof.* Notices that

$$\begin{aligned}
\frac{\hat{U}_{2,\theta}}{T^2} - \frac{U_{2,\theta}}{T^2} &= \frac{1}{T^2} \sum_{t=2}^{T} (y_{t-1}^2 - D_{t-1}^2) - \frac{1}{T^2} \sum_{t=2}^{T} \theta_{t-1}^2 \\
&= \frac{1}{T^2} \sum_{t=2}^{T} 2\theta_{t-1} e_{t-1} + \frac{1}{T^2} \sum_{t=2}^{T} (e_{t-1}^2 - D_{t-1}^2) \\
&= \frac{1}{T} \sum_{t=2}^{T} 2 \left( \frac{\theta_{t-1} e_{t-1}}{T} \right) + \frac{1}{T^2} \sum_{t=2}^{T} (e_{t-1}^2 - D_{t-1}^2).
\end{aligned} \tag{3.117}$$

For the first term, notice that $\mathbb{E}(\theta_i e_i) = \mathbb{E}(\theta_i)\,\mathbb{E}(e_i) = 0$ for all $i$, and

$$\operatorname{Cov}(\theta_i e_i, \theta_j e_j) = \mathbb{E}(\theta_i \theta_j e_i e_j) = \mathbb{E}(\theta_i \theta_j)\,\mathbb{E}(e_i e_j) = 0$$

for any $i \neq j$. Hence, the sequence of random variables $(\theta_t e_t)_{t \geq 1}$ is pairwise uncorrelated with zero mean and



$$\operatorname{Var}\left( \frac{\theta_{t-1} e_{t-1}}{T} \right) = \frac{1}{T^2}\,\mathbb{E}(\theta_{t-1}^2)\,\mathbb{E}(e_{t-1}^2) = \frac{(t-1)\sigma^2 D_{t-1}^2}{T^2} < \sigma^2 D_M^2.$$

Hence, by Theorem 2.47 (2), we can conclude that $\frac{1}{T} \sum_{t=2}^{T-1} 2 \left( \frac{\theta_{t-1}}{T} e_{t-1} \right)$ converges to zero in probability.

Since $\left(e_{t-1}^2 - D_{t-1}^2\right)_{t \geq 2}$ is a sequence of random variables with zero mean and variance $2D_{t-1}^4 < 2D_M^4$, from Theorem 2.47 (2), $\frac{1}{T^2} \sum_{t=2}^{T} (e_{t-1}^2 - D_{t-1}^2)$ converges to zero in probability.

Hence,

$$\frac{\hat{U}_{2,\theta}}{T^2} - \frac{U_{2,\theta}}{T^2} \xrightarrow{p} 0. \tag{3.118}$$

From Theorem 3.2 (2), we have

$$\frac{U_{2,\theta}}{T^2} \xrightarrow{p} \sum_{i=1}^{\infty} \gamma_i^2 Z_i'^2,$$

where $\gamma_i = (-1)^{i+1}\frac{2}{(2i-1)\pi}$ and $Z_i' \overset{iid}{\sim} N(0,\sigma^2)$. Hence, from Theorem 2.42, we can conclude that

$$\frac{\hat{U}_{2,\theta}}{T^2} = \left(\frac{\hat{U}_{2,\theta}}{T^2} - \frac{U_{2,\theta}}{T^2}\right) + \frac{U_{2,\theta}}{T^2} \xrightarrow{p} 0 + \sum_{i=1}^{\infty} \gamma_i^2 Z_i'^2 = \sum_{i=1}^{\infty} \gamma_i^2 Z_i'^2.$$

□

**Theorem 3.15.** *Under the assumption of the model (3.13) with $\rho = 1$. Then*

$$T(\hat{\rho}_\theta - 1) \xrightarrow{p} \frac{\left(\sum\limits_{i=1}^{\infty} \sqrt{2}\gamma_i Z_i'\right)^2 - \sigma^2}{2\sum\limits_{i=1}^{\infty} \gamma_i^2 Z_i'^2}, \tag{3.119}$$

*where $\gamma_i = (-1)^{i+1}\frac{2}{(2i-1)\pi}$ and $Z_i' \overset{iid}{\sim} N(0,\sigma^2)$. In particular, the asymptotic distribution of $T(\hat{\rho}_\theta - 1)$ is the same as $T(\rho_\theta - 1)$.*



*Proof.* Notice that

$$\begin{aligned} T(\hat{\rho}_\theta - 1) &= T\left(\sum_{t=2}^{T}\left(y_{t-1}^2 - D_{t-1}^2\right)\right)^{-1}\left(\sum_{t=2}^{T}\left(y_t y_{t-1} - y_{t-1}^2 + D_{t-1}^2\right)\right) \\ &= \left(\frac{1}{T^2}\sum_{t=2}^{T}\left(y_{t-1}^2 - D_{t-1}^2\right)\right)^{-1}\left(\frac{1}{T}\sum_{t=2}^{T}\left(y_t y_{t-1} - y_{t-1}^2 + D_{t-1}^2\right)\right). \end{aligned} \tag{3.120}$$

From (3.13), the numerator term of $T(\hat{\rho}_\theta - 1)$ can be simplified as:

$$\begin{aligned} \frac{1}{T}\sum_{t=2}^{T}\left(y_t y_{t-1} - y_{t-1}^2 + D_{t-1}^2\right) &= \frac{1}{T}\sum_{t=2}^{T}\left((\theta_{t-1} + e_{t-1})(\theta_t + e_t - \theta_{t-1} - e_{t-1}) + D_{t-1}^2\right) \\ &= \frac{1}{T}\sum_{t=2}^{T}\left((\theta_{t-1} + e_{t-1})(\eta_t + e_t - e_{t-1}) + D_{t-1}^2\right) \end{aligned}$$

$$= \frac{1}{T}\sum_{t=2}^{T}\theta_{t-1}\eta_t + \frac{1}{T}\sum_{t=2}^{T}\theta_{t-1}e_t - \frac{1}{T}\sum_{t=2}^{T}\theta_{t-1}e_{t-1}$$
$$+ \frac{1}{T}\sum_{t=2}^{T}\eta_t e_{t-1} + \frac{1}{T}\sum_{t=2}^{T}e_t e_{t-1} - \frac{1}{T}\sum_{t=2}^{T}(e_{t-1}^2 - D_{t-1}^2). \tag{3.121}$$

To find the limit of the term on the left hand side in (3.121), we find limits of the following terms.

(a) $\frac{1}{T}\sum_{t=2}^{T}\theta_{t-1}\eta_t$.

(b) $\frac{1}{T}\sum_{t=2}^{T}\theta_{t-1}\left(e_t - e_{t-1}\right)$.

(c) $\frac{1}{T}\sum_{t=2}^{T}\eta_t e_{t-1}$.

(d) $\frac{1}{T}\sum_{t=2}^{T}e_t e_{t-1}$.

(e) $\frac{1}{T}\sum_{t=2}^{T}(e_{t-1}^2 - D_{t-1}^2)$.

(a) From Theorem 3.2 (1) on the model (3.8), we can conclude that

$$\frac{1}{T}\sum_{t=2}^{T}\theta_{t-1}\eta_t \xrightarrow{p} \frac{1}{2}\left(\sum_{i=1}^{\infty}\sqrt{2}\gamma_i Z_i'\right)^2 - \frac{\sigma^2}{2}, \tag{3.122}$$



where $\gamma_i = (-1)^{i+1}\frac{2}{(2i-1)\pi}$, and $Z_i' \overset{iid}{\sim} N(0, \sigma^2)$.

(b) Notice from (3.13) with $\rho = 1$ that $\eta_t = \theta_t - \theta_{t-1}$ and $\theta_t = \sum_{i=1}^{t}\eta_t$. Hence, we have

$$\frac{1}{T}\sum_{t=2}^{T}\theta_{t-1}\left(e_t - e_{t-1}\right) = \frac{1}{T}\sum_{t=2}^{T}\theta_{t-1}e_t - \frac{1}{T}\sum_{t=1}^{T-1}\theta_t e_t$$
$$= \frac{\theta_{T-1}e_T}{T} - \frac{\theta_1 e_1}{T} + \frac{1}{T}\sum_{t=2}^{T-1}(\theta_{t-1} - \theta_t)e_t$$
$$= \frac{1}{T}\sum_{i=1}^{T-1}\eta_i e_T - \frac{\eta_1 e_1}{T} - \frac{1}{T}\sum_{t=2}^{T-1}\eta_t e_t. \tag{3.123}$$

Since $(\eta_i e_T)_{i\geq 1}$ is a sequence of uncorrelated random variables with zero mean and variance $\sigma^2 D_T^2 \leq \sigma^2 D_M^2$ for all $i$, from Theorem 2.47 (2), $\frac{1}{T}\sum_{i=1}^{T-1}\eta_i e_T$ converges to zero in probability.

Consider $\frac{\eta_1 e_1}{T}$. Notice that $\mathbb{E}\left(\frac{\eta_1 e_1}{T}\right) = 0$, and $\operatorname{Var}\left(\frac{\eta_1 e_1}{T}\right) = \frac{\sigma^2 D_1^2}{T^2}$. Since $\operatorname{Var}\left(\frac{\eta_1 e_1}{T}\right)$ converges to zero, by Theorem 2.48, $\frac{\eta_1 e_1}{T}$ converges to zero in probability.

Since $(\eta_t e_t)_{t\geq 1}$ is a sequence of independent random variables with zero mean and variance $\sigma^2 D_t^2 < \sigma^2 D_M^2$ for all $t$, from Theorem 2.47 (2), $\dfrac{1}{T}\sum\limits_{t=2}^{T-1}\eta_t e_t$ converges to zero in probability.

Hence, from (3.123), we can conclude that

$$\frac{1}{T}\sum_{t=2}^{T}\theta_{t-1}\left(e_t - e_{t-1}\right) \xrightarrow{p} 0. \tag{3.124}$$

(c) Notice that $(\eta_t e_{t-1})_{t\geq 2}$ is a sequence of independent random variables with zero mean and variance $\sigma^2 D_{t-1}^2 \leq \sigma^2 D_M^2$. From Theorem 2.47 (2), $\dfrac{1}{T}\sum\limits_{t=2}^{T}\eta_t e_{t-1}$ converges to zero in probability.

(d) Since $(e_t e_{t-1})_{t\geq 2}$ is a sequence of uncorrelated random variables with zero mean and variance $D_t^2 D_{t-1}^2 \leq D_M^4$, from Theorem 2.47 (2), $\dfrac{1}{T}\sum\limits_{t=2}^{T} e_t e_{t-1}$ converges to zero in probability.

(e) Since $(e_t^2 - D_t^2)_{t\geq 1}$ is a sequence of independent random variables with zero mean and variance $2D_t^2 \leq 2D_M^2$, from Theorem 2.47 (2), $\frac{1}{T}\sum\limits_{t=2}^{T}(e_{t-1}^2 - D_{t-1}^2)$ converges to zero in probability.

Hence, from (3.121) and (a)-(e),

$$\frac{1}{T}\sum_{t=2}^{T}\left(y_t y_{t-1} - y_{t-1}^2 + D_{t-1}^2\right) \xrightarrow{p} \frac{1}{2}\left(\sum_{i=1}^{\infty}\sqrt{2}\gamma_i Z_i'\right)^2 - \frac{\sigma^2}{2}. \tag{3.125}$$

Let $g(x) = \dfrac{1}{x}$. Since $g$ is a continuous function, by applying Theorem 2.41 to (3.116) with function $g$, we have

$$\frac{1}{\frac{\hat{U}_{2,\theta}}{T^2}} \xrightarrow{p} \frac{1}{\sum\limits_{i=1}^{\infty}\gamma_i^2 Z_i'^2}. \tag{3.126}$$

From (3.125), (3.126), and Theorem 2.42, we can conclude that

$$T(\hat{\rho}_\theta - 1) \xrightarrow{p} \frac{\left(\sum_{i=1}^{\infty} \sqrt{2}\gamma_i Z_i'\right)^2 - \sigma^2}{2\sum_{i=1}^{\infty} \gamma_i^2 Z_i'^2}, \tag{3.127}$$

where $\gamma_i = (-1)^{i+1}\frac{2}{(2i-1)\pi}$, and $Z_i' \overset{iid}{\sim} N(0, \sigma^2)$. From Theorem 3.2 (4), the asymptotic distribution of $T(\hat{\rho}_\theta - 1)$ and $T(\rho_\theta - 1)$ are the same. $\square$

## 3.3 Test Statistic for Stationary Test

In this section, we derive a test statistic for the stationary test for the model (3.13). However, the estimator of the variance of noise $\sigma^2$ is necessary to obtain the test statistic. Hence, we will first obtain a consistent estimator of $\sigma^2$ in Section 3.3.1. Then we use the estimate to obtain a stationary test in Section 3.3.2.

### 3.3.1 Estimator of Variance of Noises

In this section, we derive an estimator for the variance of noises $\sigma^2$ in the model (3.13). This estimator will be used to find the test statistic for the stationary test.

Assume that there is no sampling errors in the model (3.13). Dickey [5] showed that the estimator for $\sigma^2$ is $U_{4,\theta}$, defined in (3.19). In addition, the estimator $U_{4,\theta}$ is a consistent estimator.



**Proposition 3.16** ([5], Corollary 4.2, page 43)**.** *The statistic $U_{4,\theta}$ defined in (3.19) is a consistent estimator of $\sigma^2$ for all $|\rho| \leq 1$.*

However, if we ignore sampling errors and replace $\theta_t$ with $y_t$, there is a bias term in the naive estimator of $\sigma^2$, denoted by $U_{4,y}$. Therefore, the naive estimator $U_{4,y}$ is not unbiased. In Theorem 3.17, we derive a consistent estimator of $\sigma^2$. To simplify notations, we define the following functions of a random variable $P$.

$$U_\theta(P) = \frac{1}{T-2}\sum_{t=2}^{T}(\theta_t - P\theta_{t-1})^2,$$
$$U_y(P) = \frac{1}{T-2}\sum_{t=2}^{T}(y_t - Py_{t-1})^2. \tag{3.128}$$

Notice that $U_\theta(\rho_\theta) = U_{4,\theta}$ and $U_y(\rho_y) = U_{4,y}$. The next theorem will show the consistent estimator of $\sigma^2$. Some notations are followed from (3.19).

**Theorem 3.17.** *Define the statistic $\hat{U}_{4,\theta}$ by*

$$\hat{U}_{4,\theta} = \frac{1}{T-2}\sum_{t=2}^{T}(y_t - \hat{\rho}_\theta y_{t-1})^2 - \frac{1}{T-2}\sum_{t=2}^{T}\left(D_t^2 + \hat{\rho}_\theta^2 D_{t-1}^2\right).$$

*Then $\hat{U}_{4,\theta}$ is a consistent estimator of $\sigma^2$ for all $|\rho| \leq 1$.*

*Proof.* The statistic $\hat{U}_{4,\theta}$ can be written as $U_y(\hat{\rho}_\theta) - \frac{1}{T-2}\sum_{t=2}^{T}\left(D_t^2 + \hat{\rho}_\theta^2 D_{t-1}^2\right)$. From the model (3.13), we notice that

$$\begin{aligned}
(T-2)U_y(\hat{\rho}_\theta) &= \sum_{t=2}^{T}(y_t - \hat{\rho}_\theta y_{t-1})^2 \\
&= \sum_{t=2}^{T}\left(\theta_t + e_t - \hat{\rho}_\theta\theta_{t-1} - \hat{\rho}_\theta e_{t-1}\right)^2 \\
&= \sum_{t=2}^{T}\left(\theta_t - \hat{\rho}_\theta\theta_{t-1}\right)^2 + \sum_{t=2}^{T}\left(e_t - \hat{\rho}_\theta e_{t-1}\right)^2 \\
&\quad + 2\sum_{t=2}^{T}\left(\theta_t - \hat{\rho}_\theta\theta_{t-1}\right)\left(e_t - \hat{\rho}_\theta e_{t-1}\right) \\
&= \sum_{t=2}^{T}\left(\theta_t - \rho_\theta\theta_{t-1} + (\rho_\theta - \hat{\rho}_\theta)\theta_{t-1}\right)^2 + \sum_{t=2}^{T}\left(e_t - \hat{\rho}_\theta e_{t-1}\right)^2 \\
&\quad + 2\sum_{t=2}^{T}\left(\theta_t - \rho\theta_{t-1} + (\rho - \hat{\rho}_\theta)\theta_{t-1}\right)\left(e_t - \hat{\rho}_\theta e_{t-1}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{t=2}^{T} (\theta_t - \rho_\theta \theta_{t-1})^2 + (\rho_\theta - \hat{\rho}_\theta)^2 \sum_{t=2}^{T} \theta_{t-1}^2 \\
&\quad + 2(\rho_\theta - \hat{\rho}_\theta) \sum_{t=2}^{T} (\theta_t - \rho_\theta \theta_{t-1}) \theta_{t-1} + \sum_{t=2}^{T} (e_t - \hat{\rho}_\theta e_{t-1})^2 \\
&\quad + 2 \sum_{t=2}^{T} (\eta_t + (\rho - \hat{\rho}_\theta)\theta_{t-1}) (e_t - \hat{\rho}_\theta e_{t-1}) \\
&= (T-2) U_\theta(\rho_\theta) + (\rho_\theta - \hat{\rho}_\theta)^2 \sum_{t=2}^{T} \theta_{t-1}^2 \\
&\quad + 2(\rho_\theta - \hat{\rho}_\theta) \left( \sum_{t=2}^{T} \theta_t \theta_{t-1} - \rho_\theta \sum_{t=2}^{T} \theta_{t-1}^2 \right) + \sum_{t=2}^{T} (e_t - \hat{\rho}_\theta e_{t-1})^2 \\
&\quad + 2 \sum_{t=2}^{T} (\eta_t + (\rho - \hat{\rho}_\theta)\theta_{t-1}) (e_t - \hat{\rho}_\theta e_{t-1}) \\
&= (T-2) U_\theta(\rho_\theta) + (\rho_\theta - \hat{\rho}_\theta)^2 U_{2,\theta} + 2(\rho_\theta - \hat{\rho}_\theta) (U_{1,\theta} - \rho_\theta U_{2,\theta}) \\
&\quad + \sum_{t=2}^{T} (e_t - \hat{\rho}_\theta e_{t-1})^2 + 2 \sum_{t=2}^{T} (\eta_t + (\rho - \hat{\rho}_\theta)\theta_{t-1}) (e_t - \hat{\rho}_\theta e_{t-1}) \\
&= (T-2) U_\theta(\rho_\theta) + (\rho_\theta - \hat{\rho}_\theta)^2 U_{2,\theta} + \sum_{t=2}^{T} (e_t - \hat{\rho}_\theta e_{t-1})^2 \\
&\quad + 2 \sum_{t=2}^{T} (\eta_t + (\rho - \hat{\rho}_\theta)\theta_{t-1}) (e_t - \hat{\rho}_\theta e_{t-1}), \qquad (3.129)
\end{aligned}
$$

where we apply the fact that $\rho_\theta = U_{1,\theta} U_{2,\theta}^{-1}$ to obtain (3.129).

Hence, from (3.129), we can simplify the statistic $\hat{U}_{4,\theta} - U_{4,\theta}$ as follows.

$$
\begin{aligned}
\hat{U}_{4,\theta} - U_{4,\theta} &= U_y(\hat{\rho}_\theta) - \frac{1}{T-2} \sum_{t=2}^{T} \left( D_t^2 + \hat{\rho}_\theta^2 D_{t-1}^2 \right) - U_\theta(\rho_\theta) \\
&= \frac{1}{T-2} (\rho_\theta - \hat{\rho}_\theta)^2 U_{2,\theta} + \frac{1}{T-2} \sum_{t=2}^{T} (e_t - \hat{\rho}_\theta e_{t-1})^2 \\
&\quad - \frac{1}{T-2} \sum_{t=2}^{T} \left( D_t^2 + \hat{\rho}_\theta^2 D_{t-1}^2 \right) \\
&\quad + \frac{2}{T-2} \sum_{t=2}^{T} (\eta_t + (\rho - \hat{\rho}_\theta)\theta_{t-1}) (e_t - \hat{\rho}_\theta e_{t-1}). \qquad (3.130)
\end{aligned}
$$

To investigate the convergence in probability of the term on the right hand side of (3.130), we find the convergence in probability of the following terms.

(a) $\frac{1}{T-2}(\rho_\theta - \hat{\rho}_\theta)^2 U_{2,\theta}$,

(b) $\frac{1}{T-2}\sum_{t=2}^{T}(e_t - \hat{\rho}_\theta e_{t-1})^2 - \frac{1}{T-2}\sum_{t=2}^{T}\left(D_t^2 + \hat{\rho}_\theta^2 D_{t-1}^2\right)$,

(c) $\frac{2}{T-2}\sum_{t=2}^{T}(\eta_t + (\rho - \hat{\rho}_\theta)\theta_{t-1})(e_t - \hat{\rho}_\theta e_{t-1})$.

For (a), we will consider into two cases: $|\rho| < 1$, and $|\rho| = 1$. Consider the case $|\rho| < 1$. From Theorem 3.1 (2), $\frac{U_{2,\theta}}{T}$ converges in probability to $\frac{\sigma^2}{1-\rho^2}$. By applying Theorem 2.42 (2) to (3.113), we can conclude that

$$\frac{1}{T-2}(\rho_\theta - \hat{\rho}_\theta)^2 U_{2,\theta} = \frac{T}{T-2}(\rho_\theta - \hat{\rho}_\theta)^2 \frac{U_{2,\theta}}{T} \xrightarrow{p} 0.$$

Consider the case $|\rho| = 1$. From Theorem 3.2 (2),

$$\frac{U_{2,\theta}}{T^2} \xrightarrow{p} \sum_{i=1}^{\infty} \gamma_i^2 Z_i'^2,$$

where $Z_i' \overset{iid}{\sim} N(0,\sigma^2)$ and $\gamma_i = (-1)^{i+1}\frac{2}{(2i-1)\pi}$. From Corollary 3.13 (2), $T(\rho_\theta - \hat{\rho}_\theta) = O_p(1)$. In addition, from (3.113), $\rho_\theta - \hat{\rho}_\theta = o_p(1)$. By applying Corollary 2.61 (3), it follows that $T(\rho_\theta - \hat{\rho}_\theta)^2 = O_p(1)o_p(1) = o_p(1)$. Hence, we can conclude that

$$\frac{1}{T-2}(\rho_\theta - \hat{\rho}_\theta)^2 U_{2,\theta} = \left(\frac{T}{T-2}\right) T(\rho_\theta - \hat{\rho}_\theta)^2 \frac{U_{2,\theta}}{T^2} = o_p(1).$$

Therefore, $\frac{1}{T-2}(\rho_\theta - \hat{\rho}_\theta)^2 U_{2,\theta}$ converges to zero in probability.

For (b), notice that

$$\begin{aligned}&\frac{1}{T-2}\sum_{t=2}^{T}(e_t - \hat{\rho}_\theta e_{t-1})^2 - \frac{1}{T-2}\sum_{t=2}^{T}\left(D_t^2 + \hat{\rho}_\theta^2 D_{t-1}^2\right)\\ &\quad = \frac{1}{T-2}\sum_{t=2}^{T}\left(e_t^2 - D_t^2\right) + \hat{\rho}_\theta^2 \frac{1}{T-2}\sum_{t=2}^{T}\left(e_{t-1}^2 - D_{t-1}^2\right) + 2\hat{\rho}_\theta \frac{1}{T-2}\sum_{t=2}^{T} e_t e_{t-1}.\end{aligned} \tag{3.131}$$

From Theorem 3.12, $\hat{\rho}_\theta$ converges in probability to $\rho$. Since $(e_t^2 - D_t^2)_{t\geq 1}$ is a sequence of independent random variables with zero mean and variance $2D_t^4 \leq 2D_M^4$, from Theorem 2.47 (2), we have

$$\frac{1}{T-2}\sum_{t=2}^{T}\left(e_t^2 - D_t^2\right) + \hat{\rho}_\theta^2 \frac{1}{T-2}\sum_{t=2}^{T}\left(e_{t-1}^2 - D_{t-1}^2\right) \xrightarrow{p} 0.$$

Notice that $(e_t e_{t-1})_{t\geq 1}$ is a sequence of uncorrelated random variables with zero mean and variance $D_t^2 D_{t-1}^2 \leq D_M^4$, from Theorem 2.47 (2), we have

$$2\hat{\rho}_\theta \frac{1}{T-2}\sum_{t=2}^{T} e_t e_{t-1} \xrightarrow{p} 0.$$

Therefore, $\frac{1}{T-2}\sum_{t=2}^{T}\left(e_t - \hat{\rho}_\theta e_{t-1}\right)^2 - \frac{1}{T-2}\sum_{t=2}^{T}\left(D_t^2 + \hat{\rho}_\theta^2 D_{t-1}^2\right)$ converges to zero in probability.

For (c), notice that

$$\begin{aligned}
&\frac{2}{T-2}\sum_{t=2}^{T}\left(\eta_t + (\rho - \hat{\rho}_\theta)\theta_{t-1}\right)\left(e_t - \hat{\rho}_\theta e_{t-1}\right) \\
&= \frac{2}{T-2}\sum_{t=2}^{T}\eta_t e_t - \frac{2\hat{\rho}_\theta}{T-2}\sum_{t=2}^{T}\eta_t e_{t-1} + \frac{2(\rho - \hat{\rho}_\theta)}{T-2}\sum_{t=2}^{T}\theta_{t-1}e_t - \frac{2\hat{\rho}_\theta(\rho - \hat{\rho}_\theta)}{T-2}\sum_{t=2}^{T}\theta_{t-1}e_{t-1}.
\end{aligned} \tag{3.132}$$



Since $(\eta_t e_t)_{t\geq 1}$ is a sequence of independent random variables with zero mean and variance $D_t^2\sigma^2 \leq \sigma^2 D_M^2$, from Theorem 2.47 (2),

$$\frac{2}{T-2}\sum_{t=2}^{T}\eta_t e_t \xrightarrow{p} 0, \tag{3.133}$$

for $|\rho| \leq 1$.

Similarly for the sequence of independent random variables $(\eta_t e_{t-1})_{t\geq 2}$. Since $\hat{\rho}_\theta$ is a consistent estimator of $\rho$, from Theorem 2.42 and Theorem 2.47 (2),

$$\frac{2\hat{\rho}_\theta}{T-2}\sum_{t=2}^{T}\eta_t e_{t-1} \xrightarrow{p} 0, \tag{3.134}$$

for $|\rho| \leq 1$.

For the last two terms in (3.132), we will consider into two cases: $|\rho| < 1$, and $|\rho| = 1$.

Consider the case $|\rho| < 1$. By Theorem 3.12, $\rho - \hat{\rho}_\theta$ and $\hat{\rho}_\theta(\rho - \hat{\rho}_\theta)$ both converge to zero in probability. Recall from Section 3.1.1 that $\text{Var}(\theta_t) = \dfrac{\sigma^2}{1-\rho^2}$ for all $t \geq 1$. Since $(\theta_t)_{t\geq 1}$ and $(e_t)_{t\geq 1}$ are pairwise independent, $(\theta_{t-1}e_t)_{t\geq 2}$ is a sequence of uncorrelated random variables with zero mean and variance $\dfrac{\sigma^2 D_t^2}{1-\rho^2} \leq \dfrac{\sigma^2 D_M^2}{1-\rho^2}$. Hence, from Theorem 2.42 and Theorem 2.47 (2),

$$\frac{2(\rho - \hat{\rho}_\theta)}{T-2}\sum_{t=2}^{T}\theta_{t-1}e_t \xrightarrow{p} 0. \tag{3.135}$$

Similarly, $(\theta_{t-1}e_{t-1})_{t\geq 2}$ is a sequence of uncorrelated random variables with zero mean and variance $\dfrac{\sigma^2 D_{t-1}^2}{1-\rho^2} \leq \dfrac{\sigma^2 D_M^2}{1-\rho^2}$. Hence, from Theorem 2.42 and Theorem 2.47 (2),

$$\frac{2\hat{\rho}_\theta(\rho - \hat{\rho}_\theta)}{T-2}\sum_{t=2}^{T}\theta_{t-1}e_{t-1} \xrightarrow{p} 0. \tag{3.136}$$

Consider the case $|\rho| = 1$. From Theorem 3.12, we notice that $\rho - \hat{\rho}_\theta = o_p(1)$ and $\hat{\rho}_\theta(\rho - \hat{\rho}_\theta) = o_p(1)$. First, we consider $\frac{1}{T-2}\sum\limits_{t=2}^{T}\theta_{t-1}e_t$.

Recall from Section 3.1.1 that $\mathbb{E}(\theta_t^2) = 0$ and $\text{Var}(\theta_t) = t\sigma^2$ for all $t \geq 1$. From $(\theta_{t-1}e_t)_{t\geq 2}$ is a sequence of uncorrelated random variables and $\theta_{t-1}$, and $e_t$ are independent, we have

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{T-2}\sum_{t=2}^{T}\theta_{t-1}e_t\right)^2 &= \frac{1}{(T-2)^2}\,\mathbb{E}\left(\sum_{t=2}^{T}\theta_{t-1}^2 e_t^2\right) \\
&\quad + \frac{2}{(T-2)^2}\,\mathbb{E}\left(\sum_{2\leq i<j\leq T}\theta_{i-1}e_i\theta_{j-1}e_j\right) \\
&= \frac{1}{(T-2)^2}\sum_{t=2}^{T}\mathbb{E}\left(\theta_{t-1}^2 e_t^2\right)
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{(T-2)^2}\sum_{t=2}^{T}\mathbb{E}\left(\theta_{t-1}^2\right)\mathbb{E}\left(e_t^2\right) \\
&= \frac{1}{(T-2)^2}\sum_{t=2}^{T} t\sigma^2 D_t^2 \\
&\le \frac{\sigma^2 D_M^2}{(T-2)^2}\frac{(T-1)T}{2} \\
&= O(1). \qquad (3.137)
\end{aligned}$$

Hence, from Corollary 2.62, $\frac{1}{T-2}\sum_{t=2}^{T}\theta_{t-1}e_t = O_p(1)$. Therefore, from Lemma 2.61(3),

$$\frac{2(\rho-\hat{\rho}_\theta)}{T-2}\sum_{t=2}^{T}\theta_{t-1}e_t = o_p(1)O_p(1) = o_p(1). \qquad (3.138)$$

Similarly for the term $\frac{1}{T-2}\sum_{t=2}^{T}\theta_{t-1}e_{t-1}$, from $(\theta_{t-1}e_{t-1})_{t\ge 2}$ is a sequence of uncorrelated random variables and $\theta_{t-1}$, and $e_t$ are independent, we have

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{T-2}\sum_{t=2}^{T}\theta_{t-1}e_{t-1}\right)^2 &= \frac{1}{(T-2)^2}\mathbb{E}\left(\sum_{t=2}^{T}\theta_{t-1}^2 e_{t-1}^2\right) \\
&\quad + \frac{2}{(T-2)^2}\mathbb{E}\left(\sum_{2\le i<j\le T}\theta_{i-1}e_{i-1}\theta_{j-1}e_{j-1}\right) \\
&= \frac{1}{(T-2)^2}\sum_{t=2}^{T}\mathbb{E}\left(\theta_{t-1}^2 e_{t-1}^2\right) \\
&= \frac{1}{(T-2)^2}\sum_{t=2}^{T}\mathbb{E}\left(\theta_{t-1}^2\right)\mathbb{E}\left(e_{t-1}^2\right) \\
&= \frac{1}{(T-2)^2}\sum_{t=2}^{T} t\sigma^2 D_{t-1}^2 \\
&\le \frac{\sigma^2 D_M^2}{(T-2)^2}\frac{(T-1)T}{2} \\
&= O(1). \qquad (3.139)
\end{aligned}$$

Hence, $\frac{1}{T-2}\sum_{t=2}^{T}\theta_{t-1}e_{t-1} = O_p(1)$. Therefore, from Lemma 2.61 (3),

$$\frac{2\hat{\rho}_\theta(\rho-\hat{\rho}_\theta)}{T-2}\sum_{t=2}^{T}\theta_{t-1}e_{t-1} = o_p(1)O_p(1) = o_p(1). \tag{3.140}$$

Consequently, from (3.133)-(3.136), (3.138) (3.140) to (3.132), we can conclude that

$$\frac{2}{T-2}\sum_{t=2}^{T}\left(\eta_t + (\rho-\hat{\rho}_\theta)\theta_{t-1}\right)\left(e_t - \hat{\rho}_\theta e_{t-1}\right) \xrightarrow{p} 0.$$

From (3.130) and (a)-(c), we can conclude that

$$\hat{U}_{4,\theta} - U_{4,\theta} \xrightarrow{p} 0. \tag{3.141}$$

Therefore, from Theorem (3.16),

$$\hat{U}_{4,\theta} = \hat{U}_{4,\theta} - U_{4,\theta} + U_{4,\theta} \xrightarrow{p} \sigma^2. \tag{3.142}$$

□

### 3.3.2 Representation of the Test Statistic

In previous sections, we have obtained the estimators of $\rho$ and $\sigma^2$ under model (3.13). In this section, we investigate the test statistic for stationary test $\tau_\theta$, under hypothesis $\rho = 1$.

Recall from Section 3.1.1 that the sequence of true variables $\{\theta_t\}_{t\geq 1}$ is not a stationary time series if $|\rho| = 1$. The stationary test is the test of the null hypothesis that $|\rho| = 1$ against the alternative hypothesis that $|\rho| < 1$. That is,

$$H_0 : |\rho| = 1 \quad \text{or nonstationary AR(1)},$$

against

$$H_1 : |\rho| < 1 \quad \text{or stationary AR(1)}.$$

However, Dickey [5] considers only the asymptotic distribution of $\rho_\theta$ and the test statistic under the null hypothesis that $\rho = 1$. The asymptotic distribution for the case $\rho = 1$ can be extended to $\rho = -1$. In addition, the asymptotic distribution of $\rho_\theta$ given $\rho = -1$ is equal to the mirror image of the asymptotic distribution of $\rho_\theta$ given $\rho = 1$ ([8]).

**Theorem 3.18** ([5], Theorem 5.3, page 66)**.** *Let $\rho_{\theta,1}$ and $\rho_{\theta,-1}$ be the estimator of $\rho$ on the AR(1) model (3.1) with $\rho = 1$ and $\rho = -1$, respectively. Then, for any real values $a$,*

$$P\left(\rho_{\theta,1} - 1 > a\right) = P\left(\rho_{\theta,-1} + 1 < -a\right).$$

In a similar way, the asymptotic distribution of test statistic $\tau_\theta$ under $\rho = -1$ is identical to the mirror image of the asymptotic distribution of $\tau_\theta$ under $\rho = 1$ ([8]).

The $t$-test statistic, for the null hypothesis that $\rho = 1$, under the AR(1) model (3.8) is

$$\tau_\theta = \frac{\rho_\theta - 1}{S_{\rho_\theta}} = \frac{(\rho_\theta - 1)\sqrt{U_{2,\theta}}}{\sqrt{U_{4,\theta}}}, \tag{3.143}$$

where $S_{\rho_\theta} = \sqrt{U_{4,\theta} U_{2,\theta}^{-1}}$ is the standard deviation of $\rho_\theta$ for $\rho = 1$. The null hypothesis rejected if $\tau_\theta$ is sufficient large.

From the expression of $\tau_\theta$, we can see that $U_{4,\theta}$ and $U_{2,\theta}$ are unknown because the true variables $\{\theta_t\}_{t\geq 1}$ are unobserved. Recall from Theorem 3.10 that $\hat{U}_{2,\theta}$ is an unbiased estimator of $U_{2,\theta}$. In addition, from (3.113) and (3.141), $\hat{\rho}_\theta - \rho_\theta$ and $\hat{U}_{4,\theta} - U_{4,\theta}$ both converge to zero in probability. Hence, we replace $\rho_\theta$, $U_{2,\theta}$, and $U_{4,\theta}$ by their estimates. That is

$$\hat{\tau}_\theta = \frac{(\hat{\rho}_\theta - 1)\sqrt{\hat{U}_{2,\theta}}}{\sqrt{\hat{U}_{4,\theta}}}. \tag{3.144}$$

The next theorem will show that our statistic $\hat{\tau}_\theta$ can be used to represent the test statistic $\tau_\theta$, under the null hypothesis that $\rho = 1$. In addition, $\hat{\tau}_\theta$ has the same asymptotic distribution as $\tau_\theta$.

**Corollary 3.19.** *Under the conditions that $\hat{U}_{2,\theta}$ and $\hat{U}_{4,\theta}$ are both positive, let $\hat{\tau}_\theta$ be the statistic defined in (3.144). Then*

$$\hat{\tau}_\theta \xrightarrow{p} \frac{\left(\sum_{i=1}^{\infty} \sqrt{2}\gamma_i Z_i\right)^2 - 1}{2\sqrt{\sum_{i=1}^{\infty} \gamma_i^2 Z_i^2}},$$

*where $\gamma_i = (-1)^{i+1}\frac{2}{(2i-1)\pi}$ and $Z_i \overset{iid}{\sim} N(0,1)$. In addition, $\hat{\tau}_\theta$ has the same asymptotic distribution as $\tau_\theta$.*

*Proof.* The test statistic $\hat{\tau}_\theta$ can be simplified as follows.

$$\hat{\tau}_\theta = \frac{(\hat{\rho}_\theta - 1)\sqrt{\hat{U}_{2,\theta}}}{\sqrt{\hat{U}_{4,\theta}}} = \frac{T(\hat{\rho}_\theta - 1)\sqrt{\frac{\hat{U}_{2,\theta}}{T^2}}}{\sqrt{\hat{U}_{4,\theta}}}. \tag{3.145}$$

By applying Theorem 2.41 on (3.116) with continuous function $f(x) = \sqrt{x}$, we have

$$\sqrt{\frac{\hat{U}_{2,\theta}}{T^2}} \xrightarrow{p} \sqrt{2\sum_{i=1}^{\infty} \gamma_i^2 Z_i'^2}. \tag{3.146}$$

Hence, by applying Theorem 2.42 on (3.127) and (3.146), we have

$$T(\hat{\rho}_\theta - 1)\sqrt{\frac{\hat{U}_{2,\theta}}{T^2}} \xrightarrow{p} \frac{\left(\sum_{i=1}^{\infty} \sqrt{2}\gamma_i Z_i'\right)^2 - \sigma^2}{\sqrt{2\sum_{i=1}^{\infty} \gamma_i^2 Z_i'^2}}, \tag{3.147}$$

where $\gamma_i = (-1)^{i+1}\frac{2}{(2i-1)\pi}$, and $Z_i' \overset{iid}{\sim} N(0,\sigma^2)$.

From Theorem 3.17, $\hat{U}_{4,\theta}$ is a consistent estimator of $\sigma^2$. By applying Theorem (2.41) with function $\frac{1}{\sqrt{x}}$, we have

$$\frac{1}{\sqrt{\hat{U}_{4,\theta}}} \xrightarrow{p} \frac{1}{\sqrt{\sigma^2}}. \tag{3.148}$$

Hence, from (3.147), we have

$$\hat{\tau}_\theta = \frac{(\hat{\rho}_\theta - 1)\sqrt{\hat{U}_{2,\theta}}}{\sqrt{\hat{U}_{4,\theta}}} \xrightarrow{p} \frac{\left(\sum_{i=1}^{\infty}\sqrt{2}\gamma_i Z_i'\right)^2 - \sigma^2}{\sqrt{2\sigma^2\sum_{i=1}^{\infty}\gamma_i^2 Z_i'^2}} = \frac{\left(\sum_{i=1}^{\infty}\sqrt{2}\gamma_i \frac{Z_i'}{\sigma^2}\right)^2 - 1}{\sqrt{2\sum_{i=1}^{\infty}\gamma_i^2 \frac{Z_i'^2}{\sigma^2}}}. \quad (3.149)$$

Since $Z_i' \overset{iid}{\sim} N(0,\sigma^2)$, $\frac{Z_i'}{\sigma} \overset{iid}{\sim} N(0,1)$. Hence, $\hat{\tau}_\theta$ converges in probability to $\frac{\left(\sum_{i=1}^{\infty}\sqrt{2}\gamma_i Z_i\right)^2 - 1}{2\sqrt{\sum_{i=1}^{\infty}\gamma_i^2 Z_i^2}}$, where $Z_i \overset{iid}{\sim} N(0,1)$. From Theorem 3.2 (4), the asymptotic distribution of $\hat{\tau}_\theta$ and $\tau_\theta$ are the same. □

## 3.4 Positive Adjustments to the Estimators

In previous section, we have obtained the test statistic under the null hypothesis $\rho = 1$ on the model (3.13). This test statistic is well-defined under the condition that $\hat{U}_{2,\theta}$ and $\hat{U}_{4,\theta}$ are both positive. However, this condition is an unnatural setting for general cases. Our adjusted estimator $\hat{U}_{2,\theta}$ and $\hat{U}_{4,\theta}$ can be negative so these two estimators are not well-defined under a square root function. This situation will happen if the sampling variance of $D_t^2$'s are extremely large. In this section, we introduce an adjustment to the variable selection criterion by suggesting strictly positive approximations under the specific conditions.

In 2015, Lahiri and Suntornchost [12] advised an adjustment to the variable selection criterion by the positive approximations $h$-function presented in Chatterjee and Lahiri [4]. The $h$-function is a positive approximation of $x - y$, where $x$ and $y$ are positive. Later, Angkunsit and Suntornchost [2] also introduced the another positive approximations $g$-function. The constructions of the $h$-function and $g$-function are given in the following theorem.

**Lemma 3.20** ([2, 13])**.** *Let* $h : \mathbb{R}^2 \to \mathbb{R}$ *be a function defined as*

$$h(x,y) = \frac{2x}{1+\exp(\frac{2y}{x})}.$$

*Then* $h$ *is a positive approximation of* $x-y$*, where* $x, y > 0$ *and* $y < \frac{\pi}{2}x$*. Moreover, the error in approximation of* $h$*, denoted by* $R_h$*, is*

$$R_h(x,y) = h(x,y) - (x-y) = -x\left(\tanh\left(\frac{y}{x}\right) - \frac{y}{x}\right). \tag{3.150}$$

**Lemma 3.21** ([2])**.** *Let* $g : \mathbb{R}^2 \to \mathbb{R}$ *be a function defined as*

$$g(x,y) = x + \frac{2x^3\left(1-\exp\left(\left(\frac{y}{x}\right)^3\right)\right)}{y^2\left(1+\exp\left(\left(\frac{y}{x}\right)^3\right)\right)}.$$

*Then* $g$ *is a positive approximation of* $x-y$*, where* $0 < y < \sqrt[3]{\pi}x$*. Moreover, the error in approximation of* $g$*, denoted by* $R_g$*, is*

$$R_g(x,y) = g(x,y) - (x-y) = -\frac{2x^3}{y^2}\left(\tanh\left(\frac{1}{2}\left(\frac{y}{x}\right)^3\right) - \frac{1}{2}\left(\frac{y}{x}\right)^3\right). \tag{3.151}$$

The positive approximations of $\hat{U}_{2,\theta}$ by the $h$-function and the $g$-function are as follows.

$$\hat{U}_{2,\theta,h} = h\left(\sum_{t=2}^{T} y_{t-1}^2, \sum_{t=2}^{T} D_{t-1}^2\right) = \frac{2\sum_{t=2}^{T} y_{t-1}^2}{1+\exp\left(\frac{2\sum_{t=2}^{T} D_{t-1}^2}{\sum_{t=2}^{T} y_{t-1}^2}\right)}, \tag{3.152}$$

and

$$\begin{aligned}\hat{U}_{2,\theta,g} &= g\left(\sum_{t=2}^{T} y_{t-1}^2, \sum_{t=2}^{T} D_{t-1}^2\right) \\ &= \left(\sum_{t=2}^{T} y_{t-1}^2\right) + \frac{2\left(\sum_{t=2}^{T} y_{t-1}^2\right)^3\left(1-\exp\left(\left(\frac{\sum_{t=2}^{T} D_{t-1}^2}{\sum_{t=2}^{T} y_{t-1}^2}\right)^3\right)\right)}{\left(\sum_{t=2}^{T} D_{t-1}^2\right)^2\left(1+\exp\left(\left(\frac{\sum_{t=2}^{T} D_{t-1}^2}{\sum_{t=2}^{T} y_{t-1}^2}\right)^3\right)\right)}.\end{aligned} \tag{3.153}$$

Similarly for the estimators $\hat{U}_{4,\theta}$ defined in Theorem 3.17, we define the positive approximation of $\hat{U}_{4,\theta}$ in the following:

$$\hat{U}_{4,\theta,h} = h\left(\frac{1}{T-2}\sum_{t=2}^{T}(y_t - \hat{\rho}_\theta y_{t-1})^2, \frac{1}{T-2}\sum_{t=2}^{T}\left(D_t^2 + \hat{\rho}_\theta^2 D_{t-1}^2\right)\right), \tag{3.154}$$

and

$$\hat{U}_{4,\theta,g} = g\left(\frac{1}{T-2}\sum_{t=2}^{T}(y_t - \hat{\rho}_\theta y_{t-1})^2, \frac{1}{T-2}\sum_{t=2}^{T}\left(D_t^2 + \hat{\rho}_\theta^2 D_{t-1}^2\right)\right). \tag{3.155}$$

However, applying either one of these positive adjustments will introduce bias to the estimates. Therefore, we suggest to apply a positive adjustment only if the unbiased estimator is negative. Since the probability that either $\hat{U}_{2,\theta}$ or $\hat{U}_{4,\theta}$ is negative approachs to zero as $T$ goes to infinity, the new estimates are still consistent estimates. The truncated version of estimator of $U_{2,\theta}$ and $U_{4,\theta}$ as follows.

$$\hat{U}_{2,\theta,htrun} = \begin{cases} \hat{U}_{2,\theta}, & \text{if } \hat{U}_{2,\theta} > 0, \\ \hat{U}_{2,\theta,h}, & \text{otherwise.} \end{cases} \tag{3.156}$$

$$\hat{U}_{2,\theta,gtrun} = \begin{cases} \hat{U}_{2,\theta}, & \text{if } \hat{U}_{2,\theta} > 0, \\ \hat{U}_{2,\theta,g}, & \text{otherwise.} \end{cases} \tag{3.157}$$

$$\hat{U}_{4,\theta,htrun} = \begin{cases} \hat{U}_{4,\theta}, & \text{if } \hat{U}_{4,\theta} > 0, \\ \hat{U}_{4,\theta,h}, & \text{otherwise.} \end{cases} \tag{3.158}$$

$$\hat{U}_{4,\theta,gtrun} = \begin{cases} \hat{U}_{4,\theta}, & \text{if } \hat{U}_{4,\theta} > 0, \\ \hat{U}_{4,\theta,g}, & \text{otherwise.} \end{cases} \tag{3.159}$$

For the test statistic $\hat{\tau}_\theta$ defined in (3.144). If either $\hat{U}_{2,\theta}$ or $\hat{U}_{4,\theta}$ negative, we substitute all negative estimators in $\hat{\tau}_\theta$ with the positive approximation defined in (3.156)-(3.159).

## 3.5 Numerical Simulation

In this section, we investigate the performances of new estimators and test statistic comparing to the naive estimators. Recall that the first order autoregressive model subject to sampling errors is

$$\begin{aligned} y_t &= \theta_t + e_t, \qquad e_t \overset{ind}{\sim} N(0, D_t^2), \\ \theta_t &= \rho\theta_{t-1} + \eta_t, \quad \eta_t \overset{iid}{\sim} N(0, \sigma^2), \end{aligned} \tag{3.160}$$

where $y_0 = \theta_0 = 0$ and $e_t$, and $\eta_t$ are independent normal random variables with zero mean and variances $\sigma^2$ and $D_t^2$, respectively. For the simulation experiment, one thousand samples of size $T = 10000$ were generated for different values of $\rho$ and variances of $e_t$ in (3.160). The variance $\sigma^2$ is set to be 1 throughout our numerical study. The method for generating data is as follows.

(1) Generate the variables of interest $\{\theta_t\}$ from the model

$$\theta_t = \rho\theta_{t-1} + \eta_t,$$

where $\eta_t \sim N(0, 1)$.

(2) Generate the observed variables $\{y_t\}$ from the model

$$y_t = \theta_t + e_t,$$

where $e_t \sim N(0, D_t^2)$.



### 1.) Autoregressive Coefficient

In this section, we consider the performances of all estimators of autoregressive coefficient $\rho$ in two aspects (1) the asymptotic bias and (2) the relationship between sampling variances and the variance of white noises. We consider the cases where $\rho = 0.4$ and 0.7, respectively. For the values of sampling variances of $e_t$, we study different settings of $D_t^2$ as follows.

Case 1. The variance of $\{e_t\}$ is a constant not depending on time $t$. We perform simulation in three settings.

(1.1) $D_t^2 = 0.5$,

(1.2) $D_t^2 = 1$,

(1.3) $D_t^2 = 1.5$.

Case 2. The variance of the sequence $\{e_t\}$ follows a uniform distribution. We perform simulation in three settings.

(2.1) $D_t^2 \sim Uni(0.25, 0.75)$,

(2.2) $D_t^2 \sim Uni(0.75, 1.25)$,

(2.3) $D_t^2 \sim Uni(1.25, 1.75)$,

where $Uni(a, b)$ is the continuous uniform distribution over interval $(a, b)$.

The results shown in the following tables and figures are presented using the following notations, some notations are followed from Theorem 3.10, (3.19), (3.152)-(3.153), and (3.156)-(3.157).

(1) $\rho_{true} = U_{1,\theta}U_{2,\theta}^{-1}$, the Dickey-Fuller estimator or estimator of $\rho$ based on the variables of interest $\{\theta_t\}$,

(2) $\rho_{naive} = U_{1,y}U_{2,y}^{-1}$, the estimator of $\rho$ based on observed variables $\{y_t\}$ and ignore sampling errors,

(3) $\rho_{hat} = \hat{U}_{1,\theta}\hat{U}_{2,\theta}^{-1}$, an adjustment to naive estimator of $\rho$ that the denominator can be negative,

(4) $\rho_{hat,h} = \hat{U}_{1,\theta}\hat{U}_{2,\theta,h}^{-1}$, an adjustment to naive estimator of $\rho$ by the $h$-function,

(5) $\rho_{hat,g} = \hat{U}_{1,\theta}\hat{U}_{2,\theta,g}^{-1}$, an adjustment to naive estimator of $\rho$ by the $g$-function,

(6) $\rho_{hat,htrun} = \hat{U}_{1,\theta}\hat{U}_{2,\theta,htrun}^{-1}$, a truncation version of the $h$-function approximation,

(7) $\rho_{hat,gtrun} = \hat{U}_{1,\theta}\hat{U}_{2,\theta,gtrun}^{-1}$, a truncation version of the $g$-function approximation.

Tables 3.1-3.4 and Figures 3.2 - 3.5 show comparisons between all estimates of $\rho$ in terms of sampling variances relative to the variance of white noises ($\sigma^2 = 1$). We consider two different settings of sampling variances $D_t^2$: (1) the sampling variances are constant, and (2) the sampling variances are generated from a uniform distribution.

| | Percentile | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 10 | 25 | 50 | 75 | 90 | 100 |
| **Case** $\rho = 0.4,\ D_t^2 = 0.5$ | | | | | | | |
| $\rho_{true}$ | 0.3782 | 0.3873 | 0.3932 | 0.3992 | 0.4059 | 0.4117 | 0.4296 |
| $\rho_{naive}$ | 0.2566 | 0.2686 | 0.2747 | 0.2813 | 0.2883 | 0.2945 | 0.3140 |
| $\rho_{hat}$ | 0.3658 | 0.3823 | 0.3909 | 0.3997 | 0.4088 | 0.4175 | 0.4434 |
| $\rho_{hat,h}$ | 0.3614 | 0.3779 | 0.3863 | 0.3949 | 0.4041 | 0.4128 | 0.4384 |
| $\rho_{hat,g}$ | 0.3658 | 0.3823 | 0.3909 | 0.3997 | 0.4088 | 0.4175 | 0.4434 |
| $\rho_{hat,htrun}$ | 0.3658 | 0.3823 | 0.3909 | 0.3997 | 0.4088 | 0.4175 | 0.4434 |
| $\rho_{hat,gtrun}$ | 0.3658 | 0.3823 | 0.3909 | 0.3997 | 0.4088 | 0.4175 | 0.4434 |
| **Case** $\rho = 0.4,\ D_t^2 = 1$ | | | | | | | |
| $\rho_{true}$ | 0.3762 | 0.3881 | 0.3939 | 0.4000 | 0.4064 | 0.4127 | 0.4293 |
| $\rho_{naive}$ | 0.1936 | 0.2041 | 0.2101 | 0.2171 | 0.2247 | 0.2301 | 0.2467 |
| $\rho_{hat}$ | 0.3568 | 0.3775 | 0.3875 | 0.3995 | 0.4119 | 0.4228 | 0.4541 |
| $\rho_{hat,h}$ | 0.3383 | 0.3579 | 0.3674 | 0.3789 | 0.3917 | 0.4012 | 0.4295 |
| $\rho_{hat,g}$ | 0.3566 | 0.3772 | 0.3872 | 0.3993 | 0.4117 | 0.4226 | 0.4538 |
| $\rho_{hat,htrun}$ | 0.3568 | 0.3775 | 0.3875 | 0.3995 | 0.4119 | 0.4228 | 0.4541 |
| $\rho_{hat,gtrun}$ | 0.3568 | 0.3775 | 0.3875 | 0.3995 | 0.4119 | 0.4228 | 0.4541 |
| **Case** $\rho = 0.4,\ D_t^2 = 1.5$ | | | | | | | |
| $\rho_{true}$ | 0.3766 | 0.3881 | 0.3940 | 0.4004 | 0.4063 | 0.4116 | 0.4263 |
| $\rho_{naive}$ | 0.1553 | 0.1646 | 0.1699 | 0.1773 | 0.1841 | 0.1903 | 0.2106 |
| $\rho_{hat}$ | 0.3523 | 0.3719 | 0.3844 | 0.3999 | 0.4164 | 0.4301 | 0.4692 |
| $\rho_{hat,h}$ | 0.3150 | 0.3333 | 0.3443 | 0.3582 | 0.3730 | 0.3843 | 0.4198 |
| $\rho_{hat,g}$ | 0.3512 | 0.3708 | 0.3832 | 0.3986 | 0.4151 | 0.4287 | 0.4675 |
| $\rho_{hat,htrun}$ | 0.3523 | 0.3719 | 0.3844 | 0.3999 | 0.4164 | 0.4301 | 0.4692 |
| $\rho_{hat,gtrun}$ | 0.3523 | 0.3719 | 0.3844 | 0.3999 | 0.4164 | 0.4301 | 0.4692 |

Table 3.1: Comparisons of different estimators of $\rho$ varying by different values of the constant $D_t^2$'s for the case $\rho = 0.4$, $\sigma^2 = 1$, and $T = 10000$.

| | Percentile | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 10 | 25 | 50 | 75 | 90 | 100 |
| **Case** $\rho = 0.4,\ D_t^2 \sim Uni(0.25, 0.75)$ | | | | | | | |
| $\rho_{true}$ | 0.3792 | 0.3884 | 0.3936 | 0.3998 | 0.4054 | 0.4103 | 0.4336 |
| $\rho_{naive}$ | 0.2589 | 0.2692 | 0.2746 | 0.2808 | 0.2876 | 0.2937 | 0.3173 |
| $\rho_{hat}$ | 0.3691 | 0.3834 | 0.3907 | 0.3988 | 0.4078 | 0.4159 | 0.4449 |
| $\rho_{hat,h}$ | 0.3647 | 0.3789 | 0.3861 | 0.3940 | 0.4030 | 0.4109 | 0.4402 |
| $\rho_{hat,g}$ | 0.3691 | 0.3834 | 0.3907 | 0.3988 | 0.4078 | 0.4159 | 0.4449 |
| $\rho_{hat,htrun}$ | 0.3691 | 0.3834 | 0.3907 | 0.3988 | 0.4078 | 0.4159 | 0.4449 |
| $\rho_{hat,gtrun}$ | 0.3691 | 0.3834 | 0.3907 | 0.3988 | 0.4078 | 0.4159 | 0.4449 |
| **Case** $\rho = 0.4,\ D_t^2 \sim Uni(0.75, 1.25)$ | | | | | | | |
| $\rho_{true}$ | 0.3794 | 0.3886 | 0.3942 | 0.4001 | 0.4061 | 0.4110 | 0.4297 |
| $\rho_{naive}$ | 0.1942 | 0.2051 | 0.2109 | 0.2175 | 0.2246 | 0.2299 | 0.2515 |
| $\rho_{hat}$ | 0.3606 | 0.3782 | 0.3890 | 0.4009 | 0.4126 | 0.4215 | 0.4616 |
| $\rho_{hat,h}$ | 0.3412 | 0.3588 | 0.3689 | 0.3800 | 0.3911 | 0.4000 | 0.4383 |
| $\rho_{hat,g}$ | 0.3604 | 0.3780 | 0.3887 | 0.4006 | 0.4123 | 0.4212 | 0.4614 |
| $\rho_{hat,htrun}$ | 0.3606 | 0.3782 | 0.3890 | 0.4009 | 0.4126 | 0.4215 | 0.4616 |
| $\rho_{hat,gtrun}$ | 0.3606 | 0.3782 | 0.3890 | 0.4009 | 0.4126 | 0.4215 | 0.4616 |
| **Case** $\rho = 0.4,\ D_t^2 \sim Uni(1.25, 1.75)$ | | | | | | | |
| $\rho_{true}$ | 0.3794 | 0.3877 | 0.3940 | 0.4002 | 0.4062 | 0.4113 | 0.4234 |
| $\rho_{naive}$ | 0.1513 | 0.1637 | 0.1702 | 0.1774 | 0.1836 | 0.1900 | 0.2162 |
| $\rho_{hat}$ | 0.3441 | 0.3716 | 0.3863 | 0.4004 | 0.4138 | 0.4282 | 0.4879 |
| $\rho_{hat,h}$ | 0.3065 | 0.3322 | 0.3457 | 0.3591 | 0.3709 | 0.3847 | 0.4373 |
| $\rho_{hat,g}$ | 0.3431 | 0.3703 | 0.3850 | 0.3992 | 0.4125 | 0.4270 | 0.4864 |
| $\rho_{hat,htrun}$ | 0.3441 | 0.3716 | 0.3863 | 0.4004 | 0.4138 | 0.4282 | 0.4879 |
| $\rho_{hat,gtrun}$ | 0.3441 | 0.3716 | 0.3863 | 0.4004 | 0.4138 | 0.4282 | 0.4879 |

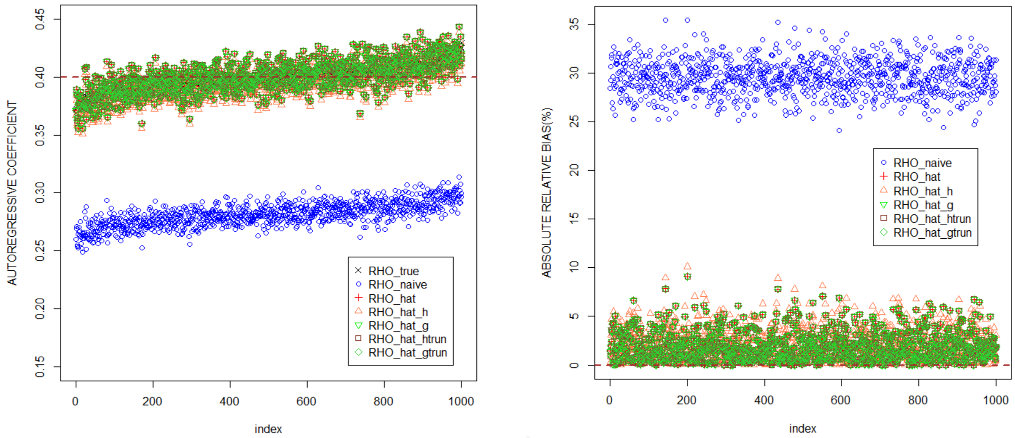
Table 3.2: Comparisons of different estimators of $\rho$ varying by different values of the uniformly distributed $D_t^2$'s for the case $\rho = 0.4$, $\sigma^2 = 1$, and $T = 10000$.

| | Percentile | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 10 | 25 | 50 | 75 | 90 | 100 |
| **Case** $\rho = 0.7,\ D_t^2 = 0.5$ | | | | | | | |
| $\rho_{true}$ | 0.6828 | 0.6904 | 0.6953 | 0.6997 | 0.7045 | 0.7084 | 0.7216 |
| $\rho_{naive}$ | 0.5362 | 0.5458 | 0.5511 | 0.5572 | 0.5640 | 0.5693 | 0.5898 |
| $\rho_{hat}$ | 0.6771 | 0.6883 | 0.6932 | 0.6996 | 0.7060 | 0.7118 | 0.7318 |
| $\rho_{hat,h}$ | 0.6745 | 0.6858 | 0.6907 | 0.6972 | 0.7036 | 0.7093 | 0.7297 |
| $\rho_{hat,g}$ | 0.6771 | 0.6883 | 0.6932 | 0.6996 | 0.7060 | 0.7118 | 0.7318 |
| $\rho_{hat,htrun}$ | 0.6771 | 0.6883 | 0.6932 | 0.6996 | 0.7060 | 0.7118 | 0.7318 |
| $\rho_{hat,gtrun}$ | 0.6771 | 0.6883 | 0.6932 | 0.6996 | 0.7060 | 0.7118 | 0.7318 |
| **Case** $\rho = 0.7,\ D_t^2 = 1$ | | | | | | | |
| $\rho_{true}$ | 0.6835 | 0.6913 | 0.6955 | 0.7003 | 0.7052 | 0.7092 | 0.7218 |
| $\rho_{naive}$ | 0.4404 | 0.4507 | 0.4571 | 0.4644 | 0.4713 | 0.4773 | 0.4912 |
| $\rho_{hat}$ | 0.6716 | 0.6843 | 0.6921 | 0.7002 | 0.7091 | 0.7156 | 0.7369 |
| $\rho_{hat,h}$ | 0.6591 | 0.6714 | 0.6794 | 0.6878 | 0.6961 | 0.7028 | 0.7234 |
| $\rho_{hat,g}$ | 0.6716 | 0.6842 | 0.6920 | 0.7002 | 0.7090 | 0.7155 | 0.7368 |
| $\rho_{hat,htrun}$ | 0.6716 | 0.6843 | 0.6921 | 0.7002 | 0.7091 | 0.7156 | 0.7369 |
| $\rho_{hat,gtrun}$ | 0.6716 | 0.6843 | 0.6921 | 0.7002 | 0.7091 | 0.7156 | 0.7369 |
| **Case** $\rho = 0.7,\ D_t^2 = 1.5$ | | | | | | | |
| $\rho_{true}$ | 0.6843 | 0.6908 | 0.6947 | 0.6999 | 0.7050 | 0.7090 | 0.7201 |
| $\rho_{naive}$ | 0.3714 | 0.3830 | 0.3894 | 0.3967 | 0.4038 | 0.4106 | 0.4327 |
| $\rho_{hat}$ | 0.6634 | 0.6804 | 0.6894 | 0.7007 | 0.7101 | 0.7197 | 0.7439 |
| $\rho_{hat,h}$ | 0.6339 | 0.6508 | 0.6599 | 0.6705 | 0.6799 | 0.6894 | 0.7158 |
| $\rho_{hat,g}$ | 0.6631 | 0.6801 | 0.6891 | 0.7004 | 0.7099 | 0.7194 | 0.7437 |
| $\rho_{hat,htrun}$ | 0.6634 | 0.6804 | 0.6894 | 0.7007 | 0.7101 | 0.7197 | 0.7439 |
| $\rho_{hat,gtrun}$ | 0.6634 | 0.6804 | 0.6894 | 0.7007 | 0.7101 | 0.7197 | 0.7439 |

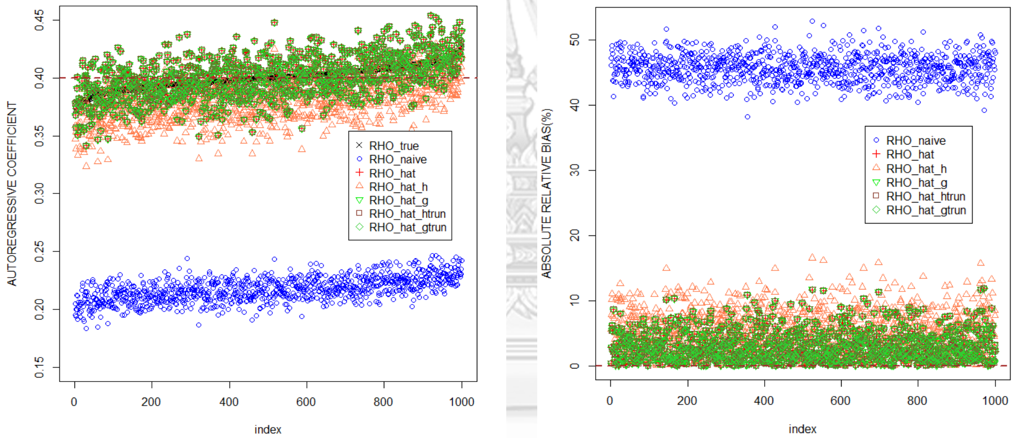
Table 3.3: Comparisons of different estimators of $\rho$ varying by different values of the constant $D_t^2$'s for the case $\rho = 0.7$, $\sigma^2 = 1$, and $T = 10000$.

| | Percentile | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 10 | 25 | 50 | 75 | 90 | 100 |
| **Case** $\rho = 0.7,\ D_t^2 \sim Uni(0.25, 0.75)$ | | | | | | | |
| $\rho_{true}$ | 0.6838 | 0.6901 | 0.6948 | 0.7000 | 0.7048 | 0.7085 | 0.7224 |
| $\rho_{naive}$ | 0.5362 | 0.5449 | 0.5510 | 0.5574 | 0.5640 | 0.5697 | 0.5882 |
| $\rho_{hat}$ | 0.6792 | 0.6873 | 0.6933 | 0.6999 | 0.7063 | 0.7117 | 0.7348 |
| $\rho_{hat,h}$ | 0.6768 | 0.6847 | 0.6909 | 0.6975 | 0.7039 | 0.7094 | 0.7324 |
| $\rho_{hat,g}$ | 0.6792 | 0.6873 | 0.6933 | 0.6999 | 0.7063 | 0.7117 | 0.7348 |
| $\rho_{hat,htrun}$ | 0.6792 | 0.6873 | 0.6933 | 0.6999 | 0.7063 | 0.7117 | 0.7348 |
| $\rho_{hat,gtrun}$ | 0.6792 | 0.6873 | 0.6933 | 0.6999 | 0.7063 | 0.7117 | 0.7348 |
| **Case** $\rho = 0.7,\ D_t^2 \sim Uni(0.75, 1.25)$ | | | | | | | |
| $\rho_{true}$ | 0.6835 | 0.6903 | 0.6952 | 0.6998 | 0.7045 | 0.7089 | 0.7242 |
| $\rho_{naive}$ | 0.4401 | 0.4500 | 0.4565 | 0.4642 | 0.4708 | 0.4760 | 0.4942 |
| $\rho_{hat}$ | 0.6707 | 0.6832 | 0.6912 | 0.6999 | 0.7084 | 0.7154 | 0.7369 |
| $\rho_{hat,h}$ | 0.6569 | 0.6705 | 0.6783 | 0.6875 | 0.6955 | 0.7026 | 0.7241 |
| $\rho_{hat,g}$ | 0.6706 | 0.6831 | 0.6911 | 0.6999 | 0.7083 | 0.7154 | 0.7368 |
| $\rho_{hat,htrun}$ | 0.6707 | 0.6832 | 0.6912 | 0.6999 | 0.7084 | 0.7154 | 0.7369 |
| $\rho_{hat,gtrun}$ | 0.6707 | 0.6832 | 0.6912 | 0.6999 | 0.7084 | 0.7154 | 0.7369 |
| **Case** $\rho = 0.7,\ D_t^2 \sim Uni(1.25, 1.75)$ | | | | | | | |
| $\rho_{true}$ | 0.6834 | 0.6902 | 0.6953 | 0.7002 | 0.7051 | 0.7089 | 0.7204 |
| $\rho_{naive}$ | 0.3705 | 0.3812 | 0.3887 | 0.3965 | 0.4042 | 0.4111 | 0.4294 |
| $\rho_{hat}$ | 0.6624 | 0.6790 | 0.6894 | 0.6997 | 0.7101 | 0.7206 | 0.7497 |
| $\rho_{hat,h}$ | 0.6333 | 0.6495 | 0.6595 | 0.6700 | 0.6803 | 0.6904 | 0.7167 |
| $\rho_{hat,g}$ | 0.6620 | 0.6787 | 0.6891 | 0.6994 | 0.7098 | 0.7203 | 0.7493 |
| $\rho_{hat,htrun}$ | 0.6624 | 0.6790 | 0.6894 | 0.6997 | 0.7101 | 0.7206 | 0.7497 |
| $\rho_{hat,gtrun}$ | 0.6624 | 0.6790 | 0.6894 | 0.6997 | 0.7101 | 0.7206 | 0.7497 |

Table 3.4: Comparisons of different estimators of $\rho$ varying by different values of the uniformly distributed $D_t^2$'s for the case $\rho = 0.7$, $\sigma^2 = 1$, and $T = 10000$.

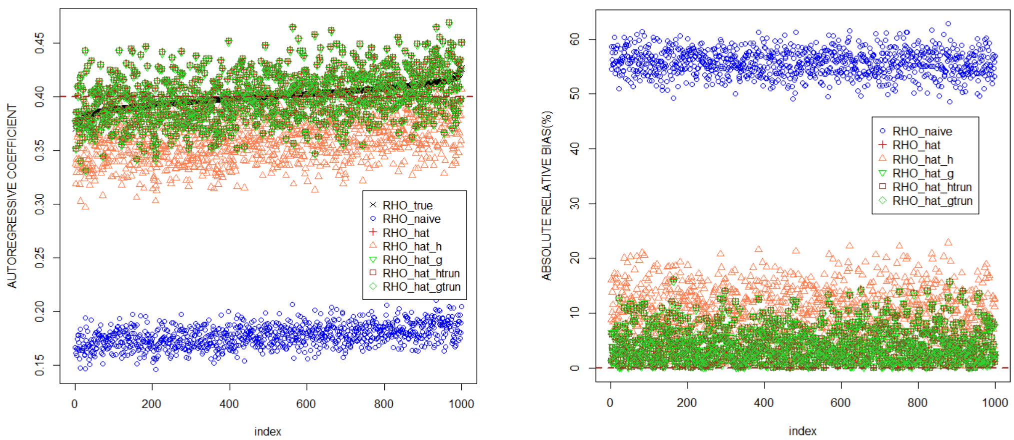



(a) $\rho = 0.4,\ D_t^2 = 0.5$



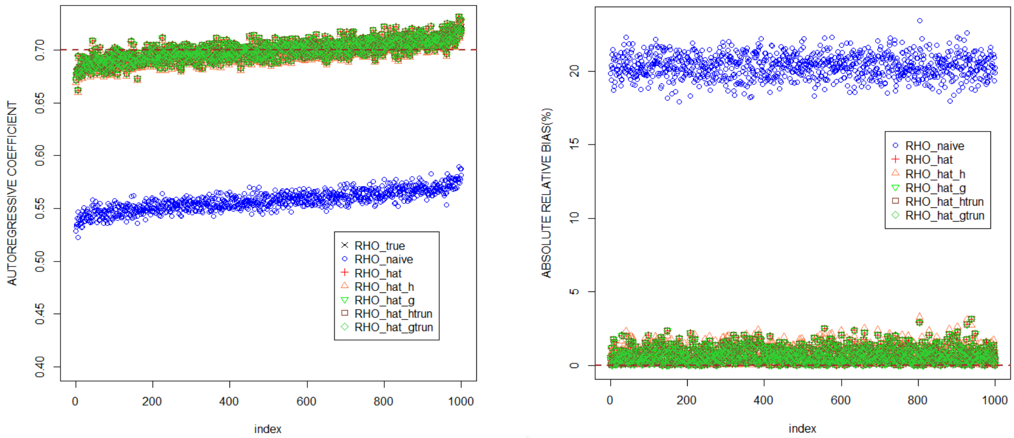

(b) $\rho = 0.4,\ D_t^2 = 1$



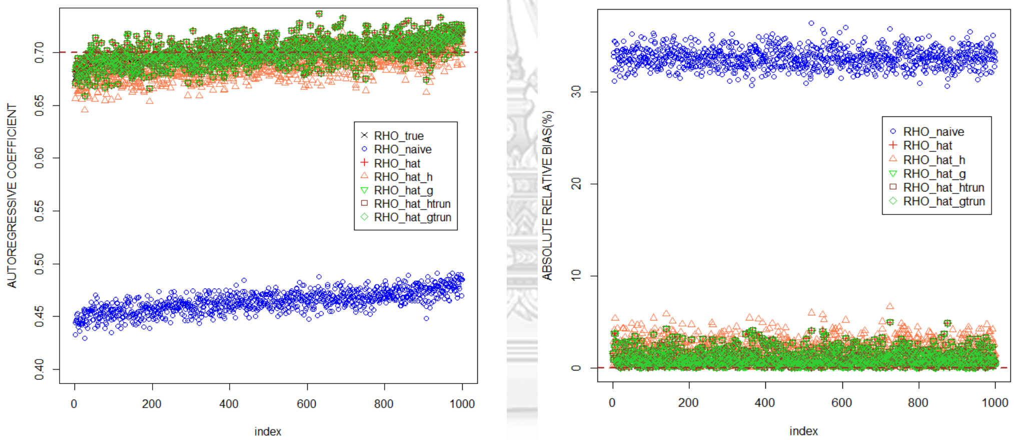



(c) $\rho = 0.4,\ D_t^2 = 1.5$

Figure 3.2: Plot of all different estimator (left) and absolute relative bias (right) of $\rho$ for the case $\rho = 0.4$, $\sigma^2 = 1$, $T = 10000$, and different values of the constant $D_t^2$.



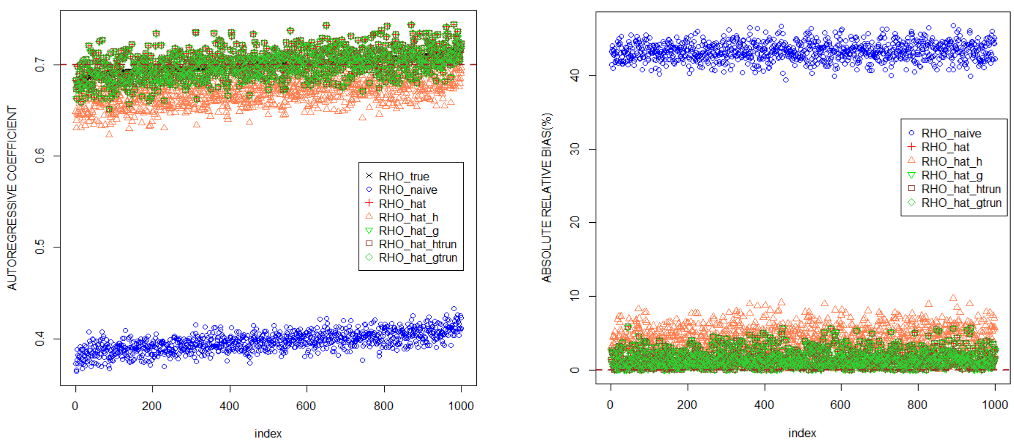

(a) $\rho = 0.7,\ D_t^2 = 0.5$





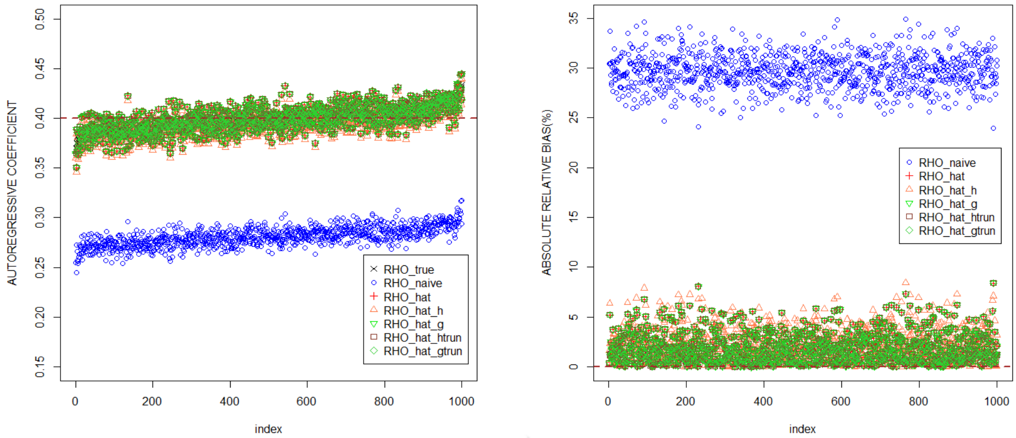

(b) $\rho = 0.7,\ D_t^2 = 1$



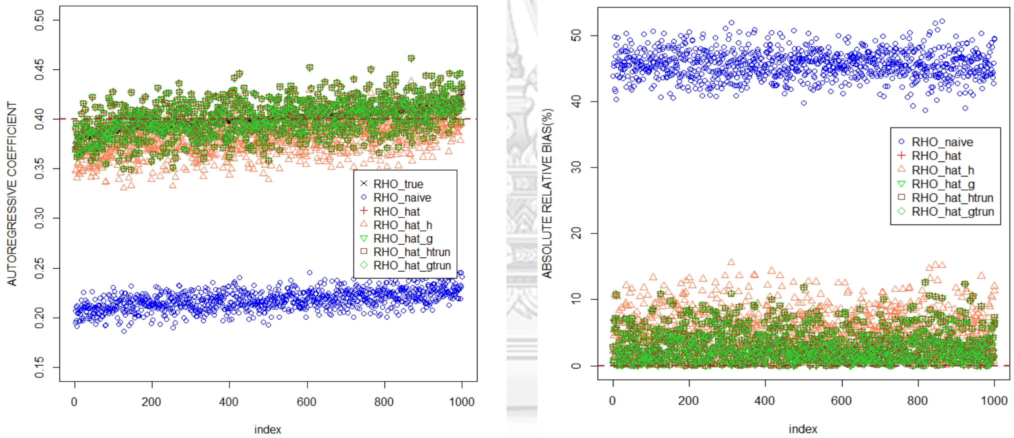



(c) $\rho = 0.7,\ D_t^2 = 1.5$

Figure 3.3: Plot of all different estimator (left) and absolute relative bias (right) of $\rho$ for the case $\rho = 0.7$, $\sigma^2 = 1$, $T = 10000$, and different values of the constant $D_t^2$.

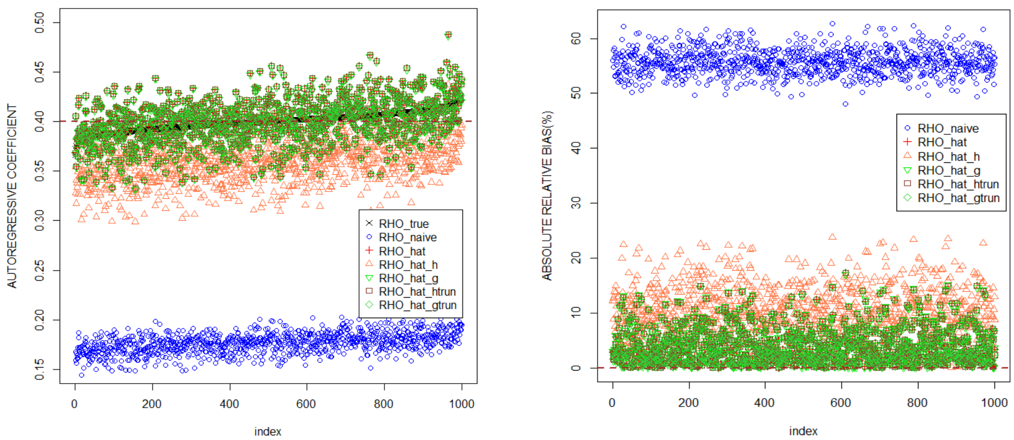



(a) $\rho = 0.4,\ D_t^2 \sim Uni(0.25, 0.75)$



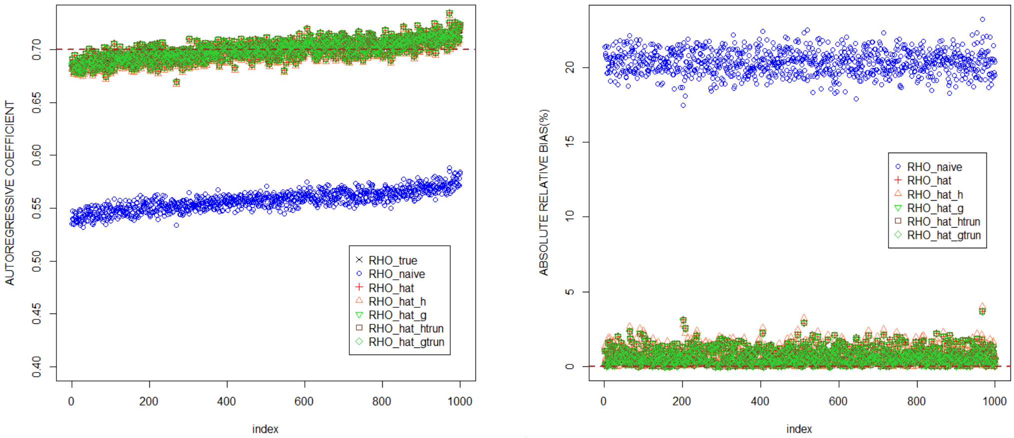

(b) $\rho = 0.4,\ D_t^2 \sim Uni(0.75, 1.25)$



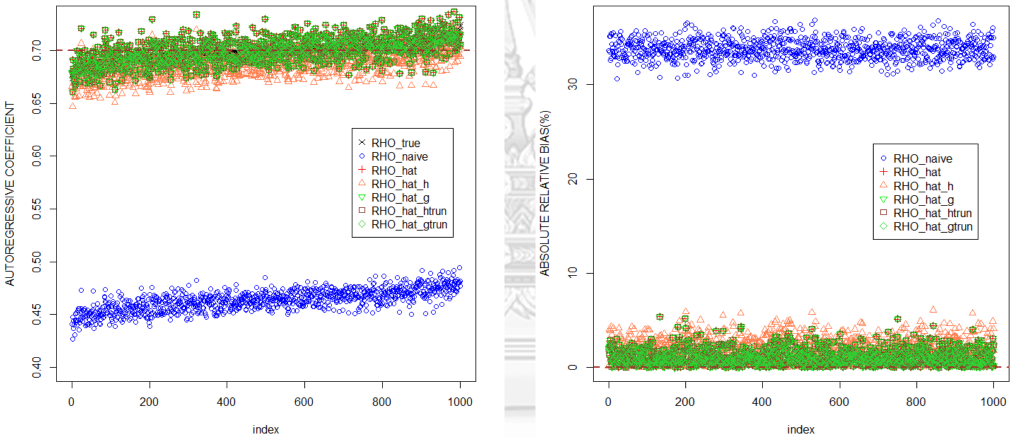



(c) $\rho = 0.4,\ D_t^2 \sim Uni(1.25, 1.75)$

Figure 3.4: Plot of all different estimator (left) and absolute relative bias (right) of $\rho$ for the case $\rho = 0.4$, $\sigma^2 = 1$, $T = 10000$, and different values of the uniformly distributed $D_t^2$.

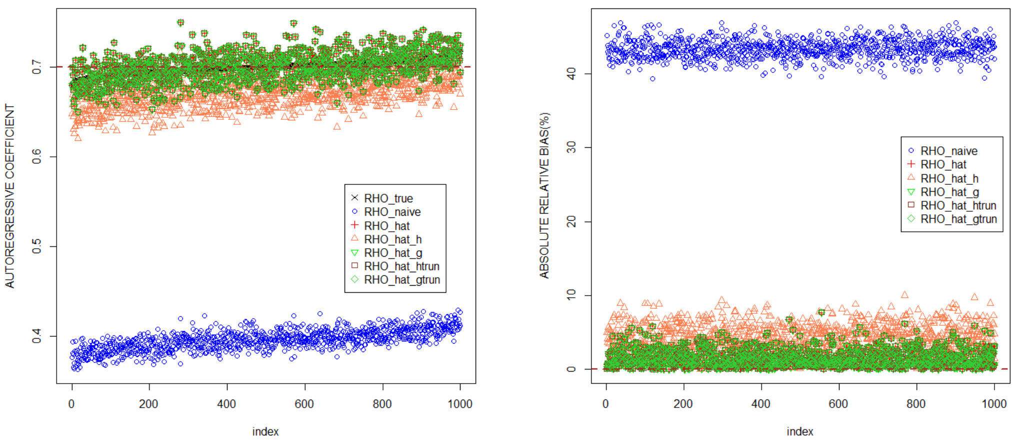



(a) $\rho = 0.7,\ D_t^2 \sim Uni(0.25, 0.75)$




(b) $\rho = 0.7,\ D_t^2 \sim Uni(0.75, 1.25)$






(c) $\rho = 0.7,\ D_t^2 \sim Uni(1.25, 1.75)$

Figure 3.5: Plot of all different estimator (left) and absolute relative bias (right) of $\rho$ for the case $\rho = 0.7$, $\sigma^2 = 1$, $T = 10000$, and different values of the uniformly distributed $D_t^2$

From Tables 3.1-3.4 and Figures 3.2-3.5, we can see that the naive estimator $\rho_{naive}$ always underestimate $\rho_{true}$ for all cases of sampling variances. In addition, the differences between the true values and the naive estimates increase when sampling variance increases.

On the other hand, other adjusted estimates of $\rho$ can reduce the bias terms, and their values are close to the true value estimate $\rho_{true}$ for each cases of sampling variances. However, these adjusted estimators have high variability when the sampling variances are greater than the variance of noise. For example, consider the case $\rho = 0.7$ form Table 3.4, the range of the values of $\rho_{hat}$ is $[0.6707, 0.7369]$ when $D_t^2 \sim Uni(0.75, 1.25)$, but $[0.6624, 0.7497]$ when $D_t^2 \sim Uni(1.25, 1.75)$.

Comparing among values of these adjusted estimators of $\rho$ in Tables 3.1 - 3.4, only the positive adjusted estimates $\rho_{hat,h}$ is slightly underestimated of $\rho_{true}$ when the sampling variances are large. In contrast, the positive adjusted estimates $\rho_{hat,g}$ are close to the estimates $\rho_{hat}$ even when sampling variances are large. However, we will use the adjusted estimates only when $U_{2,\theta}$ is negative. Hence, the truncation under the $g$-function is recommended as the estimate of autoregressive coefficient $\rho$.

To investigate asymptotic behavior of the estimators, we vary the length of time $T$. All estimators of $\rho$ except $\rho_{true}$ are then measured via average absolute bias ($\overline{\text{AB}}$), average absolute relative bias ($\overline{\text{ARB}}$), and average root mean squared error ($\overline{\text{RMSE}}$), defined as the following.

$$\overline{\text{AB}}_\rho(\hat{\rho}) = \frac{1}{1000}\sum_{i=1}^{1000} \left|\rho_{true,i} - \hat{\rho}_i\right|,$$

$$\overline{\text{ARB}}_\rho(\hat{\rho}) = \frac{1}{1000}\sum_{i=1}^{1000} \left|\frac{\rho_{true,i} - \hat{\rho}_i}{\rho_{true,i}}\right|,$$

$$\overline{\text{RMSE}}_\rho(\hat{\rho}) = \sqrt{\frac{1}{1000}\sum_{i=1}^{1000} \left(\rho_{true,i} - \hat{\rho}_i\right)^2},$$

where $\rho_{true,i}$ (for $i = 1, 2, \ldots, 1000$) is the value $\rho_{true}$ from $i$-th simulation and $\hat{\rho}_i$ is the other estimates of $\rho$.

The three measures are presented in Table 3.5-3.16, where all measures are presented in terms of percentage. The corresponding plots are presented in Figure 3.6-3.8.

| Quality Measure | Size Group $T$ | Estimator | | | | | |
|---|---|---|---|---|---|---|---|
| | | $\rho_{naive}$ | $\rho_{hat}$ | $\rho_{hat,h}$ | $\rho_{hat,g}$ | $\rho_{htrun}$ | $\rho_{gtrun}$ |
| **Case** $\rho = 0.4$, $D_t^2 = 0.5$ | | | | | | | |
| $\overline{\text{AB}}_\rho(\%)$ | 50 | 13.196 | 11.849 | 11.548 | 11.845 | 11.849 | 11.849 |
| | 100 | 12.056 | 8.350 | 8.229 | 8.349 | 8.350 | 8.350 |
| | 250 | 11.811 | 5.267 | 5.216 | 5.267 | 5.267 | 5.267 |
| | 500 | 11.691 | 3.575 | 3.527 | 3.575 | 3.575 | 3.575 |
| | 1000 | 11.849 | 2.484 | 2.478 | 2.484 | 2.484 | 2.484 |
| | 2500 | 11.762 | 1.544 | 1.547 | 1.544 | 1.544 | 1.544 |
| | 5000 | 11.805 | 1.123 | 1.160 | 1.123 | 1.123 | 1.123 |
| | 10000 | 11.813 | 0.785 | 0.857 | 0.785 | 0.785 | 0.785 |
| $\overline{\text{ARB}}_\rho(\%)$ | 50 | 39.914 | 41.689 | 40.493 | 41.670 | 41.689 | 41.689 |
| | 100 | 32.044 | 23.572 | 23.198 | 23.570 | 23.572 | 23.572 |
| | 250 | 30.100 | 13.650 | 13.511 | 13.649 | 13.650 | 13.650 |
| | 500 | 29.474 | 9.118 | 8.989 | 9.117 | 9.118 | 9.118 |
| | 1000 | 29.623 | 6.237 | 6.222 | 6.237 | 6.237 | 6.237 |
| | 2500 | 29.377 | 3.866 | 3.877 | 3.866 | 3.866 | 3.866 |
| | 5000 | 29.514 | 2.810 | 2.905 | 2.810 | 2.810 | 2.810 |
| | 10000 | 29.569 | 1.966 | 2.146 | 1.966 | 1.966 | 1.966 |
| $\overline{\text{RMSE}}_\rho(\%)$ | 50 | 15.820 | 15.403 | 14.934 | 15.393 | 15.403 | 15.403 |
| | 100 | 13.987 | 10.537 | 10.381 | 10.536 | 10.537 | 10.537 |
| | 250 | 12.773 | 6.696 | 6.621 | 6.696 | 6.696 | 6.696 |
| | 500 | 12.154 | 4.488 | 4.440 | 4.488 | 4.488 | 4.488 |
| | 1000 | 12.069 | 3.076 | 3.078 | 3.076 | 3.076 | 3.076 |
| | 2500 | 11.846 | 1.936 | 1.943 | 1.936 | 1.936 | 1.936 |
| | 5000 | 11.850 | 1.401 | 1.449 | 1.401 | 1.401 | 1.401 |
| | 10000 | 11.835 | 0.981 | 1.071 | 0.981 | 0.981 | 0.981 |

Table 3.5: Percent average absolute bias ($\overline{\text{AB}}$), percent average absolute relative bias ($\overline{\text{ARB}}$), and percent average root mean squared error ($\overline{\text{RMSE}}$) of all different estimators of $\rho$ for the case $\rho = 0.4$ and $D_t^2 = 0.5$ varying by length of time $T$.

| Quality Measure | Size Group $T$ | Estimator | | | | | |
|---|---|---|---|---|---|---|---|
| | | $\rho_{naive}$ | $\rho_{hat}$ | $\rho_{hat,h}$ | $\rho_{hat,g}$ | $\rho_{htrun}$ | $\rho_{gtrun}$ |
| **Case** $\rho = 0.4,\ D_t^2 = 1$ | | | | | | | |
| $\overline{\text{AB}}_\rho(\%)$ | 50 | 18.5956 | 20.8498 | 17.9551 | 20.2218 | 20.8498 | 20.8498 |
| | 100 | 17.7395 | 13.1450 | 12.2057 | 13.1163 | 13.1450 | 13.1450 |
| | 250 | 17.9433 | 7.9968 | 7.5951 | 7.9853 | 7.9968 | 7.9968 |
| | 500 | 18.2464 | 5.5277 | 5.3720 | 5.5206 | 5.5277 | 5.5277 |
| | 1000 | 18.0763 | 3.9679 | 3.9800 | 3.9627 | 3.9679 | 3.9679 |
| | 2500 | 18.1836 | 2.4711 | 2.8397 | 2.4683 | 2.4711 | 2.4711 |
| | 5000 | 18.2651 | 1.6690 | 2.3947 | 1.6684 | 1.6690 | 1.6690 |
| | 10000 | 18.2618 | 1.2342 | 2.1541 | 1.2334 | 1.2342 | 1.2342 |
| $\overline{\text{ARB}}_\rho(\%)$ | 50 | 57.9181 | 78.3299 | 67.9744 | 76.3763 | 78.3299 | 78.3299 |
| | 100 | 46.6252 | 37.3658 | 34.6463 | 37.2820 | 37.3658 | 37.3658 |
| | 250 | 45.3425 | 20.6605 | 19.6156 | 20.6310 | 20.6605 | 20.6605 |
| | 500 | 45.7559 | 13.9702 | 13.5858 | 13.9524 | 13.9702 | 13.9702 |
| | 1000 | 45.3763 | 9.9935 | 10.0111 | 9.9804 | 9.9935 | 9.9935 |
| | 2500 | 45.4685 | 6.1777 | 7.1099 | 6.1708 | 6.1777 | 6.1777 |
| | 5000 | 45.6824 | 4.1735 | 5.9967 | 4.1723 | 4.1735 | 4.1735 |
| | 10000 | 45.6610 | 3.0898 | 5.3887 | 3.0880 | 3.0898 | 3.0898 |
| $\overline{\text{RMSE}}_\rho(\%)$ | 50 | 21.9693 | 34.0338 | 23.2929 | 28.5757 | 34.0338 | 34.0338 |
| | 100 | 19.8635 | 16.9864 | 15.6511 | 16.9376 | 16.9864 | 16.9864 |
| | 250 | 18.8545 | 10.1295 | 9.5740 | 10.1133 | 10.1295 | 10.1295 |
| | 500 | 18.6878 | 6.9762 | 6.8552 | 6.9686 | 6.9762 | 6.9762 |
| | 1000 | 18.3134 | 5.0136 | 5.0305 | 5.0070 | 5.0136 | 5.0136 |
| | 2500 | 18.2709 | 3.1385 | 3.5263 | 3.1347 | 3.1385 | 3.1385 |
| | 5000 | 18.3065 | 2.1213 | 2.8724 | 2.1198 | 2.1213 | 2.1213 |
| | 10000 | 18.2842 | 1.5349 | 2.5075 | 1.5338 | 1.5349 | 1.5349 |

Table 3.6: Percent average absolute bias ($\overline{\text{AB}}$), percent average absolute relative bias ($\overline{\text{ARB}}$), and percent average root mean squared error ($\overline{\text{RMSE}}$) of all different estimators of $\rho$ for the case $\rho = 0.4$ and $D_t^2 = 1$ varying by length of time $T$.

| Quality Measure | Size Group $T$ | Estimator | | | | | |
|---|---|---|---|---|---|---|---|
| | | $\rho_{naive}$ | $\rho_{hat}$ | $\rho_{hat,h}$ | $\rho_{hat,g}$ | $\rho_{htrun}$ | $\rho_{gtrun}$ |
| **Case** $\rho = 0.4$, $D_t^2 = 1.5$ | | | | | | | |
| $\overline{\text{AB}}_\rho(\%)$ | 50 | 22.3948 | 33.8336 | 22.7192 | 29.4781 | 28.9331 | 30.7957 |
| | 100 | 21.7262 | 20.0998 | 15.7475 | 18.4206 | 18.5997 | 18.7461 |
| | 250 | 21.9727 | 11.0414 | 10.0456 | 10.9549 | 11.0414 | 11.0414 |
| | 500 | 22.1719 | 7.6666 | 7.3880 | 7.6189 | 7.6666 | 7.6666 |
| | 1000 | 22.3043 | 5.5020 | 5.9614 | 5.4769 | 5.5020 | 5.5020 |
| | 2500 | 22.3447 | 3.3482 | 4.6800 | 3.3370 | 3.3482 | 3.3482 |
| | 5000 | 22.3016 | 2.3906 | 4.3030 | 2.3782 | 2.3906 | 2.3906 |
| | 10000 | 22.2903 | 1.6969 | 4.1503 | 1.6917 | 1.6969 | 1.6969 |
| $\overline{\text{ARB}}_\rho(\%)$ | 50 | 62.2634 | 115.6101 | 73.4403 | 101.1043 | 94.8840 | 106.0011 |
| | 100 | 56.6097 | 57.0189 | 43.4967 | 51.1635 | 51.6033 | 52.0998 |
| | 250 | 56.1526 | 28.8708 | 26.2538 | 28.6458 | 28.8708 | 28.8708 |
| | 500 | 56.0147 | 19.5958 | 18.8004 | 19.4711 | 19.5958 | 19.5958 |
| | 1000 | 55.9563 | 13.8574 | 14.9727 | 13.7927 | 13.8574 | 13.8574 |
| | 2500 | 55.9171 | 8.4090 | 11.7277 | 8.3810 | 8.4090 | 8.4090 |
| | 5000 | 55.8264 | 5.9902 | 10.7698 | 5.9589 | 5.9902 | 5.9902 |
| | 10000 | 55.7237 | 4.2445 | 10.3730 | 4.2312 | 4.2445 | 4.2445 |
| $\overline{\text{RMSE}}_\rho(\%)$ | 50 | 25.9091 | 105.1508 | 29.1586 | 47.6093 | 41.6825 | 50.8013 |
| | 100 | 24.0056 | 45.2227 | 19.7739 | 23.7619 | 24.2622 | 24.6852 |
| | 250 | 22.8869 | 14.2000 | 12.6314 | 14.0401 | 14.2000 | 14.2000 |
| | 500 | 22.6407 | 9.7014 | 9.3479 | 9.6335 | 9.7014 | 9.7014 |
| | 1000 | 22.5476 | 6.8136 | 7.3548 | 6.7811 | 6.8136 | 6.8136 |
| | 2500 | 22.4337 | 4.1727 | 5.6124 | 4.1559 | 4.1727 | 4.1727 |
| | 5000 | 22.3501 | 2.9726 | 4.9481 | 2.9627 | 2.9726 | 2.9726 |
| | 10000 | 22.3144 | 2.1114 | 4.5328 | 2.1030 | 2.1114 | 2.1114 |

Table 3.7: Percent average absolute bias ($\overline{\text{AB}}$), percent average absolute relative bias ($\overline{\text{ARB}}$), and percent average root mean squared error ($\overline{\text{RMSE}}$) of all different estimators of $\rho$ for the case $\rho = 0.4$ and $D_t^2 = 1.5$ varying by length of time $T$.

| Quality Measure | Size Group $T$ | Estimator | | | | | |
|---|---|---|---|---|---|---|---|
| | | $\rho_{naive}$ | $\rho_{hat}$ | $\rho_{hat,h}$ | $\rho_{hat,g}$ | $\rho_{htrun}$ | $\rho_{gtrun}$ |
| **Case** $\rho = 0.4$, $D_t^2 \sim Uni(0.25, 0.75)$ | | | | | | | |
| $\overline{\text{AB}}_\rho(\%)$ | 50 | 12.7601 | 11.3700 | 11.0826 | 11.3674 | 11.3700 | 11.3700 |
| | 100 | 11.8462 | 7.7284 | 7.5674 | 7.7276 | 7.7284 | 7.7284 |
| | 250 | 11.7228 | 4.9414 | 4.8729 | 4.9412 | 4.9414 | 4.9414 |
| | 500 | 11.8107 | 3.5386 | 3.5161 | 3.5385 | 3.5386 | 3.5386 |
| | 1000 | 11.8971 | 2.5132 | 2.5311 | 2.5132 | 2.5132 | 2.5132 |
| | 2500 | 11.8427 | 1.5825 | 1.6005 | 1.5824 | 1.5825 | 1.5825 |
| | 5000 | 11.8106 | 1.0951 | 1.1381 | 1.0950 | 1.0951 | 1.0951 |
| | 10000 | 11.8483 | 0.7569 | 0.8510 | 0.7569 | 0.7569 | 0.7569 |
| $\overline{\text{ARB}}_\rho(\%)$ | 50 | 35.2404 | 34.9973 | 33.9998 | 34.9868 | 34.9973 | 34.9973 |
| | 100 | 30.4746 | 21.2131 | 20.7616 | 21.2107 | 21.2131 | 21.2131 |
| | 250 | 29.6697 | 12.7503 | 12.5838 | 12.7497 | 12.7503 | 12.7503 |
| | 500 | 29.7417 | 8.9744 | 8.9228 | 8.9742 | 8.9744 | 8.9744 |
| | 1000 | 29.8598 | 6.3521 | 6.3945 | 6.3520 | 6.3521 | 6.3521 |
| | 2500 | 29.6762 | 3.9760 | 4.0208 | 3.9758 | 3.9760 | 3.9760 |
| | 5000 | 29.5983 | 2.7484 | 2.8566 | 2.7483 | 2.7484 | 2.7484 |
| | 10000 | 29.6560 | 1.8957 | 2.1313 | 1.8958 | 1.8957 | 1.8957 |
| $\overline{\text{RMSE}}_\rho(\%)$ | 50 | 15.5073 | 14.6545 | 14.2749 | 14.6503 | 14.6545 | 14.6545 |
| | 100 | 13.6325 | 9.7061 | 9.5325 | 9.7052 | 9.7061 | 9.7061 |
| | 250 | 12.5596 | 6.1814 | 6.0958 | 6.1811 | 6.1814 | 6.1814 |
| | 500 | 12.2514 | 4.3911 | 4.3635 | 4.3910 | 4.3911 | 4.3911 |
| | 1000 | 12.1271 | 3.1292 | 3.1469 | 3.1292 | 3.1292 | 3.1292 |
| | 2500 | 11.9354 | 1.9814 | 2.0207 | 1.9813 | 1.9814 | 1.9814 |
| | 5000 | 11.8542 | 1.3849 | 1.4397 | 1.3848 | 1.3849 | 1.3849 |
| | 10000 | 11.8699 | 0.9527 | 1.0637 | 0.9527 | 0.9527 | 0.9527 |

Table 3.8: Percent average absolute bias ($\overline{\text{AB}}$), percent average absolute relative bias ($\overline{\text{ARB}}$), and percent average root mean squared error ($\overline{\text{RMSE}}$) of all different estimators of $\rho$ for the case $\rho = 0.4$ and $D_t^2 \sim Uni(0.25, 0.75)$ varying by length of time $T$.

| Quality Measure | Size Group $T$ | Estimator | | | | | |
|---|---|---|---|---|---|---|---|
| | | $\rho_{naive}$ | $\rho_{hat}$ | $\rho_{hat,h}$ | $\rho_{hat,g}$ | $\rho_{htrun}$ | $\rho_{gtrun}$ |
| **Case** $\rho = 0.4$, $D_t^2 \sim Uni(0.75, 1.25)$ | | | | | | | |
| $\overline{\text{AB}}_\rho(\%)$ | 50 | 19.5712 | 20.1129 | 17.7338 | 20.0203 | 19.9907 | 20.2510 |
| | 100 | 18.7165 | 13.0279 | 12.0745 | 12.9871 | 13.0279 | 13.0279 |
| | 250 | 18.3246 | 7.8534 | 7.5509 | 7.8429 | 7.8534 | 7.8534 |
| | 500 | 18.3933 | 5.4180 | 5.3366 | 5.4112 | 5.4180 | 5.4180 |
| | 1000 | 18.3542 | 3.8714 | 4.0590 | 3.8680 | 3.8714 | 3.8714 |
| | 2500 | 18.2654 | 2.5342 | 2.9100 | 2.5311 | 2.5342 | 2.5342 |
| | 5000 | 18.2121 | 1.7378 | 2.3673 | 1.7354 | 1.7378 | 1.7378 |
| | 10000 | 18.2328 | 1.2140 | 2.1065 | 1.2120 | 1.2140 | 1.2140 |
| $\overline{\text{ARB}}_\rho(\%)$ | 50 | 58.1296 | 74.8150 | 64.1680 | 74.8425 | 73.7287 | 76.0426 |
| | 100 | 48.0176 | 35.6953 | 32.9963 | 35.5796 | 35.6953 | 35.6953 |
| | 250 | 45.9722 | 20.2077 | 19.3601 | 20.1793 | 20.2077 | 20.2077 |
| | 500 | 46.0252 | 13.7418 | 13.5018 | 13.7237 | 13.7418 | 13.7418 |
| | 1000 | 45.8720 | 9.7420 | 10.1947 | 9.7336 | 9.7420 | 9.7420 |
| | 2500 | 45.7380 | 6.3670 | 7.3000 | 6.3590 | 6.3670 | 6.3670 |
| | 5000 | 45.5669 | 4.3554 | 5.9282 | 4.3493 | 4.3554 | 4.3554 |
| | 10000 | 45.5844 | 3.0380 | 5.2657 | 3.0328 | 3.0380 | 3.0380 |
| $\overline{\text{RMSE}}_\rho(\%)$ | 50 | 22.5732 | 27.4523 | 22.6933 | 28.0029 | 26.9247 | 28.6702 |
| | 100 | 20.6735 | 16.6150 | 15.2944 | 16.5401 | 16.6150 | 16.6150 |
| | 250 | 19.2138 | 9.9221 | 9.5046 | 9.9068 | 9.9221 | 9.9221 |
| | 500 | 18.8047 | 6.7938 | 6.6966 | 6.7844 | 6.7938 | 6.7938 |
| | 1000 | 18.5782 | 4.9445 | 5.1204 | 4.9402 | 4.9445 | 4.9445 |
| | 2500 | 18.3581 | 3.1706 | 3.6175 | 3.1674 | 3.1706 | 3.1706 |
| | 5000 | 18.2578 | 2.2435 | 2.8695 | 2.2402 | 2.2435 | 2.2435 |
| | 10000 | 18.2550 | 1.5402 | 2.4573 | 1.5378 | 1.5402 | 1.5402 |

Table 3.9: Percent average absolute bias ($\overline{\text{AB}}$), percent average absolute relative bias ($\overline{\text{ARB}}$), and percent average root mean squared error ($\overline{\text{RMSE}}$) of all different estimators of $\rho$ for the case $\rho = 0.4$ and $D_t^2 \sim Uni(0.75, 1.25)$ varying by length of time $T$.

| Quality Measure | Size Group $T$ | Estimator | | | | | |
|---|---|---|---|---|---|---|---|
| | | $\rho_{naive}$ | $\rho_{hat}$ | $\rho_{hat,h}$ | $\rho_{hat,g}$ | $\rho_{htrun}$ | $\rho_{gtrun}$ |
| **Case** $\rho = 0.4$, $D_t^2 \sim Uni(1.25, 1.75)$ | | | | | | | |
| $\overline{\text{AB}}_\rho(\%)$ | 50 | 21.6036 | 32.8251 | 21.8282 | 29.8036 | 29.3024 | 31.9217 |
| | 100 | 21.4317 | 17.8818 | 14.5661 | 17.4496 | 17.8818 | 17.8818 |
| | 250 | 22.0191 | 10.4730 | 9.7439 | 10.4030 | 10.4730 | 10.4730 |
| | 500 | 22.1768 | 7.5952 | 7.4512 | 7.5537 | 7.5952 | 7.5952 |
| | 1000 | 22.1247 | 5.1817 | 5.6525 | 5.1521 | 5.1817 | 5.1817 |
| | 2500 | 22.2159 | 3.2534 | 4.5840 | 3.2387 | 3.2534 | 3.2534 |
| | 5000 | 22.2702 | 2.2612 | 4.2286 | 2.2485 | 2.2612 | 2.2612 |
| | 10000 | 22.2644 | 1.5734 | 4.1258 | 1.5672 | 1.5734 | 1.5734 |
| $\overline{\text{ARB}}_\rho(\%)$ | 50 | 60.5304 | 107.8606 | 71.8233 | 99.1004 | 96.9785 | 106.1050 |
| | 100 | 55.7268 | 50.9047 | 40.9392 | 49.5256 | 50.9047 | 50.9047 |
| | 250 | 56.3069 | 27.4730 | 25.5495 | 27.2856 | 27.4730 | 27.4730 |
| | 500 | 56.2194 | 19.5283 | 19.0869 | 19.4186 | 19.5283 | 19.5283 |
| | 1000 | 55.6842 | 13.1467 | 14.3083 | 13.0712 | 13.1467 | 13.1467 |
| | 2500 | 55.6541 | 8.1591 | 11.4886 | 8.1223 | 8.1591 | 8.1591 |
| | 5000 | 55.7428 | 5.6631 | 10.5860 | 5.6309 | 5.6631 | 5.6631 |
| | 10000 | 55.6929 | 3.9386 | 10.3245 | 3.9231 | 3.9386 | 3.9386 |
| $\overline{\text{RMSE}}_\rho(\%)$ | 50 | 24.9895 | 87.2364 | 28.9711 | 58.6614 | 52.6000 | 68.6579 |
| | 100 | 23.4657 | 23.9078 | 18.6928 | 22.9033 | 23.9078 | 23.9078 |
| | 250 | 22.8498 | 13.5859 | 12.2469 | 13.4530 | 13.5859 | 13.5859 |
| | 500 | 22.6176 | 9.4726 | 9.2971 | 9.4171 | 9.4726 | 9.4726 |
| | 1000 | 22.3390 | 6.5041 | 6.9564 | 6.4656 | 6.5041 | 6.5041 |
| | 2500 | 22.3066 | 4.1054 | 5.4607 | 4.0857 | 4.1054 | 4.1054 |
| | 5000 | 22.3136 | 2.8471 | 4.8362 | 2.8343 | 2.8471 | 2.8471 |
| | 10000 | 22.2847 | 1.9846 | 4.4624 | 1.9750 | 1.9846 | 1.9846 |

Table 3.10: Percent average absolute bias ($\overline{\text{AB}}$), percent average absolute relative bias ($\overline{\text{ARB}}$), and percent average root mean squared error ($\overline{\text{RMSE}}$) of all different estimators of $\rho$ for the case $\rho = 0.4$ and $D_t^2 \sim Uni(1.25, 1.75)$ varying by length of time $T$.

| Quality Measure | Size Group $T$ | Estimator | | | | | |
|---|---|---|---|---|---|---|---|
| | | $\rho_{naive}$ | $\rho_{hat}$ | $\rho_{hat,h}$ | $\rho_{hat,g}$ | $\rho_{htrun}$ | $\rho_{gtrun}$ |
| **Case** $\rho = 0.7$, $D_t^2 = 0.5$ | | | | | | | |
| $\overline{\text{AB}}_\rho(\%)$ | 50 | 14.7645 | 7.7053 | 7.5985 | 7.7050 | 7.7053 | 7.7053 |
| | 100 | 14.5199 | 5.2627 | 5.2296 | 5.2626 | 5.2627 | 5.2627 |
| | 250 | 14.3579 | 3.2586 | 3.2508 | 3.2586 | 3.2586 | 3.2586 |
| | 500 | 14.2938 | 2.2654 | 2.2629 | 2.2654 | 2.2654 | 2.2654 |
| | 1000 | 14.2622 | 1.6186 | 1.6228 | 1.6186 | 1.6186 | 1.6186 |
| | 2500 | 14.2105 | 1.0240 | 1.0333 | 1.0240 | 1.0240 | 1.0240 |
| | 5000 | 14.2173 | 0.7035 | 0.7221 | 0.7035 | 0.7035 | 0.7035 |
| | 10000 | 14.2281 | 0.4893 | 0.5278 | 0.4893 | 0.4893 | 0.4893 |
| $\overline{\text{ARB}}_\rho(\%)$ | 50 | 22.8455 | 12.3418 | 12.1646 | 12.3411 | 12.3418 | 12.3418 |
| | 100 | 21.5755 | 7.9241 | 7.8739 | 7.9240 | 7.9241 | 7.9241 |
| | 250 | 20.8985 | 4.7808 | 4.7695 | 4.7808 | 4.7808 | 4.7808 |
| | 500 | 20.5742 | 3.2676 | 3.2649 | 3.2676 | 3.2676 | 3.2676 |
| | 1000 | 20.4508 | 2.3258 | 2.3317 | 2.3258 | 2.3258 | 2.3258 |
| | 2500 | 20.3473 | 1.4674 | 1.4804 | 1.4674 | 1.4674 | 1.4674 |
| | 5000 | 20.3337 | 1.0061 | 1.0326 | 1.0061 | 1.0061 | 1.0061 |
| | 10000 | 20.3368 | 0.6993 | 0.7544 | 0.6993 | 0.6993 | 0.6993 |
| $\overline{\text{RMSE}}_\rho(\%)$ | 50 | 16.6124 | 10.1094 | 9.9939 | 10.1089 | 10.1094 | 10.1094 |
| | 100 | 15.5048 | 6.7846 | 6.7465 | 6.7845 | 6.7846 | 6.7846 |
| | 250 | 14.7688 | 4.1668 | 4.1578 | 4.1668 | 4.1668 | 4.1668 |
| | 500 | 14.4962 | 2.8544 | 2.8606 | 2.8544 | 2.8544 | 2.8544 |
| | 1000 | 14.3635 | 2.0213 | 2.0305 | 2.0213 | 2.0213 | 2.0213 |
| | 2500 | 14.2477 | 1.2617 | 1.2720 | 1.2617 | 1.2617 | 1.2617 |
| | 5000 | 14.2363 | 0.8857 | 0.9119 | 0.8857 | 0.8857 | 0.8857 |
| | 10000 | 14.2372 | 0.6091 | 0.6548 | 0.6091 | 0.6091 | 0.6091 |

Table 3.11: Percent average absolute bias ($\overline{\text{AB}}$), percent average absolute relative bias ($\overline{\text{ARB}}$), and percent average root mean squared error ($\overline{\text{RMSE}}$) of all different estimators of $\rho$ for the case $\rho = 0.7$ and $D_t^2 = 0.5$ varying by length of time $T$.

| Quality Measure | Size Group $T$ | Estimator | | | | | |
|---|---|---|---|---|---|---|---|
| | | $\rho_{naive}$ | $\rho_{hat}$ | $\rho_{hat,h}$ | $\rho_{hat,g}$ | $\rho_{htrun}$ | $\rho_{gtrun}$ |
| **Case** $\rho = 0.7,\ D_t^2 = 1$ | | | | | | | |
| $\overline{\text{AB}}_\rho(\%)$ | 50 | 23.8169 | 13.5272 | 12.5907 | 13.4735 | 13.5272 | 13.5272 |
| | 100 | 23.6410 | 8.4927 | 8.1822 | 8.4854 | 8.4927 | 8.4927 |
| | 250 | 23.6071 | 5.2145 | 5.1000 | 5.2128 | 5.2145 | 5.2145 |
| | 500 | 23.6440 | 3.5243 | 3.5643 | 3.5237 | 3.5243 | 3.5243 |
| | 1000 | 23.5270 | 2.5067 | 2.5780 | 2.5062 | 2.5067 | 2.5067 |
| | 2500 | 23.6041 | 1.5810 | 1.8524 | 1.5808 | 1.5810 | 1.5810 |
| | 5000 | 23.5998 | 1.1328 | 1.4991 | 1.1324 | 1.1328 | 1.1328 |
| | 10000 | 23.6199 | 0.7690 | 1.3490 | 0.7689 | 0.7690 | 0.7690 |
| $\overline{\text{ARB}}_\rho(\%)$ | 50 | 36.3232 | 21.7372 | 20.0888 | 21.6316 | 21.7372 | 21.7372 |
| | 100 | 34.9040 | 12.6696 | 12.2061 | 12.6577 | 12.6696 | 12.6696 |
| | 250 | 34.1648 | 7.5740 | 7.4108 | 7.5714 | 7.5740 | 7.5740 |
| | 500 | 33.9746 | 5.0764 | 5.1350 | 5.0755 | 5.0764 | 5.0764 |
| | 1000 | 33.6355 | 3.5889 | 3.6912 | 3.5881 | 3.5889 | 3.5889 |
| | 2500 | 33.7075 | 2.2593 | 2.6464 | 2.2589 | 2.2593 | 2.2593 |
| | 5000 | 33.7108 | 1.6190 | 2.1424 | 1.6185 | 1.6190 | 1.6190 |
| | 10000 | 33.7357 | 1.0987 | 1.9273 | 1.0985 | 1.0987 | 1.0987 |
| $\overline{\text{RMSE}}_\rho(\%)$ | 50 | 25.9306 | 18.7904 | 16.6199 | 18.5753 | 18.7904 | 18.7904 |
| | 100 | 24.6636 | 10.9064 | 10.3980 | 10.8832 | 10.9064 | 10.9064 |
| | 250 | 24.0422 | 6.6075 | 6.5148 | 6.6056 | 6.6075 | 6.6075 |
| | 500 | 23.8531 | 4.4669 | 4.5310 | 4.4662 | 4.4669 | 4.4669 |
| | 1000 | 23.6290 | 3.1352 | 3.2383 | 3.1344 | 3.1352 | 3.1352 |
| | 2500 | 23.6463 | 1.9843 | 2.2831 | 1.9839 | 1.9843 | 1.9843 |
| | 5000 | 23.6202 | 1.4118 | 1.8433 | 1.4115 | 1.4118 | 1.4118 |
| | 10000 | 23.6297 | 0.9783 | 1.5815 | 0.9781 | 0.9783 | 0.9783 |

Table 3.12: Percent average absolute bias ($\overline{\text{AB}}$), percent average absolute relative bias ($\overline{\text{ARB}}$), and percent average root mean squared error ($\overline{\text{RMSE}}$) of all different estimators of $\rho$ for the case $\rho = 0.7$ and $D_t^2 = 1$ varying by length of time $T$.

| Quality Measure | Size Group $T$ | Estimator | | | | | |
|---|---|---|---|---|---|---|---|
| | | $\rho_{naive}$ | $\rho_{hat}$ | $\rho_{hat,h}$ | $\rho_{hat,g}$ | $\rho_{htrun}$ | $\rho_{gtrun}$ |
| **Case** $\rho = 0.7,\ D_t^2 = 1.5$ | | | | | | | |
| $\overline{\text{AB}}_\rho(\%)$ | 50 | 30.6366 | 19.9933 | 16.3248 | 19.0284 | 19.9933 | 19.9933 |
| | 100 | 30.4874 | 11.7795 | 10.9668 | 11.7016 | 11.7795 | 11.7795 |
| | 250 | 30.2825 | 7.2772 | 6.9669 | 7.2625 | 7.2772 | 7.2772 |
| | 500 | 30.4551 | 4.9912 | 5.2543 | 4.9843 | 4.9912 | 4.9912 |
| | 1000 | 30.3684 | 3.5047 | 4.0670 | 3.4994 | 3.5047 | 3.5047 |
| | 2500 | 30.3258 | 2.2014 | 3.3267 | 2.1994 | 2.2014 | 2.2014 |
| | 5000 | 30.2916 | 1.5742 | 3.0217 | 1.5716 | 1.5742 | 1.5742 |
| | 10000 | 30.3264 | 1.1008 | 2.9817 | 1.1006 | 1.1008 | 1.1008 |
| $\overline{\text{ARB}}_\rho(\%)$ | 50 | 46.9834 | 33.1100 | 26.4462 | 31.2291 | 33.1100 | 33.1100 |
| | 100 | 44.8480 | 17.6650 | 16.4189 | 17.5401 | 17.6650 | 17.6650 |
| | 250 | 43.6956 | 10.5631 | 10.0991 | 10.5409 | 10.5631 | 10.5631 |
| | 500 | 43.7064 | 7.1769 | 7.5563 | 7.1668 | 7.1769 | 7.1769 |
| | 1000 | 43.5003 | 5.0261 | 5.8316 | 5.0184 | 5.0261 | 5.0261 |
| | 2500 | 43.4070 | 3.1541 | 4.7678 | 3.1513 | 3.1541 | 3.1541 |
| | 5000 | 43.3132 | 2.2524 | 4.3220 | 2.2486 | 2.2524 | 2.2524 |
| | 10000 | 43.3356 | 1.5734 | 4.2618 | 1.5731 | 1.5734 | 1.5734 |
| $\overline{\text{RMSE}}_\rho(\%)$ | 50 | 32.8568 | 35.0098 | 21.9438 | 28.7818 | 35.0098 | 35.0098 |
| | 100 | 31.5662 | 15.7031 | 14.0274 | 15.4320 | 15.7031 | 15.7031 |
| | 250 | 30.6875 | 9.0761 | 8.7959 | 9.0552 | 9.0761 | 9.0761 |
| | 500 | 30.6649 | 6.3282 | 6.6707 | 6.3192 | 6.3282 | 6.3282 |
| | 1000 | 30.4699 | 4.3821 | 5.1092 | 4.3772 | 4.3821 | 4.3821 |
| | 2500 | 30.3688 | 2.7906 | 3.9745 | 2.7876 | 2.7906 | 2.7906 |
| | 5000 | 30.3129 | 1.9886 | 3.4753 | 1.9857 | 1.9886 | 1.9886 |
| | 10000 | 30.3367 | 1.3720 | 3.2363 | 1.3704 | 1.3720 | 1.3720 |

Table 3.13: Percent average absolute bias ($\overline{\text{AB}}$), percent average absolute relative bias ($\overline{\text{ARB}}$), and percent average root mean squared error ($\overline{\text{RMSE}}$) of all different estimators of $\rho$ for the case $\rho = 0.7$ and $D_t^2 = 1.5$ varying by length of time $T$.

| Quality Measure | Size Group $T$ | Estimator | | | | | |
|---|---|---|---|---|---|---|---|
| | | $\rho_{naive}$ | $\rho_{hat}$ | $\rho_{hat,h}$ | $\rho_{hat,g}$ | $\rho_{htrun}$ | $\rho_{gtrun}$ |
| **Case** $\rho = 0.7,\ D_t^2 \sim Uni(0.25, 0.75)$ | | | | | | | |
| $\overline{\text{AB}}_\rho(\%)$ | 50 | 15.0413 | 8.2216 | 8.1254 | 8.2211 | 8.2216 | 8.2216 |
| | 100 | 14.6797 | 5.4744 | 5.4364 | 5.4743 | 5.4744 | 5.4744 |
| | 250 | 14.6387 | 3.2826 | 3.2898 | 3.2826 | 3.2826 | 3.2826 |
| | 500 | 14.4304 | 2.2700 | 2.2778 | 2.2700 | 2.2700 | 2.2700 |
| | 1000 | 14.3056 | 1.5897 | 1.5973 | 1.5897 | 1.5897 | 1.5897 |
| | 2500 | 14.2719 | 0.9956 | 1.0225 | 0.9956 | 0.9956 | 0.9956 |
| | 5000 | 14.2568 | 0.6783 | 0.7086 | 0.6783 | 0.6783 | 0.6783 |
| | 10000 | 14.2298 | 0.4915 | 0.5203 | 0.4915 | 0.4915 | 0.4915 |
| $\overline{\text{ARB}}_\rho(\%)$ | 50 | 23.4334 | 13.3956 | 13.2277 | 13.3947 | 13.3956 | 13.3956 |
| | 100 | 21.6722 | 8.1885 | 8.1326 | 8.1883 | 8.1885 | 8.1885 |
| | 250 | 21.2412 | 4.7985 | 4.8091 | 4.7985 | 4.7985 | 4.7985 |
| | 500 | 20.7778 | 3.2804 | 3.2916 | 3.2803 | 3.2804 | 3.2804 |
| | 1000 | 20.5288 | 2.2838 | 2.2947 | 2.2838 | 2.2838 | 2.2838 |
| | 2500 | 20.4382 | 1.4259 | 1.4646 | 1.4259 | 1.4259 | 1.4259 |
| | 5000 | 20.3884 | 0.9701 | 1.0133 | 0.9701 | 0.9701 | 0.9701 |
| | 10000 | 20.3386 | 0.7024 | 0.7437 | 0.7024 | 0.7024 | 0.7024 |
| $\overline{\text{RMSE}}_\rho(\%)$ | 50 | 16.9678 | 10.6896 | 10.5459 | 10.6888 | 10.6896 | 10.6896 |
| | 100 | 15.7017 | 6.9452 | 6.9072 | 6.9451 | 6.9452 | 6.9452 |
| | 250 | 15.0627 | 4.1888 | 4.2058 | 4.1888 | 4.1888 | 4.1888 |
| | 500 | 14.6330 | 2.8674 | 2.8788 | 2.8674 | 2.8674 | 2.8674 |
| | 1000 | 14.4024 | 2.0104 | 2.0209 | 2.0104 | 2.0104 | 2.0104 |
| | 2500 | 14.3096 | 1.2453 | 1.2702 | 1.2453 | 1.2453 | 1.2453 |
| | 5000 | 14.2749 | 0.8619 | 0.9012 | 0.8619 | 0.8619 | 0.8619 |
| | 10000 | 14.2391 | 0.6129 | 0.6572 | 0.6129 | 0.6129 | 0.6129 |

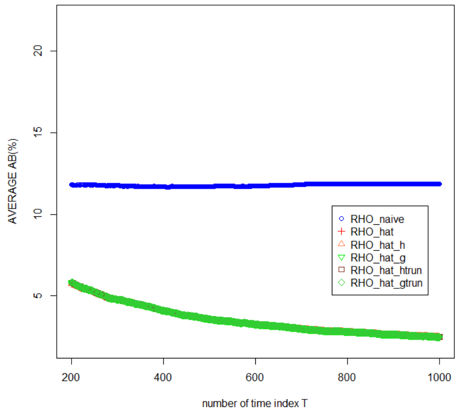
Table 3.14: Percent average absolute bias ($\overline{\text{AB}}$), percent average absolute relative bias ($\overline{\text{ARB}}$), and percent average root mean squared error ($\overline{\text{RMSE}}$) of all different estimators of $\rho$ for the case $\rho = 0.7$ and $D_t^2 \sim Uni(0.25, 0.75)$ varying by length of time $T$.

| Quality Measure | Size Group $T$ | Estimator | | | | | |
|---|---|---|---|---|---|---|---|
| | | $\rho_{naive}$ | $\rho_{hat}$ | $\rho_{hat,h}$ | $\rho_{hat,g}$ | $\rho_{htrun}$ | $\rho_{gtrun}$ |
| **Case** $\rho = 0.7,\ D_t^2 \sim Uni(0.75, 1.25)$ | | | | | | | |
| $\overline{\mathrm{AB}}_\rho(\%)$ | 50 | 23.6508 | 13.1177 | 12.1957 | 13.0673 | 13.1177 | 13.1177 |
| | 100 | 23.7205 | 8.6817 | 8.3482 | 8.6755 | 8.6817 | 8.6817 |
| | 250 | 23.5758 | 5.2865 | 5.2058 | 5.2851 | 5.2865 | 5.2865 |
| | 500 | 23.7174 | 3.6037 | 3.6318 | 3.6031 | 3.6037 | 3.6037 |
| | 1000 | 23.6786 | 2.4555 | 2.6209 | 2.4551 | 2.4555 | 2.4555 |
| | 2500 | 23.5950 | 1.5691 | 1.8579 | 1.5689 | 1.5691 | 1.5691 |
| | 5000 | 23.6215 | 1.1619 | 1.5541 | 1.1617 | 1.1619 | 1.1619 |
| | 10000 | 23.6320 | 0.8116 | 1.3735 | 0.8115 | 0.8116 | 0.8116 |
| $\overline{\mathrm{ARB}}_\rho(\%)$ | 50 | 36.0095 | 20.9421 | 19.3673 | 20.8466 | 20.9421 | 20.9421 |
| | 100 | 34.9014 | 13.0637 | 12.5413 | 13.0536 | 13.0637 | 13.0637 |
| | 250 | 34.1010 | 7.7143 | 7.5869 | 7.7123 | 7.7143 | 7.7143 |
| | 500 | 34.0972 | 5.2002 | 5.2406 | 5.1994 | 5.2002 | 5.2002 |
| | 1000 | 33.9147 | 3.5214 | 3.7618 | 3.5209 | 3.5214 | 3.5214 |
| | 2500 | 33.7293 | 2.2445 | 2.6576 | 2.2442 | 2.2445 | 2.2445 |
| | 5000 | 33.7652 | 1.6619 | 2.2227 | 1.6616 | 1.6619 | 1.6619 |
| | 10000 | 33.7734 | 1.1601 | 1.9633 | 1.1600 | 1.1601 | 1.1601 |
| $\overline{\mathrm{RMSE}}_\rho(\%)$ | 50 | 25.7863 | 17.6181 | 16.0368 | 17.4420 | 17.6181 | 17.6181 |
| | 100 | 24.7938 | 11.2387 | 10.7695 | 11.2283 | 11.2387 | 11.2387 |
| | 250 | 24.0086 | 6.6489 | 6.5380 | 6.6469 | 6.6489 | 6.6489 |
| | 500 | 23.9220 | 4.5284 | 4.5984 | 4.5276 | 4.5284 | 4.5284 |
| | 1000 | 23.7795 | 3.1135 | 3.3082 | 3.1131 | 3.1135 | 3.1135 |
| | 2500 | 23.6348 | 1.9931 | 2.2972 | 1.9927 | 1.9931 | 1.9931 |
| | 5000 | 23.6431 | 1.4594 | 1.8971 | 1.4591 | 1.4594 | 1.4594 |
| | 10000 | 23.6421 | 1.0182 | 1.6184 | 1.0181 | 1.0182 | 1.0182 |

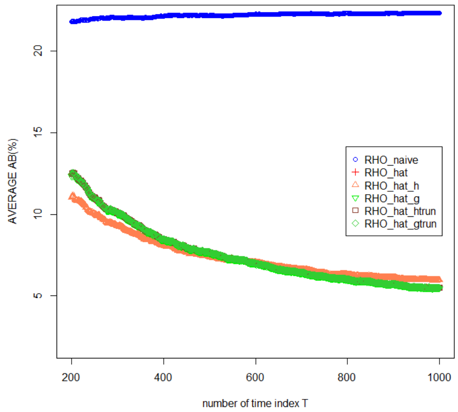
Table 3.15: Percent average absolute bias ($\overline{\mathrm{AB}}$), percent average absolute relative bias ($\overline{\mathrm{ARB}}$), and percent average root mean squared error ($\overline{\mathrm{RMSE}}$) of all different estimators of $\rho$ for the case $\rho = 0.7$ and $D_t^2 \sim Uni(0.75, 1.25)$ varying by length of time $T$.

| Quality Measure | Size Group $T$ | Estimator | | | | | |
|---|---|---|---|---|---|---|---|
| | | $\rho_{naive}$ | $\rho_{hat}$ | $\rho_{hat,h}$ | $\rho_{hat,g}$ | $\rho_{htrun}$ | $\rho_{gtrun}$ |
| **Case** $\rho = 0.7$, $D_t^2 \sim Uni(1.25, 1.75)$ | | | | | | | |
| $\overline{\text{AB}}_\rho(\%)$ | 50 | 30.0685 | 19.8833 | 16.3296 | 18.6314 | 18.8088 | 18.9404 |
| | 100 | 30.4052 | 12.0051 | 11.2764 | 11.9636 | 12.0051 | 12.0051 |
| | 250 | 30.3175 | 7.3817 | 7.3645 | 7.3722 | 7.3817 | 7.3817 |
| | 500 | 30.2480 | 4.9692 | 5.1701 | 4.9618 | 4.9692 | 4.9692 |
| | 1000 | 30.2899 | 3.4504 | 4.0170 | 3.4459 | 3.4504 | 3.4504 |
| | 2500 | 30.2992 | 2.1740 | 3.2174 | 2.1713 | 2.1740 | 2.1740 |
| | 5000 | 30.3028 | 1.5134 | 3.0314 | 1.5124 | 1.5134 | 1.5134 |
| | 10000 | 30.3567 | 1.1251 | 3.0278 | 1.1249 | 1.1251 | 1.1251 |
| $\overline{\text{ARB}}_\rho(\%)$ | 50 | 45.3090 | 31.3297 | 25.4217 | 29.2347 | 29.5968 | 29.8091 |
| | 100 | 44.5711 | 17.9554 | 16.8488 | 17.8886 | 17.9554 | 17.9554 |
| | 250 | 43.7558 | 10.7139 | 10.6911 | 10.6998 | 10.7139 | 10.7139 |
| | 500 | 43.4057 | 7.1496 | 7.4432 | 7.1389 | 7.1496 | 7.1496 |
| | 1000 | 43.3781 | 4.9483 | 5.7659 | 4.9419 | 4.9483 | 4.9483 |
| | 2500 | 43.3190 | 3.1080 | 4.6030 | 3.1041 | 3.1080 | 3.1080 |
| | 5000 | 43.2873 | 2.1621 | 4.3335 | 2.1607 | 2.1621 | 2.1621 |
| | 10000 | 43.3722 | 1.6078 | 4.3282 | 1.6076 | 1.6078 | 1.6078 |
| $\overline{\text{RMSE}}_\rho(\%)$ | 50 | 32.3378 | 43.1538 | 21.5865 | 25.8485 | 26.4618 | 26.8042 |
| | 100 | 31.5271 | 15.6459 | 14.5596 | 15.5638 | 15.6459 | 15.6459 |
| | 250 | 30.7817 | 9.3743 | 9.2586 | 9.3579 | 9.3743 | 9.3743 |
| | 500 | 30.4588 | 6.2049 | 6.4462 | 6.1949 | 6.2049 | 6.2049 |
| | 1000 | 30.3965 | 4.3088 | 5.0167 | 4.3038 | 4.3088 | 4.3088 |
| | 2500 | 30.3399 | 2.7052 | 3.8821 | 2.7016 | 2.7052 | 2.7052 |
| | 5000 | 30.3240 | 1.9270 | 3.4593 | 1.9247 | 1.9270 | 1.9270 |
| | 10000 | 30.3677 | 1.4034 | 3.2903 | 1.4029 | 1.4034 | 1.4034 |

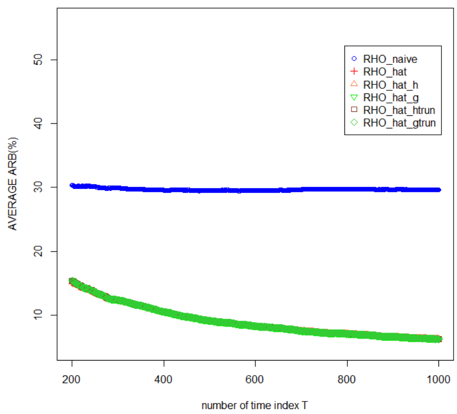
Table 3.16: Percent average absolute bias ($\overline{\text{AB}}$), percent average absolute relative bias ($\overline{\text{ARB}}$), and percent average root mean squared error ($\overline{\text{RMSE}}$) of all different estimators of $\rho$ for the case $\rho = 0.7$ and $D_t^2 \sim Uni(1.25, 1.75)$ varying by length of time $T$.



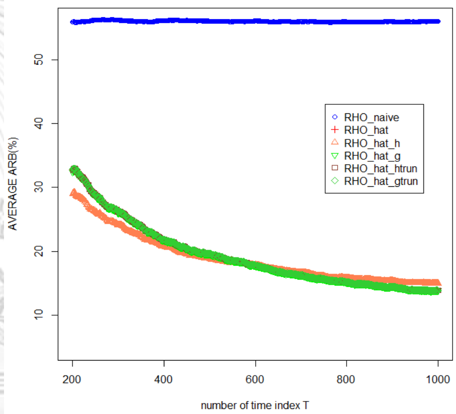

(a) $\overline{\mathrm{AB}}_\rho(\%)$, $D_t^2 = 0.5$




(b) $\overline{\mathrm{AB}}_\rho(\%)$, $D_t^2 = 1.5$

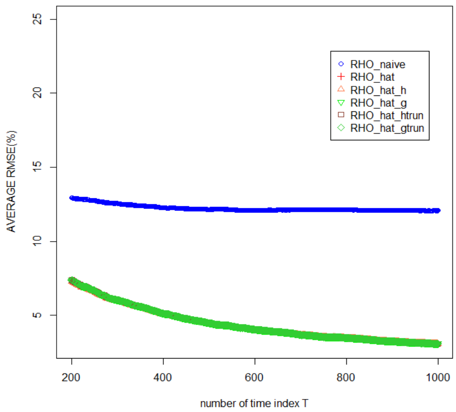



(c) $\overline{\mathrm{ARB}}_\rho(\%)$, $D_t^2 = 0.5$

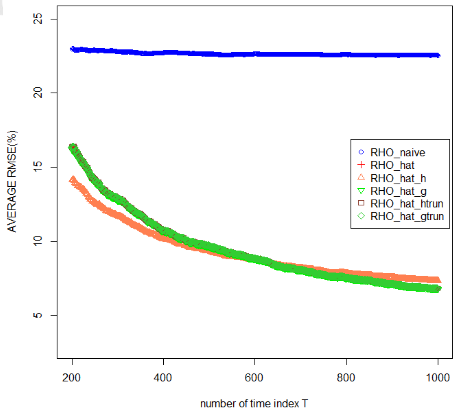



(d) $\overline{\mathrm{ARB}}_\rho(\%)$, $D_t^2 = 1.5$



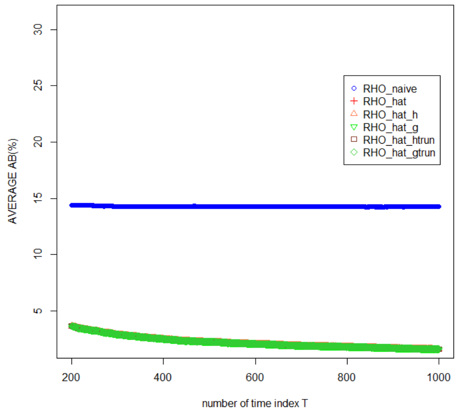



(e) $\overline{\mathrm{RMSE}}_\rho(\%)$, $D_t^2 = 0.5$

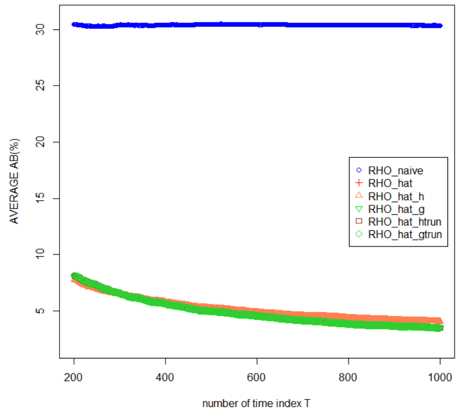



(f) $\overline{\mathrm{RMSE}}_\rho(\%)$, $D_t^2 = 1.5$

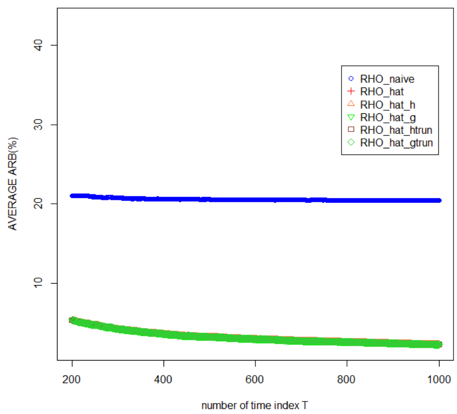
Figure 3.6: Plot of percent average absolute bias ($\overline{\mathrm{AB}}$), percent average absolute relative bias ($\overline{\mathrm{ARB}}$), and percent average root mean squared error ($\overline{\mathrm{RMSE}}$) of all different estimators of $\rho$ varying by length of time $T$ from 200 to 1000 for the case $\rho = 0.4$ and different values of the constant $D_t^2$.



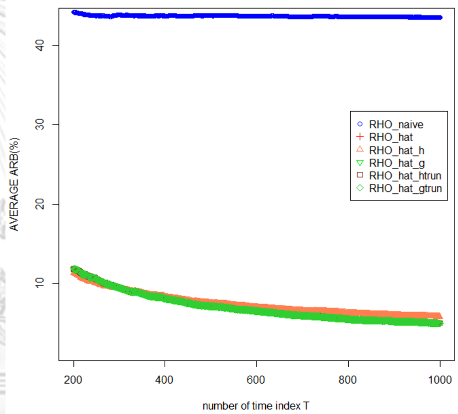

(a) $\overline{\mathrm{AB}}_{\rho}(\%)$, $D_t^2 = 0.5$




(b) $\overline{\mathrm{AB}}_{\rho}(\%)$, $D_t^2 = 1.5$

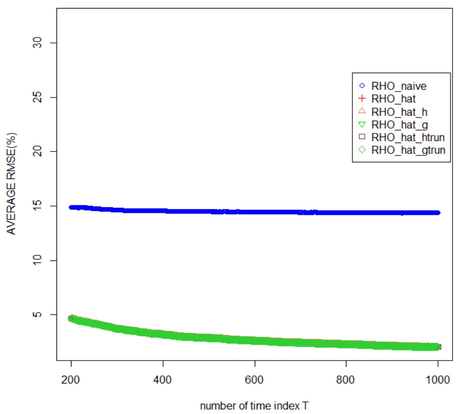



(c) $\overline{\mathrm{ARB}}_{\rho}(\%)$, $D_t^2 = 0.5$

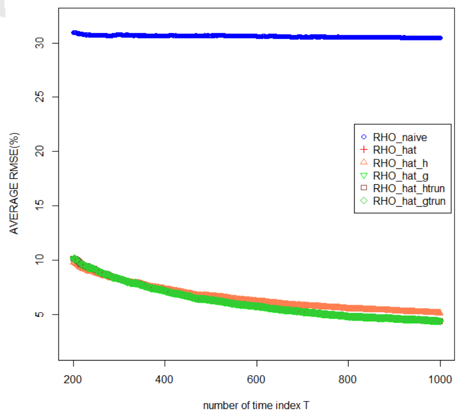



(d) $\overline{\mathrm{ARB}}_{\rho}(\%)$, $D_t^2 = 1.5$



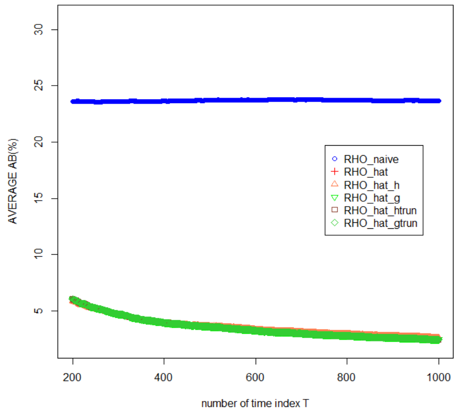



(e) $\overline{\mathrm{RMSE}}_{\rho}(\%)$, $D_t^2 = 0.5$

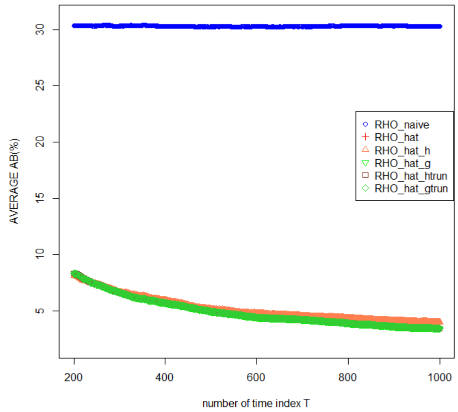



(f) $\overline{\mathrm{RMSE}}_{\rho}(\%)$, $D_t^2 = 1.5$

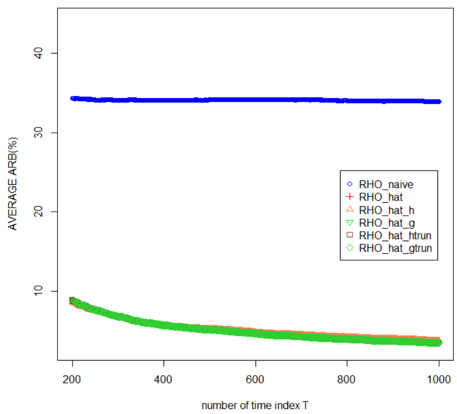
Figure 3.7: Plot of percent average absolute bias ($\overline{\mathrm{AB}}$), percent average absolute relative bias ($\overline{\mathrm{ARB}}$), and percent average root mean squared error ($\overline{\mathrm{RMSE}}$) of all different estimators of $\rho$ varying by length of time $T$ from 200 to 1000 for the case $\rho = 0.7$ and different values of the constant $D_t^2$.



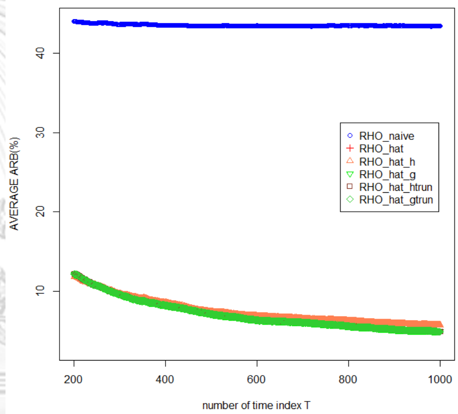

(a) $\overline{\mathrm{AB}}_\rho(\%)$, $D_t^2 \sim Uni(0.75, 1.25)$




(b) $\overline{\mathrm{AB}}_\rho(\%)$, $D_t^2 \sim Uni(1.25, 1.75)$

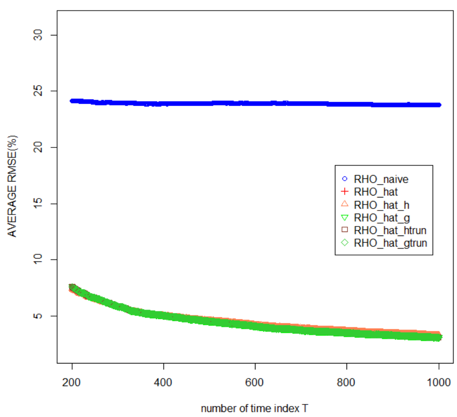



(c) $\overline{\mathrm{ARB}}_\rho(\%)$, $D_t^2 \sim Uni(0.75, 1.25)$

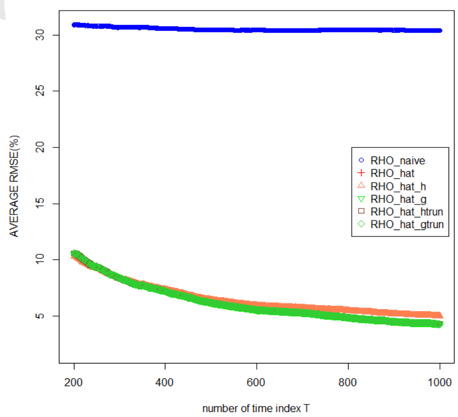



(d) $\overline{\mathrm{ARB}}_\rho(\%)$, $D_t^2 \sim Uni(1.25, 1.75)$






(e) $\overline{\mathrm{RMSE}}_\rho(\%)$, $D_t^2 \sim Uni(0.75, 1.25)$




(f) $\overline{\mathrm{RMSE}}_\rho(\%)$, $D_t^2 \sim Uni(1.25, 1.75)$

Figure 3.8: Plot of percent average absolute bias ($\overline{\mathrm{AB}}$), percent average absolute relative bias ($\overline{\mathrm{ARB}}$), and percent average root mean squared error ($\overline{\mathrm{RMSE}}$) of all different estimators of $\rho$ varying by length of time $T$ from 200 to 1000 for the case $\rho = 0.7$ and different values of the uniformly distributed $D_t^2$.

From Tables 3.5-3.16 and Figures 3.6 - 3.8, three quality measures are smaller for all adjusted estimates of $\rho$ when $T$ is large. The values of three quality measures for the naive estimator $\rho_{naive}$ are approximately the same for all values of $T$ and do not go to zero even for a large value of $T$. The result agrees with Proposition 3.6.

In contrast, three measures for other adjusted estimates are smaller and go to zero as $T$ increases. Notice from the case $\rho = 4$ in Table 3.7 that some values of three measures for other adjusted estimates are extremely higher than the naive estimates when $T = 50$. For a small values of $T$, the high value of sampling variance, compared with the variance of noise, causes the negative value of $\hat{U}_{2,\theta}$. In this case, we recommend to use the positive adjusted estimator $\rho_{hat,h}$ and $\rho_{hat,g}$ instead of the estimator $\rho_{hat}$.

However, only the positive adjusted estimator $\rho_{hat,h}$ has large values of three quality measures when sampling variances are large, compared with other adjusted estimates. For example, consider the case $\rho = 0.7$ and sampling variances $D_t^2 \sim Uni(1.25, 1.75)$, the percent average absolute relative bias of $\rho_{hat,h}$ is 3.2903 when $T = 10000$. However, the percent average absolute relative bias of $\rho_{hat}$, $\rho_{hat,g}$, $\rho_{hat,htrun}$, and $\rho_{hat,gtrun}$ are 1.4034, 1.4029, 1.4034, and 1.4034, respectively. Comparing among these quality measures in Table 3.5 and Figure 3.6, the estimator $\rho_{hat}$ can be represented for the estimator $\rho_\theta$, when $\theta_t$'s are unobserved. In spite of the fact that the values of three measures of $\rho_{hat,g}$ is less than $\rho_{hat}$, the $g$-function in Lemma 3.21 is applied when $U_{2,\theta}$ is negative. Hence, we can conclude that the estimator $\rho_{hat,gtrun}$ is a good estimator for represent for the estimator $\rho_\theta$, when $\theta_t$'s are unobserved.

**2.) Test Statistics for the Stationary Test**

In this section, we consider the performances of the test statistic for the stationary test. The stationary test in this section is the test of the null hypothesis that $\rho = 1$, against the alternative hypothesis that $|\rho| < 1$. That is,

$$H_0 : \rho = 1 \quad \text{or nonstationary AR(1)},$$

against

$$H_1 : |\rho| < 1 \quad \text{or stationary AR(1)}.$$

In this section, we consider only the case $\rho = 1$. For the values of sampling variances of $e_t$, we study different values of $D_t^2$ as follows.

**1.** The variances of $\{e_t\}$ are a constant not depending on time $t$. We perform simulations in three settings.

(1.1) $D_t^2 = 0.5$,

(1.2) $D_t^2 = 1$,

(1.3) $D_t^2 = 1.5$.

**2.** The variances of $\{e_t\}$ follow a uniform distribution. We perform simulations in three settings.



(2.1) $D_t^2 \sim Uni(0.25, 0.75)$,

(2.2) $D_t^2 \sim Uni(0.75, 1.25)$,

(2.3) $D_t^2 \sim Uni(1.25, 1.75)$,

where $Uni(a, b)$ is the continuous uniform distribution over interval $[a, b]$.

The results shown in Tables 3.18-3.19 and Figures 3.9-3.10 are presented using the following notations, some notations are followed from Theorem 3.10, Theorem 3.17, (3.19), (3.152)-(3.153), (3.154)-(3.155), and (3.156)-(3.159).

(1) $\tau_{true} = (\rho_{true} - 1)U_{2,\theta}^{\frac{1}{2}}U_{4,\theta}^{-\frac{1}{2}}$, the Dickey-Fuller test statistic based on the variables of interest $\{\theta_t\}$,

(2) $\tau_{naive} = (\rho_{naive} - 1)U_{2,y}^{\frac{1}{2}}U_{4,y}^{-\frac{1}{2}}$, the naive test statistic based on observed variables $\{y_t\}$ and ignore sampling errors,

(3) $\tau_{hat} = (\rho_{hat} - 1)\hat{U}_{2,\theta}^{\frac{1}{2}}\hat{U}_{4,\theta}^{-\frac{1}{2}}$, an adjustment to the naive test statistic,

(4) $\tau_{hat,h} = (\rho_{hat,h} - 1)\hat{U}_{2,\theta,h}^{\frac{1}{2}}\hat{U}_{4,\theta,h}^{-\frac{1}{2}}$, an adjustment to the naive test statistic by the $h$-function,

(5) $\tau_{hat,g} = (\rho_{hat,g} - 1)\hat{U}_{2,\theta,g}^{\frac{1}{2}}\hat{U}_{4,\theta,g}^{-\frac{1}{2}}$, an adjustment to the naive test statistic by the $g$-function,

(6) $\tau_{hat,htrun} = (\rho_{hat,htrun} - 1)\hat{U}_{2,\theta,htrun}^{\frac{1}{2}}\hat{U}_{4,\theta,htrun}^{-\frac{1}{2}}$, a truncation version of the $h$-function approximation,

(7) $\tau_{hat,gtrun} = (\rho_{hat,gtrun} - 1)\hat{U}_{2,\theta,gtrun}^{\frac{1}{2}}\hat{U}_{4,\theta,gtrun}^{-\frac{1}{2}}$, a truncation version of the $g$-function approximation.

| $T$ | Probability of a Small Values | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0.01 | 0.025 | 0.05 | 0.1 | 0.9 | 0.95 | 0.975 | 0.99 |
| 25 | -2.66 | -2.26 | -1.95 | -1.60 | 0.92 | 1.33 | 1.70 | 2.16 |
| 50 | -2.62 | -2.25 | -1.95 | -1.61 | 0.91 | 1.31 | 1.66 | 2.07 |
| 100 | -2.60 | -2.24 | -1.95 | -1.61 | 0.90 | 1.29 | 1.64 | 2.03 |
| 250 | -2.58 | -2.24 | -1.95 | -1.62 | 0.89 | 1.28 | 1.63 | 2.01 |
| 500 | -2.58 | -2.24 | -1.95 | -1.62 | 0.89 | 1.28 | 1.62 | 2.00 |
| 750 | -2.58 | -2.24 | -1.95 | -1.62 | 0.89 | 1.28 | 1.62 | 2.00 |
| $\infty$ | -2.58 | -2.23 | -1.95 | -1.62 | 0.89 | 1.28 | 1.62 | 1.99 |

Table 3.17: ([5], Page 58) Empirical percentiles for $\tau$ statistics.

Under the null hypothesis $H_0 : \rho = 1$, the null hypothesis with be rejected under the level of significance $\alpha$ if the test statistic $\tau_\theta$ satisfies $\sup P_{\tau_\theta}(\tau_\theta \in R) \leq \alpha$, where $R$ is the rejection region. Since this test is a two-sided test, the rejection region $R$ is $(-\infty, c_1) \cup (c_2, \infty)$, for some constants $c_1$ and $c_2$. The percentiles for the distribution in Theorem 3.2 (4) were given in Table 3.17 by Dickey [5]. Hence, the constants $c_1$ and $c_2$ are applied for evaluating the probability $P_{\tau_\theta}(\tau_\theta \in R)$ when the level of significance $\alpha$ is fixed. For example, under the level of significance $\alpha = 0.05$ and $T = 10000$, the null hypothesis is rejected if $\tau_\theta < -2.23$ or $\tau_\theta > 1.62$.

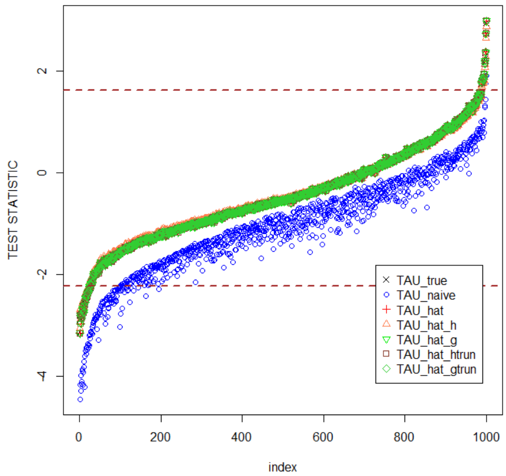
To consider the test statistic, Tables 3.18-3.20 and Figures 3.9-3.10 show comparisons between all test statistics for 1000 generated data with $\rho = 1$, $T = 10000$ by varying sampling variances.

| | Percentile | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 10 | 25 | 50 | 75 | 90 | 100 |
| **Scenarios 1:** $\rho = 1$, $D_t^2 = 0.5$ | | | | | | | |
| $\tau_{true}$ | -2.6980 | -1.5583 | -1.0902 | -0.5424 | 0.1810 | 0.8474 | 2.9414 |
| $\tau_{naive}$ | -3.8615 | -2.3186 | -1.6834 | -1.0579 | -0.3845 | 0.2124 | 1.9103 |
| $\tau_{hat}$ | -2.6268 | -1.5757 | -1.0956 | -0.5510 | 0.1801 | 0.8460 | 2.9826 |
| $\tau_{hat,h}$ | -2.5226 | -1.5176 | -1.0567 | -0.5317 | 0.1738 | 0.8161 | 2.8694 |
| $\tau_{hat,g}$ | -2.6246 | -1.5746 | -1.0948 | -0.5506 | 0.1801 | 0.8456 | 2.9808 |
| $\tau_{hat,htrun}$ | -2.6268 | -1.5757 | -1.0956 | -0.5510 | 0.1801 | 0.8460 | 2.9826 |
| $\tau_{hat,gtrun}$ | -2.6268 | -1.5757 | -1.0956 | -0.5510 | 0.1801 | 0.8460 | 2.9826 |
| **Scenarios 1:** $\rho = 1$, $D_t^2 = 1$ | | | | | | | |
| $\tau_{true}$ | -2.5641 | -1.5681 | -1.0677 | -0.5306 | 0.2207 | 0.8961 | 3.0485 |
| $\tau_{naive}$ | -4.5144 | -3.0040 | -2.1097 | -1.3178 | -0.6212 | -0.0541 | 1.4246 |
| $\tau_{hat}$ | -2.5182 | -1.5714 | -1.0700 | -0.5241 | 0.2227 | 0.8802 | 3.0667 |
| $\tau_{hat,h}$ | -2.2383 | -1.4096 | -0.9587 | -0.4657 | 0.1996 | 0.7870 | 2.7293 |
| $\tau_{hat,g}$ | -2.4982 | -1.5611 | -1.0627 | -0.5198 | 0.2212 | 0.8740 | 3.0429 |
| $\tau_{hat,htrun}$ | -2.5182 | -1.5714 | -1.0700 | -0.5241 | 0.2227 | 0.8802 | 3.0667 |
| $\tau_{hat,gtrun}$ | -2.5182 | -1.5714 | -1.0700 | -0.5241 | 0.2227 | 0.8802 | 3.0667 |
| **Scenarios 1:** $\rho = 1$, $D_t^2 = 1.5$ | | | | | | | |
| $\tau_{true}$ | -2.7453 | -1.6867 | -1.1229 | -0.5550 | 0.1617 | 0.8735 | 2.8335 |
| $\tau_{naive}$ | -5.5339 | -3.6182 | -2.6520 | -1.6997 | -0.8976 | -0.3242 | 1.0637 |
| $\tau_{hat}$ | -2.6928 | -1.6533 | -1.1331 | -0.5493 | 0.1594 | 0.8817 | 2.7462 |
| $\tau_{hat,h}$ | -2.2131 | -1.3646 | -0.9476 | -0.4586 | 0.1303 | 0.7300 | 2.2543 |
| $\tau_{hat,g}$ | -2.6307 | -1.6148 | -1.1095 | -0.5382 | 0.1556 | 0.8651 | 2.6827 |
| $\tau_{hat,htrun}$ | -2.6928 | -1.6533 | -1.1331 | -0.5493 | 0.1594 | 0.8817 | 2.7462 |
| $\tau_{hat,gtrun}$ | -2.6928 | -1.6533 | -1.1331 | -0.5493 | 0.1594 | 0.8817 | 2.7462 |

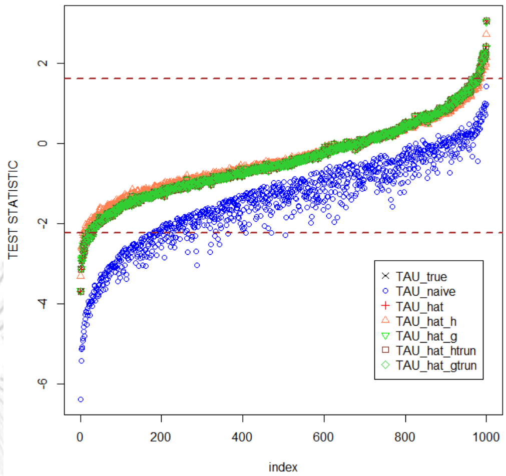
Table 3.18: Values of different estimators of all test statistics varying by different constants $D_t^2$'s for the case $\rho = 1$, $\sigma^2 = 1$, and $T = 10000$.

| | Percentile | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 10 | 25 | 50 | 75 | 90 | 100 |
| **Scenarios 2:** $\rho = 1,\ D_t^2 \sim Uni(0.25, 0.75)$ | | | | | | | |
| $\tau_{true}$ | -2.6568 | -1.5909 | -1.0862 | -0.5133 | 0.2159 | 0.9279 | 2.6899 |
| $\tau_{naive}$ | -3.7574 | -2.3841 | -1.7460 | -1.0285 | -0.3321 | 0.2416 | 1.6350 |
| $\tau_{hat}$ | -2.6396 | -1.5929 | -1.0930 | -0.5058 | 0.2150 | 0.9077 | 2.6869 |
| $\tau_{hat,h}$ | -2.5470 | -1.5357 | -1.0521 | -0.4876 | 0.2074 | 0.8764 | 2.5941 |
| $\tau_{hat,g}$ | -2.6380 | -1.5918 | -1.0923 | -0.5055 | 0.2143 | 0.9072 | 2.6861 |
| $\tau_{hat,htrun}$ | -2.6396 | -1.5929 | -1.0930 | -0.5058 | 0.2150 | 0.9077 | 2.6869 |
| $\tau_{hat,gtrun}$ | -2.6396 | -1.5929 | -1.0930 | -0.5058 | 0.2150 | 0.9077 | 2.6869 |
| **Scenarios 2:** $\rho = 1,\ D_t^2 \sim Uni(0.75, 1.25)$ | | | | | | | |
| $\tau_{true}$ | -2.5397 | -1.6937 | -1.1263 | -0.5089 | 0.2005 | 0.8181 | 2.7249 |
| $\tau_{naive}$ | -4.5551 | -3.0786 | -2.2070 | -1.3834 | -0.6500 | -0.1110 | 1.2240 |
| $\tau_{hat}$ | -2.5515 | -1.6859 | -1.1107 | -0.5077 | 0.2120 | 0.8190 | 2.7899 |
| $\tau_{hat,h}$ | -2.2993 | -1.5102 | -0.9916 | -0.4544 | 0.1881 | 0.7266 | 2.4622 |
| $\tau_{hat,g}$ | -2.5359 | -1.6746 | -1.1012 | -0.5042 | 0.2102 | 0.8123 | 2.7638 |
| $\tau_{hat,htrun}$ | -2.5515 | -1.6859 | -1.1107 | -0.5077 | 0.2120 | 0.8190 | 2.7899 |
| $\tau_{hat,gtrun}$ | -2.5515 | -1.6859 | -1.1107 | -0.5077 | 0.2120 | 0.8190 | 2.7899 |
| **Scenarios 2:** $\rho = 1,\ D_t^2 \sim Uni(1.25, 1.75)$ | | | | | | | |
| $\tau_{true}$ | -2.7477 | -1.6911 | -1.1169 | -0.5154 | 0.2418 | 0.9348 | 2.6847 |
| $\tau_{naive}$ | -5.8113 | -3.5945 | -2.5950 | -1.5901 | -0.7971 | -0.2766 | 0.7480 |
| $\tau_{hat}$ | -2.7648 | -1.6569 | -1.1080 | -0.5151 | 0.2513 | 0.9222 | 2.6270 |
| $\tau_{hat,h}$ | -2.2963 | -1.3579 | -0.9173 | -0.4254 | 0.2080 | 0.7531 | 2.1886 |
| $\tau_{hat,g}$ | -2.7084 | -1.6191 | -1.0860 | -0.5038 | 0.2463 | 0.8994 | 2.5753 |
| $\tau_{hat,htrun}$ | -2.7648 | -1.6569 | -1.1080 | -0.5151 | 0.2513 | 0.9222 | 2.6270 |
| $\tau_{hat,gtrun}$ | -2.7648 | -1.6569 | -1.1080 | -0.5151 | 0.2513 | 0.9222 | 2.6270 |

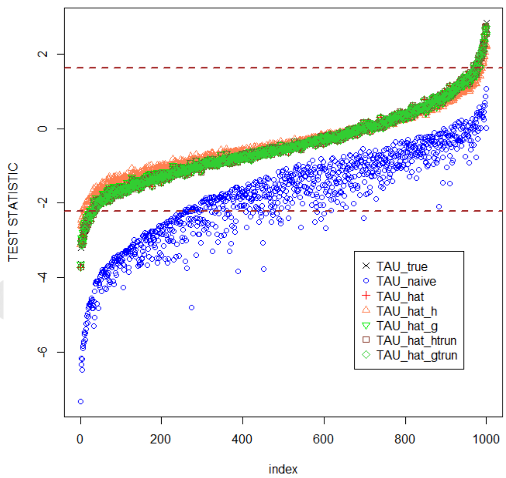
Table 3.19: Values of different estimators of all test statistics varying by different uniformly distributed $D_t^2$'s for the case $\rho = 1$, $\sigma^2 = 1$, and $T = 10000$.




(a) $D^2 = 0.5$.

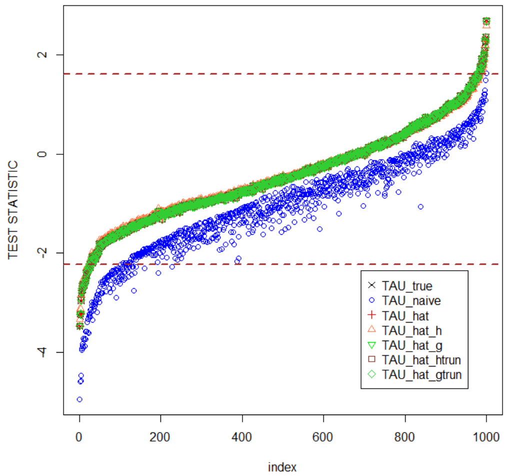



(b) $D^2 = 1$.

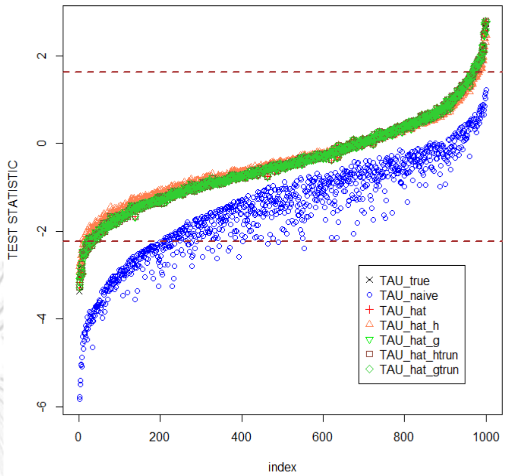



(c) $D^2 = 1.5$

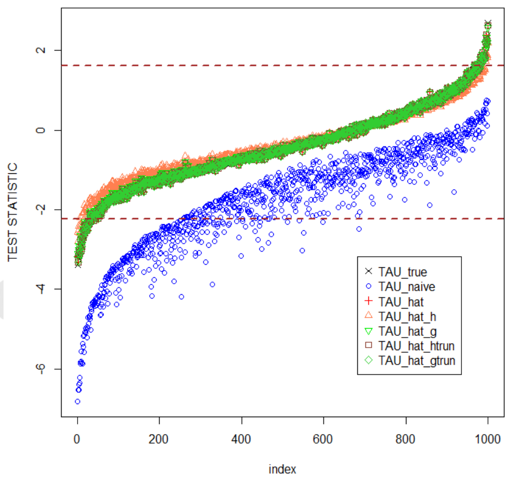
Figure 3.9: Plot of all different test statistics for the case $\rho = 1$, $\sigma^2 = 1$, $T = 10000$, and different values of the constant $D_t^2$.




(a) $D^2 \sim Uni(0.25, 0.75)$.




(b) $D^2 \sim Uni(0.75, 1.25)$.




(c) $D^2 \sim Uni(1.25, 1.75)$

Figure 3.10: Plot of all different test statistics for the case $\rho = 1$, $\sigma^2 = 1$, $T = 10000$, and different values of ranges of the $D_t^2$.

| Sampling Variance | Test Statistic | | | | | | |
|---|---|---|---|---|---|---|---|
| | $\tau_{true}$ | $\tau_{naive}$ | $\tau_{hat}$ | $\tau_{hat,h}$ | $\tau_{hat,g}$ | $\tau_{htrun}$ | $\tau_{gtrun}$ |
| **Scenario 1** | | | | | | | |
| $D_t^2 = 0.5$ | 40 | 117 | 39 | 36 | 39 | 39 | 39 |
| $D_t^2 = 1$ | 53 | 225 | 50 | 30 | 49 | 50 | 50 |
| $D_t^2 = 1.5$ | 56 | 335 | 58 | 24 | 53 | 58 | 58 |
| **Scenario 2** | | | | | | | |
| $D_t^2 \sim Uni(0.25, 0.75)$ | 46 | 125 | 45 | 39 | 45 | 45 | 45 |
| $D_t^2 \sim Uni(0.75, 1.25)$ | 62 | 244 | 63 | 35 | 62 | 63 | 63 |
| $D_t^2 \sim Uni(1.25, 1.75)$ | 56 | 329 | 54 | 19 | 50 | 54 | 54 |

Table 3.20: The number of errors in hypothesis testing from 1000 simulations, for the case $\rho = 1$ and $T = 10000$, under the level of significance 0.05.

From Table 3.20, the number of times that the null hypothesis is rejected by the test statistic $\tau_{naive}$ is greater than 50, and increases when the sampling variance increases. This result shows that the conclusion from test statistic $\tau_{naive}$ is not reliable because there are more than 5% error of the test.

On the other hand, for test statistics $\tau_{true}$, there is around 5% error of the test, which is around 50 samples from 1000 to reject the null hypothesis. In addition, from Table 3.18, the statistic $\tau_{naive}$ is smaller than $\tau_{true}$. In contrast, the statistic $\tau_{hat}$ provides the same conclusion as $\tau_{true}$, under the level of significance 0.05. When estimator $U_{2,\theta}$ or $U_{4,\theta}$ is negative, the positive adjustment by the $g$-function can provide the same result as $\tau_{true}$. However, if the estimator $U_{2,\theta}$ or $U_{4,\theta}$ are both positive, the statistic $\tau_{hat}$ can represent test statistic $\tau_{true}$ when the true variables $\theta_t$'s are unobserved. Hence, from our simulation, we can conclude that $\tau_{hat,gtrun}$ is the best test for the stationary test.

# CHAPTER IV
# CONCLUSION

In this thesis, we have shown the effect of sampling errors for parameter estimation and test statistics for the first order autoregressive model. We proposed 2 new estimators of the unknown parameters $\rho$ and $\sigma^2$ for the first order autoregressive models subject to sampling errors (3.13).

1. The estimator of autoregressive coefficient $\rho$:

$$\hat{\rho}_\theta = \frac{\hat{U}_{1,\theta}}{\hat{U}_{2,\theta}},$$

where $\hat{U}_{1,\theta} = \sum_{t=2}^{T} y_t y_{t-1}$ and $\hat{U}_{2,\theta} = \sum_{t=2}^{T} (y_{t-1}^2 - D_{t-1}^2)$.

2. The estimator of variance of noise $\sigma^2$:

$$\hat{\sigma}_\theta^2 = \frac{1}{T-2} \sum_{t=2}^{T} (y_t - \hat{\rho}_\theta y_{t-1})^2 - \frac{1}{T-2} \sum_{t=2}^{T} \left( D_t^2 + \hat{\rho}_\theta^2 D_{t-1}^2 \right).$$



Moreover, we obtain the representation of test statistic, based on the observed data $y_t$'s:

$$\hat{\tau}_\theta = \frac{(\hat{\rho}_\theta - 1)\sqrt{\hat{U}_{2,\theta}}}{\sqrt{\hat{\sigma}_\theta^2}}.$$

From our study, we found that our estimators $\hat{\rho}_\theta$ and $\hat{\sigma}_\theta^2$ can reduce the bias due to sampling errors. In addition, our statistic $\hat{\tau}_\theta$ provide the same conclusion as the test statistic based on true variables $\theta_t$'s.

However, there has a chance that $\hat{U}_{2,\theta}$ and $\hat{\sigma}_\theta^2$ are negative, when the sampling variances are extremely large. The method of positive adjustments to the estimators with $g$-function by Angkunsit and Suntornchost [2] can solve this problems.

From Tables 3.18-3.20 and Figures 3.9-3.10, the truncation version of positive adjustments of estimators under $g$-function is better than other test statistic in respect of the conclusion of hypothesis testing. To accommodate all situations of estimators, the truncation of test statistic $\tau_{hat,gtrun}$ is recommended.

# REFERENCES


[1] Anderson, T.W.: On asymptotic distributions of estimates of parameters of stochastic difference equations. *The annals of mathematical statistics,* **30**, 676–687 (1959).

[2] Angkunsit, A., Suntornchost, J.: Variable selection in multivariate linear regression models subject to sampling errors. *Chamchuri Journal of Mathematics*, **10**, 28–52 (2018).

[3] Casella G., Berger R.L.: *Statistical Inference, 2nd ed.* Duxbury Press, California, 2002.

[4] Chatterjee, S., Lahiri, P.: A Simple Computational Method for Estimating Mean Squared Prediction Error in General Small-Area Model. *JSM Proceedings*, 3486-3493 (2007).

[5] Dickey, D. A.: *Estimation and hypothesis testing in nonstationary time series.* Ph.D. dissertation, Department of Statistics, Iowa State university, 1976.

[6] Dickey, D. A., Bell, W. R., Miller, R. B.: Unit roots in time series models: tests and implications. *The american statistician,* **40**, 12–26 (1986).

[7] Dickey, D. A., Fuller, W.: Distribution of the estimators for autoregressive time series with a unit root. *J. American Statistical Association*, **74**, 427–431 (1979).

[8] Dickey, D. A., Fuller, W.: Likelihood ratio statistics for autoregressive time series with a unit root. *J. Econometrica*, **49**, 1057–1072 (1981).

[9] Fay, R.E., Herriot, R. A.: Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. *J. American Statistical Association*, **74**, 269–277 (1979).

[10] Folland, G. B.: *Advanced Calculus.* Pearson Prentice Hall: Upper Saddle River, New Jersey, 2002.

[11] Fuller, W.: *Introduction to statistical time series, 2nd ed.* John Wiley & Sons, Inc., New York, 1976.

[12] Lahiri, P., Suntornchost, J.: Variable selection for linear mixed models with applications in small area estimation. *The Indian Journal of Statistics*, **77**, 312–320 (2015).

[13] Lenghoe, N., Suntornchost, J.: Variable selection for linear regression model with general variance. *Journal of Applied Statistics and Information Technology*, **2**, 15–24 (2017).

[14] Li, Y., Lahiri, P.: A Simple Adaptation of Variable Selection Software for Regression Models to Select Variables in Nested Error Regression models. *Sankhya : The Indian Journal of Statistics*, **-B**, 312–317 (2019).

[15] Marhuenda, Y., Morales, D., Pardo, M. C.: Tests for the variance parameter in the Fay–Herriot model. *J. theoretical and applied statistics*, **50**, 27–42 (2016).

[16] Mann, H. B., Wald A.: On the statistical treatment of linear stochastic difference equations. *J. Econometrica*, **11**, 173–220 (1943).

[17] Pfeiffer, P.E.: *Probability for Applications.* Springer-Verlag 1990.

[18] Resnick, S. I.: *A probability path.* Birkhäuser Basel, New York, 2014.

[19] Rubin, H.: Consistency of maximum-likelihood estimates in the explosive case. *Cowles commission monograph,* **10**, 356–364 (1950).

[20] Said, S. E., Dickey, D. A.: Testing for unit roots in autoregressive-moving average models of unknown order. *J. Biometrika*, **71**, 599–607 (1984).

[21] Said, S. E., Dickey, D. A.: Hypothesis testing in ARIMA(p, 1, q) models. *J. American Statistical Association*, **80**, 369–374 (1985).

[22] Spivak, M.: *Calculus on Manifolds.* W.A. Benjamin, New York, 1965.

[23] Tucker, H. G.: *A graduate course in probability.* Academic Press, New York, 1967.

[24] Wei, W.W.S.: *Time series analysis univariate and multivariate methods, 2nd ed.* Pearson Addison Wesley, California, 2005.

[25] White, J. S.: The limiting distribution of the serial correlation coefficient in the explosive case. *The annals of mathematical statistics*, **29**, 1188–1197 (1958).

[26] White, J. S.: The limiting distribution of the serial correlation coefficient in the explosive case II. *The annals of mathematical statistics*, **30**, 831–834 (1959).

# VITA

| | | |
|---|---|---|
| Name | : | Mister Weerapat Rattanachadjane |
| Date of Birth | : | 10 June 1995 |
| Place of Birth | : | Yala, Thailand |
| Education | : | B.Sc. (Mathematics), (First Class Honors),<br>Prince of Songkla University, 2017 |
| Scholarship | : | Development and Promotion of Science and Technology<br>Talents Project (DPST) |