2017

Memory Properties Of Transformations Of Linear Processes And Symmetric Gini Correlation

Yongli Sang
University of Mississippi

MEMORY PROPERTIES OF TRANSFORMATIONS OF LINEAR PROCESSES AND SYMMETRIC GINI CORRELATION

DISSERTATION

A Dissertation
presented in partial fulfillment of requirements
for the degree of Doctor of Philosophy
in the Department of Mathematics
The University of Mississippi

by

YONGLI SANG

May 2017

ABSTRACT

A large class of time series processes can be modeled by linear processes, including a subset of the fractional ARIMA process. Transformation of time series is one of the most popular topics in recent years. In this dissertation, we study the memory properties of transformations of linear processes. Our results show that transformations of short-memory time series still have short-memory and transformations of long-memory time series may have different weaker memory parameters which depend on the power rank of the transformation. In particular, we provide the memory parameters of transformations of the FARIMA(p, d, q) processes. As an example, the memory properties of call option processes at different strike prices are discussed in details.

When we develop the memory properties of transformations of linear processes, we use the Pearson correlation to measure memories. Correlation analysis is another big topic in statistics, which is used to measure the dependence of stochastic processes or random variables. Standard Gini correlation is one of the correlations to measure the dependence between random variables with heavy tailed distributions. However, the asymmetry of Gini covariance and correlation brings a substantial difficulty in interpretation. In this dissertation, we propose a symmetric Gini-type covariance and correlation (ρ_g) based on the joint rank function. The proposed correlation ρ_g is symmetric and is more robust than the Pearson correlation but less robust than the Kendall's τ correlation in terms of influence functions. Furthermore, we establish the relationship between ρ_g and the linear correlation ρ for a class of random vectors in the family of elliptical distributions, which allows us to estimate ρ

based on estimation of ρ_g. We compare asymptotic efficiencies of linear correlation estimators based on the symmetric Gini, and the proposed measure ρ_g shows superior finite sample performance, which makes it attractive in applications.

DEDICATION

I dedicate my dissertation work to my loving parents, Yuping Sang and Zuolan Du, who have always loved me unconditionally and encouraged me to pursue my dreams. I will always appreciate all they have done for me. I would also like to dedicate this work and give special thanks to my husband, Shaohui Wang, and my wonderful daughter, Taylor, for being there as constant sources of support and encouragement during the graduate program. I am truly thankful for having you both in my life.

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere appreciation to my advisor, Dr. Hailin Sang, who has been a tremendous mentor for me. I would like to thank him for his continuous support of my Ph.D study and continuous encouragement of my research. His advice on my career is also invaluable.

Besides my advisor, I would also like to thank Dr. Xin Dang for her patience and advice on both research as well as on my career. Her advice helped me in my entire Ph.D program.

I would also like to thank my committee members, Dr. Jeremy Clark, Dr. Erwin Miña Díaz, and Dr. Walt Mayer for serving as my committee members even at hardship. I also want to thank them for letting my defense be an enjoyable moment, and for their brilliant comments and suggestions.

My sincere thanks also goes to Ms. Colleen Gilman, Dr. Kayla Harville, Ms. Casandra Jenkins, Dr. Iwo Labuda, Dr. Talmage James Reid, Dr. Lanzhen Song and Dr. Bing Wei who were always there to help me succeed in my graduate studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 MEMORY PROPERTIES OF TRANSFORMATIONS OF LINEAR PROCESSES

Dittmann and Granger [9] studied the memory properties of polynomial transformation of Gaussian FARIMA$(0, d, 0)$ processes

$$X_n = (1-B)^{-d}\varepsilon_n = \sum_{i\geq 0} a_i \varepsilon_{n-i}, \tag{1.1.1}$$

where $a_i = \frac{\Gamma(i+d)}{\Gamma(d)\Gamma(i+1)}$, $\varepsilon_i \sim i.i.d.\ \ N(0, \sigma^2)$, $-1 < d < 1/2$ and $d \neq 0$. $d = 0$ gives the i.i.d. process $\{\varepsilon_n\}$. They applied the orthonormality of the Hermite polynomials under the measure for the standard normal distribution. That is,

$$\int_{-\infty}^{\infty} H_m(x) H_n(x) dP(Z \leq x) = I(m = n), \tag{1.1.2}$$

where Hermite polynomials $H_j(x)$ are defined by

$$\left(\frac{d}{dx}\right)^j e^{-x^2/2} = (-1)^j \sqrt{j!} H_j(x) e^{-x^2/2}, j = 0, 1, 2, \cdots,$$

$Z \sim N(0, 1)$ and $I(m = n)$ is the indicator function, $m, n = 0, 1, 2, \cdots$. For example, see [5]. In the continuous case, [43] and [16] studied the nonlinear transformations of fractional Brownian motions. Nevertheless, this nice orthogonal property (1.1.2) does not hold in general when the distribution is not Gaussian. On the other hand, it is witnessed and well known in the financial field that quite many financial data like stock prices have heavier tails than the tail of the normal distribution, for example, see [35]. For the non-Gaussian case, based on the innovations ε_i, [19, 20] developed an expansion with orthogonal terms which is akin to the Hermite expansion for the Gaussian case.

We focus on the transformations of linear processes

$$X_n = \sum_{i=0}^{\infty} a_i \varepsilon_{n-i}, \quad n \in \mathbb{N}, \tag{1.1.3}$$

which are not necessarily Gaussian in this dissertation. To explore the memory properties of transformations $K(X_n)$ with $\mathbb{E}K^2(X_n) < \infty$ of linear processes (1.1.3), we shall apply the decomposition of $K(X_n)$ proposed by [19, 20]. See also the review paper by [24]. This method has been applied to the expansion of $K(X_n)$ for linear processes X_n in the study of many subjects, for examples, weak convergence theorems including central limit theorem, functional central limit theorem, convergence to Wiener-Ito integral and Hermite process ([19, 20, 23, 49, 51]), the kernel density estimation ([22, 48, 28]), the empirical processes of long memory sequences ([50]), the U-statistics ([21, 25]) and the moderate deviations ([52, 34]). Under the condition proposed by [51], we obtain results both in time domain and frequency domain. The results in time domain are consistent with the limit theorems in [20] and [51] via the order of normalization. The results are applicable not only to FARIMA$(0, d, 0)$ processes as studied in [9] for Gaussian case, but also to general FARIMA(p, d, q) processes for some special transformations. The results hold not only for smooth transformations, they also hold for functions which are not differentiable. In particular, we study the memory properties of option time series $(X_n - C)^+$ in finance for different strike price $C > 0$.

We also study the properties of nonlinear transformations of non-stationary time series X_n with the form

$$X_n = \sum_{j=1}^{n} Y_j, \quad \text{where} \quad Y_j = \sum_{i=0}^{\infty} a_i \varepsilon_{j-i},$$

$a_i = \frac{\Gamma(i+d-1)}{\Gamma(d-1)\Gamma(i+1)}$, $1/2 < d < 1$. Again, we do not assume that the innovations are Gaussian.

More details are discussed in Chapter 2.

1.2 SYMMETRIC GINI CORRELATION

In Chapter 3, we propose a symmetric Gini correlation. Let X and Y be two non-degenerate random variables with marginal distribution functions F and G, respectively, and a joint distribution function H. To describe the correlation between X and Y, the Pearson correlation (denoted as ρ_p) is probably the most frequently used measure. This measure is based on the covariance between two variables, which is optimal for describing the linear association between bivariate normal variables. However, the Pearson correlation performs poorly for variables with heavily-tailed or asymmetric distributions, and may be seriously impacted even by a single outlier (e.g., [42]). Under the assumption that F and G are continuous, the Spearman correlation, a robust alternative, is a multiple (twelve) of the covariance between the cumulative functions (or ranks) of two variables; the Gini correlation is based on the covariance between one variable and the cumulative distribution of the other ([2]). Two Gini correlations can be defined by

$$\gamma(X,Y) = \frac{\mathrm{cov}(X, G(Y))}{\mathrm{cov}(X, F(X))}, \quad \gamma(Y,X) = \frac{\mathrm{cov}(Y, F(X))}{\mathrm{cov}(Y, G(Y))} \tag{1.2.1}$$

to reflect different roles of X and Y. The representation of Gini correlation $\gamma(X,Y)$ indicates that it has mixed properties of those of the Pearson and Spearman correlations. It is similar to Pearson in X (the variable taken in its variate values) and similar to Spearman in Y (the variable taken in its ranks). Hence Gini correlations complement the Pearson and Spearman correlations ([36, 37, 38]). Two Gini correlations are equal if X and Y are exchangeable up to a linear transformation. However, Gini covariances are not symmetric in X and Y in general. On one hand, this asymmetrical nature is useful and can be used for testing bivariate exchangeability ([39]). On the other hand, such asymmetry violates the axioms of correlation measurement ([30]). Although some authors (e.g., [53]) dealt with asymmetry by a simple average $(\gamma(X,Y)+\gamma(Y,X))/2$, it is difficult to interpret this measure, especially when $\gamma(X,Y)$ and $\gamma(Y,X)$ have different signs. The asymmetry of $\gamma(X,Y)$ and $\gamma(Y,X)$ stems from the usage of marginal rank function $F(x)$ or $G(y)$. A remedy is to utilize a joint rank function. Based on this joint rank function, we are able to propose a symmetric Gini covariance (denoted as cov_g) and a corresponding symmetric correlation (denoted as ρ_g). That is, $\text{cov}_g(X,Y) = \text{cov}_g(Y,X)$ and $\rho_g(X,Y) = \rho_g(Y,X)$.

We study properties of the proposed Gini correlation ρ_g. In terms of the influence function, ρ_g is more robust than the Pearson correlation ρ_p. However, ρ_g is not as robust as the Spearman correlation and Kendall's τ correlation. Kendall's τ is another commonly used nonparametric measure of association. The Kendall correlation measure is more robust and more efficient than the Spearman correlation ([6]). For this reason in this dissertation we do not consider Spearman correlation for comparison.

As Kendall's τ has a relationship with the linear correlation ρ under elliptical distributions ([27, 29]), we also set up a function between ρ_g and ρ under elliptical distributions. This provides us an alternative method to estimate ρ based on estimation of ρ_g. The asymptotic normality of the estimator based on the symmetric Gini correlation is established. Its asymptotic efficiency and finite sample performance are compared with those of Pearson, Kendall's τ and the regular Gini correlation coefficients under various elliptical distributions.

As any quantity based on spatial ranks, ρ_g is only invariant under translation and homogeneous change. In order to gain the invariance property under heterogeneous changes, we provide an affine invariant version.

Simulations studies are conducted to compare our proposed symmetric Gini correlation with the Pearson correlation and the Kendall's τ, so we present the two current correlations below: the Pearson's correlation measure is

$$\rho_p(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}} = \frac{\mathbb{E}XY - \mathbb{E}X\mathbb{E}Y}{\sqrt{(\mathbb{E}X^2-(\mathbb{E}X)^2)(\mathbb{E}Y^2-(\mathbb{E}Y)^2)}}$$

and the Kendall's τ is

$$\tau(X,Y) = \mathbb{E}[\text{sign}\{(X_1 - X_2)(Y_1 - Y_2)\}],$$

where X_i and Y_i, $i = 1, 2$, are i.i.d copies of X and Y, respectively.

1.3 OVERVIEW

1.3 Contribution of the dissertation

The contribution of this dissertation is as follows:

- We derive the memory properties of nonlinear transformation of linear processes which are not necessarily Gaussian both in covariance sense and frequency domain. We conduct the simulation on the FARIMA(p, d, q) processes to confirm the theoretical results, and apply our results in econometrics and financial data analysis when the time series observations have non-Gaussian heavy tails. As an example, the memory properties of call option processes at different strike prices are discussed in details.

- We propose a symmetric Gini correlation and have studied its properties. When the scatter matrix $\mathbf{\Sigma}$ is homogeneous, the relationship between the proposed symmetric Gini correlation and the linear correlation is established. We also propose the affine invariant version of the symmetric Gini correlation to deal with the case hen $\mathbf{\Sigma}$ is heterogeneous. We calculate the influence function of the symmetric Gini correlation, which reveals that the proposed correlation is more robust than the Pearson correlation.

1.3 Dissertation Structure

The structure of this dissertation is organized in the following way. In Chapter 2 we discuss the memory properties of transformations of linear processes for both stationary and non-stationary processes. The conducted simulation study and the application to option processes in finance are illustrated to confirm the theoretical results. The symmetric Gini correlation is proposed in Chapter 3. We present properties of the new proposed correlation in terms of robustness and efficiency, and we also present a real data application of the proposed correlation in Chapter 3.

2 MEMORY PROPERTIES OF TRANSFORMATIONS OF LINEAR PROCESSES

2.1 PRELIMINARIES

In this section, we introduce some basic ideas and elements of time series analysis. The definitions and remarks below about time series are from [4] and readers can refer to [4] for more details about time series analysis.

2.1 Stationarity

Loosely speaking, stationary time series are those whose statistical properties remain constant over time. It plays a crucial role in the analysis of time series. In application, many observed time series, e.g., population of the U.S.A., strikes in the U.S.A., monthly accidental deaths in the U.S.A. and so on, are not stationary in appearance but such date sets can be transformed by some techniques into series which can reasonably be modelled as realizations of some stationary process. The techniques for transforming the non stationary observations into a stationary process are out of the scope of this dissertation, and the reader can refer to [4] for more details. The theory of stationary processes can be used for the analysis, fitting and prediction of the resulting series. So stationarity is one important topic in time series. The definition of stationarity is based on the following autocovariance function.

Definition 2.1.1 (The Autocovariance Function)**.** *If $\{X_t, t \in \mathbb{Z}\}$ is a process such that $var(X_t) < \infty$ for each $t \in \mathbb{Z}$, then the autocovariance function $\gamma_X(\cdot, \cdot)$ of $\{X_t\}$ is defined by*

$$\gamma_X(r, s) = cov(X_r, X_s) = \mathbb{E}[(X_r - \mathbb{E}X_r)(X_s - \mathbb{E}X_s)], \quad r, s \in \mathbb{Z}.$$

Definition 2.1.2 (Stationarity). *The time series* $\{X_t, t \in \mathbb{Z}\}$*, with index set* $\mathbb{Z} = \{0, \pm 1, \pm 2, ...\}$*, is said to be stationary if*

$$(i)\mathbb{E}|X_t|^2 < \infty \ \text{ for all } t \in \mathbb{Z},$$

$$(ii)\mathbb{E}X_t = m \ \text{ for all } t \in \mathbb{Z},$$

$$(iii)\gamma_X(r,s) = \gamma_X(r+t, s+t) \ \text{ for all } r,s,t \in \mathbb{Z}.$$

Another important and frequently used notion of stationarity is:

Definition 2.1.3 (Strict Stationarity). *The time series* $\{X_t, t \in \mathbb{Z}\}$ *is said to be strictly stationary if the joint distributions of* $(X_{t_1}, ..., X_{t_k})$ *and* $(X_{t_1+h}, ..., X_{t_k+h})$ *are the same for all positive integers* k *and for all* $t_1, ..., t_k, h \in \mathbb{Z}$*.*

Remark 2.1.4. The relation between stationarity and strict stationarity: Under the assumption of the existence of the second moment, a strictly stationary process is stationary; The converse of the statement is not true. For example, if $\{X_t\}$ is a sequence of independent random variables such that X_t is exponentially distributed with mean one when t is odd and normally distributed with mean one and variance one when t is even, then $\{X_t\}$ is stationary with $\gamma_X(0) = 1$ and $\gamma_X(h) = 0$ for $h \neq 0$. however since X_1 and X_2 have different distributions, $\{X_t\}$ cannot be strictly stationary.

Remark 2.1.5. In the literature, Definition 2.1.2 is often referred to as weak stationarity, covariance stationarity, stationarity in the wide sense or second-order stationarity. In this dissertation, the term stationarity will always refer to the properties specified by Definition 2.1.2.

Remark 2.1.6. For a stationary process $\{X_t, t \in \mathbb{Z}\}$, the autocovariance function can be redefined as the function of just one variable

$$\gamma_X(h) = \text{cov}(X_{t+h}, X_t) \ \text{ for all } t, h \in \mathbb{Z} \ .$$

2.1 FARIMA(p, d, q) Processes

The autoregressive moving average (ARMA) processes is an extremely important class of stationary time series.

Definition 2.1.7 (The ARMA(p, q) Process). *The process $\{X_t, t = 0, \pm 1, \pm 2, ...\}$ is said to be an ARMA(p, q) process if $\{X_t\}$ is stationary and if for every t,*

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \tag{2.1.8}$$

where $\{Z_t\}$ are i.i.d with mean 0 and standard deviation σ. We say that $\{X_t\}$ is an ARMA(p, q) process with mean μ if $\{X_t - \mu\}$ is an ARMA(p, q) process.

Symbolically, we can rewrite the equation (2.1.8) as

$$\phi(B)X_t = \theta(B)Z_t, \quad t = 0, \pm 1, \pm 2, ...,$$

where ϕ and θ are the p^{th} and q^{th} degree polynomials

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$$

and

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$$

and B is the backward shift operator defined by

$$B^j X_t = X_{t-j}, \quad j = 0, \pm 1, \pm 2, ...$$

The polynomials ϕ and θ will be referred to as the autoregressive and moving average polynomials, respectively.

The well-known moving average process and autoregressive process are two examples of ARMA processes.

Example 2.1.9 (The MA(q) Processes)**.** When $\phi(z) \equiv 1$, then (2.1.8) can be rewritten as

$$X_t = \theta(B)Z_t \tag{2.1.10}$$

which is said to be a moving average process of order q (MA(q)). Obviously, (2.1.10) has a unique solution.

Example 2.1.11 (The AR(p) Processes)**.** If $\theta(Z) \equiv 1$ then

$$\phi(B)X_t = Z_t \tag{2.1.12}$$

and the process is said to be an autoregressive process of order p (AR(p)). In order to explore the existence and uniqueness of the solution for (2.1.12), we do the following investigation for a simple case with $\phi(z) = 1 - \phi_1 z$, that is,

$$X_t = Z_t + \phi_1 X_{t-1}. \tag{2.1.13}$$

Iterating (2.1.13) we get

$$X_t = Z_t + \phi_1 Z_{t-1} + \cdots + \phi_1^k Z_{t-k} + \phi_1^{k+1} X_{t-k-1}.$$

If $|\phi_1| < 1$ we can conclude that

$$X_t = \sum_{j=1}^{\infty} \phi_1^j Z_{t-j},$$

which is a linear process.

For a general ARMA(p, q) process, we use the causality to keep the convergence of a linear process form.

Definition 2.1.14 (Causality). *An ARMA(p, q) process defined by the equation* $\phi(B)X_t = \theta(B)Z_t$ *is said to be causal (or more specifically to be a causal function of* $\{Z_t\}$*) if there exists a sequence of constants* $\{\psi_j\}$ *such that* $\sum_{j=0}^{\infty} |\psi_j| < \infty$ *and*

$$X_t = \sum_{j=1}^{\infty} \psi_j Z_{t-j}, \quad t = 0, \pm 1, ...$$

When we incorporate a wide range of non-stationary series, we obtain ARIMA processes which reduce to ARMA processes after differencing finitely many times.

Definition 2.1.15 (The ARIMA(p, d, q) Process). *If* d *is a non-negative integer, then* $\{X_t\}$ *is said to be an ARIMA(p, d, q) process if* $Y_t = (1-B)^d X_t$ *is a causal ARMA(p, q) process.*

If d can be some fractional number, we have the FARIMA processes.

Definition 2.1.16 (The FARIMA(p, d, q) process). *The model of an autoregressive fractionally integrated moving average process of a time series of order* (p, d, q)*, denoted by FARIMA*(p, d, q) *is defined as*

$$\phi(B)X_t = \theta(B)(1-B)^{-d}\varepsilon_t. \tag{2.1.17}$$

Here, $-1 < d < 1/2$*,* p, q *are nonnegative integers,* $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ *is the AR polynomial and* $\theta(z) = 1 + \theta_1 z + \cdots \theta_q z^q$ *is the MA polynomial.*

Remark 2.1.18. For the Definition 2.1.16, under the conditions that $\phi(z)$ and $\theta(z)$ have no common zeros, the zeros of $\phi(\cdot)$ lie outside the closed unit disk and $-1 < d < 1/2$, the FARIMA(p, d, q) process has linear process form

$$X_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}, \quad t \in \mathbb{N},$$

with $a_i = \frac{\theta(1)}{\phi(1)} \frac{i^{d-1}}{\Gamma(d)} + \mathcal{O}(i^{-1})$ and $\mathcal{O}(i^{-1})/i^{-1}$ means all the values satisfying that $\mathcal{O}(i^{-1})/i^{-1}$ is bounded. .

2.1 Long-memory Processes

In this section we present a particular class of linear time series: *long-memory* or *long-range-dependent-processes*. Long-memory (long-range dependent) processes can be defined in many ways. Here we only focus our attention on the following two definitions in the covariance sense and in the frequency domain, respectively.

Definition 2.1.19. *Let $\gamma_X(h)$ be the autocovariance function at lag h of the stationary process $\{X_n : n \in \mathbb{Z}\}$. If $\sum_{h=-\infty}^{\infty} |\gamma_X(h)| = \infty$, then we say X_n has long memory in the covariance sense; otherwise, X_n has short memory in the covariance sense.*

There are alternatives definitions to Definition 2.1.19. In particular, long memory can be defined by specifying a hyperbolic decay of the autocovariances

$$\gamma_X(h) \sim h^{2d-1} L_1(h), \quad \text{as } h \to \infty,$$

where d is the long-memory parameter and $L_1(\cdot) > 0$ is a slowly varying function, i.e., $\lim_{x\to\infty} L(\lambda x)/L(x) = 1$ for any $\lambda > 0$.

Definition 2.1.20. *In the frequency domain, a stationary time series X_n with a spectral density function $f(\lambda)$ is called a long memory process in a restricted spectral density sense if $f(\lambda)$ is bounded on $[\delta, \pi]$ for every $\delta > 0$, and $f(\lambda) \to \infty$ as $\lambda \to 0^+$.*

The spectral density function $f(\cdot)$ is defined as

Definition 2.1.21. *If a time series $\{X_t\}$ has auto covariance γ satisfying $\sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty$, then we define its spectral density as*

$$f(\lambda) = \sum_{h=-\infty}^{\infty} \gamma_X(h) e^{-i2\pi\lambda h}$$

for $-\infty < \lambda < \infty$.

Remark 2.1.22. These two definitions 2.1.19, 2.1.21 are not always equivalent. In particular, for $d < 1/2$, [9] called a stationary time series $X_n \sim \mathrm{LM}(d)$ if the spectral density function $f(\lambda)$ behaves like a power function at low frequencies, that is as $|\lambda|^{-2d}$ as λ approaches zero. For $d \geq 1/2$, $X_n \sim \mathrm{LM}(d)$ if and only if $(1-B)^k X_n \sim \mathrm{LM}(d-k)$ for $k = [d+1/2]$, where $[x]$ denotes the largest integer smaller or equal to x. B is the backward shift operator, $BX_i = X_{i-1}$. The cases $d > 0$, $d = 0$ and $d < 0$ correspond to long memory, short memory and negative dependence (antipersistence), respectively.

Throughout this chapter, the innovations ε_i, $i \in \mathbb{Z}$, of the linear process (1.1.3) are i.i.d. random variables with mean zero and finite variances. Without loss of generality, we assume that $\mathbb{E}\varepsilon_i^2 = 1$, $i \in \mathbb{Z}$. The coefficients a_i satisfy $\sum_{i=0}^{\infty} a_i^2 < \infty$, under which the linear process (1.1.3) is well defined by the three series theorem. $||X|| = [\mathbb{E}(X^2)]^{1/2}$ is the L^2 norm of the random variable X. Define the shift process $\mathcal{F}_i = (\ldots, \varepsilon_{i-1}, \varepsilon_i)$, and let

$$\begin{aligned}
X_{n,k} &= \mathbb{E}(X_n|\mathcal{F}_k), \\
K_n(w) &= \mathbb{E}[K(w + X_n - X_{n,0})], \\
K_\infty(w) &= \mathbb{E}[K(w + X_n)], \\
K_n^{(r)}(w) &= \frac{d^r}{dw^r}\mathbb{E}[K(w + X_n - X_{n,0})], \\
K_\infty^{(r)}(w) &= \frac{d^r}{dw^r}\mathbb{E}[K(w + X_n)]
\end{aligned}$$

for any nonnegative integer r. Part of our results will be related with the following definition is from [20].

Definition 2.1.23. *A transformation $K(\cdot)$ has power rank k with respect to the linear process X_n for some positive integer k if $K_\infty^{(k)}(0) \neq 0$ and $K_\infty^{(r)}(0) = 0$ for all $1 \leq r < k$.*

We will use k to denote the power rank of $K(\cdot)$ with respect to the linear process X_n throughout this dissertation. One of our assumptions on the transformation function $K(\cdot)$ is obtained from [51].

Condition 2.1.24. *Let* $\mathbb{E}(|\varepsilon_1|^q) < \infty$ *for some* $2 < q \leq 4$ *and* $K_n \in \mathbb{C}^{k+1}(\mathbb{R})$ *for all large* n. *Assume that for some* $\lambda > 0$,

$$\sum_{\alpha=0}^{k+1} \|K_{n-1}^{(\alpha)}(X_{n,0};\lambda)\| + \sum_{\alpha=0}^{k-1} \||\varepsilon_1|^{q/2} K_{n-1}^{(\alpha)}(X_{n,1})\| + \|\varepsilon_1 K_{n-1}^{(k)}(X_{n,1})\| = \mathcal{O}(1),$$

where $K_{n-1}^{(\alpha)}(X_{n,0};\lambda) = \sup_{|y|\leq\lambda}|K_{n-1}^{(\alpha)}(X_{n,0}+y)|$ is the local maximal function for $K_{n-1}^{(\alpha)}(X_{n,0})$. As [51] mentioned, Condition 2.1.24 is quite mild, which only imposes certain smoothness requirements on K_{n-1}. Additional notations are defined as follow: we use $a_m \sim b_m$ instead of the notation $a_m/b_m \to 1$; for positive sequences, the notation $a_m \ll b_m$ or $b_m \gg a_m$ and the Vinogradov symbol $\mathcal{O}$ mean that a_m/b_m is bounded; the notation $a_m \simeq b_m$ means that there exist constants c_1 an c_2 such that $0 < c_1 b_m < a_m < c_2 b_m$ for m large enough. $C > 0$ is a generic constant which may vary in different context.

2.2 TRANSFORMATIONS OF STATIONARY PROCESSES

We first consider the transformation $K(X_n)$ of the stationary process (1.1.3). First of all, $K(X_n)$, $n \in \mathbb{N}$, is strictly stationary since the time series X_n is strictly stationary. By the condition $\mathbb{E}K^2(X_n) < \infty$, $K(X_n)$ is also (covariance) stationary.

2.2 Transformations of Long-memory Processes

We consider the case $a_0 = 1$ and $a_i = i^{-\beta}L(i), i > 0$, $1/2 < \beta < 1$, for the linear process (1.1.3). Notice that in this case the covariance function $\gamma_X(h) = \mathbb{E}X_0X_h$ of the original series X_n is regularly varying with exponent $-1 < 1-2\beta < 0$ and hence X_n has long memory in the covariance sense. The FARIMA(p,d,q) process as in (2.1.17) with $0 < d < 1/2$ is a particular example of this case, $\beta = 1-d$.

Theorem 2.2.1. *Assume that Condition 2.1.24 holds with $q = 4$ and that K has power rank $k \geq 1$. Let $a_0 = 1$ and $a_i = i^{-\beta} L(i), i > 0$, $1/2 < \beta < 1$, in model (1.1.3). If the power rank k of a transformation $K(\cdot)$ with respect to the linear process (1.1.3) satisfies $k < (2\beta - 1)^{-1}$, then $K(X_n)$ has long memory in the covariance sense. $K(X_n)$ has short memory in the covariance sense if $k > (2\beta - 1)^{-1}$.*

Proof. Define the projection operator

$$\mathcal{P}_i X = \mathbb{E}(X|\mathcal{F}_i) - \mathbb{E}(X|\mathcal{F}_{i-1}).$$

We adopt the notations from [51] as follows: For $j \geq 2$, let $A_n(j) = \sum_{t=n}^{\infty} |a_t|^j$, $\theta_n = |a_{n-1}|[|a_{n-1}| + A_n^{1/2}(4) + A_n^{k/2}(2)]$.

Recall that $K(\cdot)$ is a measurable function with the power rank of k. By [20], $K(X_n) - \mathbb{E}K(X_n)$ can be decomposed as $U(\mathcal{F}_n) + S(\mathcal{F}_n)$, where

$$U(\mathcal{F}_n) = K_\infty^{(k)}(0) \sum_{0 \leq j_1 < j_2 < \cdots < j_k < \infty} \prod_{s=1}^{k} a_{j_s} \varepsilon_{n-j_s}$$

and

$$S(\mathcal{F}_n) = K(X_n) - \mathbb{E}K(X_n) - U(\mathcal{F}_n).$$

We also have the decomposition $U(\mathcal{F}_{n+h}) + S(\mathcal{F}_{n+h})$ for $K(X_{n+h})$. Therefore, $\operatorname{cov}(K(X_n), K(X_{n+h}))$ can be represented by

$$\begin{aligned} &\operatorname{cov}(U(\mathcal{F}_n), U(\mathcal{F}_{n+h})) + \operatorname{cov}(U(\mathcal{F}_n), S(\mathcal{F}_{n+h})) \\ &+ \operatorname{cov}(S(\mathcal{F}_n), U(\mathcal{F}_{n+h})) + \operatorname{cov}(S(\mathcal{F}_n), S(\mathcal{F}_{n+h})). \end{aligned} \tag{2.2.1}$$

We first find the bounds of $\text{cov}(U(\mathcal{F}_n), U(\mathcal{F}_{n+h}))$, which are useful in the proofs of Theorem 2.2.1, Corollary 2.2.2 and Theorem 2.2.3. By Stirling's approximation, $\begin{pmatrix} h \\ k \end{pmatrix} \simeq h^k/k!$, if h is large enough. If $j_k \leq h$, the quantity $\prod_{s=1}^{k} L(j_s) j_s^{-\beta} L(h+j_s)(h+j_s)^{-\beta}$ at least has order $h^{-2k\beta} L^k(h) \min_{1 \leq s \leq h} L^k(s)$. Therefore for the lower bound,

$$\begin{aligned}
&\text{cov}(U(\mathcal{F}_n), U(\mathcal{F}_{n+h})) = [K_\infty^{(k)}(0)]^2 \sum_{0 \leq j_1 < j_2 < \cdots < j_k < \infty} \prod_{s=1}^{k} a_{j_s} a_{h+j_s} \qquad (2.2.2)\\
&\geq [K_\infty^{(k)}(0)]^2 \sum_{1 \leq j_1 < j_2 < \cdots < j_k \leq h} \prod_{s=1}^{k} L(j_s) j_s^{-\beta} L(h+j_s)(h+j_s)^{-\beta}\\
&\gg [K_\infty^{(k)}(0)]^2 \begin{pmatrix} h \\ k \end{pmatrix} h^{-2k\beta} L^k(h) \min_{1 \leq s \leq h} L^k(s)\\
&\simeq (k!)^{-1} [K_\infty^{(k)}(0)]^2 h^k h^{-2k\beta} L^k(h) \min_{1 \leq s \leq h} L^k(s)\\
&= (k!)^{-1} [K_\infty^{(k)}(0)]^2 h^{k(1-2\beta)} L^k(h) \min_{1 \leq s \leq h} L^k(s).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\text{cov}(U(\mathcal{F}_n), U(\mathcal{F}_{n+h})) \qquad (2.2.3)\\
&\leq [K_\infty^{(k)}(0)]^2 (\sum_{0 \leq i < \infty} a_i a_{h+i})^k\\
&= [K_\infty^{(k)}(0)]^2 (\sum_{i=0}^{h} a_i a_{h+i} + \sum_{i=h+1}^{\infty} a_i a_{h+i})^k\\
&= \mathcal{O}[h^{k(1-2\beta)} L^{2k}(h)].
\end{aligned}$$

We shall estimate the other covariances in (2.2.1) for each case. In the case that $k(2\beta - 1) < 1$, we first apply the projection operator to the terms in the covariances and then apply the

Cauchy-Schwarz inequality. Then

$$
\begin{aligned}
&|\text{cov}(S(\mathcal{F}_n), S(\mathcal{F}_{n+h}))| \\
&= |\sum_{i=-\infty}^{n-1} \sum_{j=-\infty}^{n+h-1} \text{cov}(\mathcal{P}_{i+1}S(\mathcal{F}_n), \mathcal{P}_{j+1}S(\mathcal{F}_{n+h}))| \\
&= |\sum_{i=-\infty}^{n-1} \sum_{j=-\infty}^{n+h-1} \mathbb{E}[\mathcal{P}_{i+1}S(\mathcal{F}_n)\mathcal{P}_{j+1}S(\mathcal{F}_{n+h})]| \\
&= |\sum_{i=-\infty}^{n-1} \mathbb{E}[\mathcal{P}_{i+1}S(\mathcal{F}_n)\mathcal{P}_{i+1}S(\mathcal{F}_{n+h})]| \qquad (2.2.4) \\
&\leq \sum_{i=-\infty}^{n-1} \|\mathcal{P}_{i+1}S(\mathcal{F}_n)\| \|\mathcal{P}_{i+1}S(\mathcal{F}_{n+h})\| \\
&= \sum_{i=-\infty}^{n-1} \|\mathcal{P}_1 S(\mathcal{F}_{n-i})\| \|\mathcal{P}_1 S(\mathcal{F}_{n+h-i})\| \\
&= \sum_{i=-\infty}^{n-1} \mathcal{O}(\theta_{n-i}\theta_{n+h-i}) = \sum_{i=1}^{\infty} \mathcal{O}(\theta_i \theta_{i+h}). \qquad (2.2.5)
\end{aligned}
$$

Equality (2.2.4) is true because if $i \neq j$, suppose $i < j$, then

$$
\begin{aligned}
&\mathbb{E}[\mathcal{P}_{i+1}S(\mathcal{F}_n)\mathcal{P}_{j+1}S(\mathcal{F}_{n+h})] = \mathbb{E}\big\{\mathbb{E}[\mathcal{P}_{i+1}S(\mathcal{F}_n)\mathcal{P}_{j+1}S(\mathcal{F}_{n+h})|\mathcal{F}_{i+1}]\big\} \\
&= \mathbb{E}\big\{\mathcal{P}_{i+1}S(\mathcal{F}_n)\mathbb{E}[\mathcal{P}_{j+1}S(\mathcal{F}_{n+h})|\mathcal{F}_{i+1}]\big\} \\
&= \mathbb{E}\big\{\mathcal{P}_{i+1}S(\mathcal{F}_n)\mathbb{E}\big[(\mathbb{E}[S(\mathcal{F}_{n+h})|\mathcal{F}_{j+1}] - \mathbb{E}[S(\mathcal{F}_{n+h})|\mathcal{F}_j])|\mathcal{F}_{i+1}\big]\big\} \\
&= \mathbb{E}\big\{\mathcal{P}_{i+1}S(\mathcal{F}_n)\big(\mathbb{E}[S(\mathcal{F}_{n+h})|\mathcal{F}_{i+1}] - \mathbb{E}[S(\mathcal{F}_{n+h})|\mathcal{F}_{i+1}]\big)\big\} = 0.
\end{aligned}
$$

Equality (2.2.5) is the result of Theorem 5 (Reduction principle) of [51]. By Karamata's theorem ([40]), $A_n(j) = \mathcal{O}[n^{1-j\beta}L^j(n)]$ for $j \geq 2$. Therefore, under the condition $k(2\beta-1) <$

1,

$$
\begin{aligned}
&\sum_{i=1}^{\infty}\mathcal{O}(\theta_i\theta_{i+h})\\
&=\sum_{i=0}^{\infty}a_ia_{i+h}\mathcal{O}\big\{[(i+1)^{1-2\beta}L^2(i+1)(i+h+1)^{1-2\beta}L^2(i+h+1)]^{k/2}\big\}\\
&=\sum_{i=1}^{\infty}\mathcal{O}[i^{-\beta+k(1-2\beta)/2}(i+h)^{-\beta+k(1-2\beta)/2}L^{k+1}(i)L^{k+1}(i+h)]\\
&=\sum_{i=1}^{h}\mathcal{O}[i^{-\beta+k(1-2\beta)/2}h^{-\beta+k(1-2\beta)/2}(1+i/h)^{-\beta+k(1-2\beta)/2}L^{k+1}(i)L^{k+1}(i+h)]\\
&+\sum_{i=h+1}^{\infty}\mathcal{O}[i^{-2\beta+k(1-2\beta)}(1+h/i)^{-\beta+k(1-2\beta)/2}L^{k+1}(i)L^{k+1}(i+h)].
\end{aligned}
$$

Applying Karamata's theorem again, we have

(i) $\sum_{i=1}^{\infty}\mathcal{O}(\theta_i\theta_{i+h})=\mathcal{O}[L^{k+1}(h)h^{\frac{k(1-2\beta)-2\beta}{2}}]$ if $(k+1)(2\beta-1)>1$;

(ii)$\sum_{i=1}^{\infty}\mathcal{O}(\theta_i\theta_{i+h})=\mathcal{O}[L^{2k+2}(h)h^{(k+1)(1-2\beta)}]$ if $(k+1)(2\beta-1)<1$;

(iii) $\sum_{i=1}^{\infty}\mathcal{O}(\theta_i\theta_{i+h}) = \max\big\{\mathcal{O}[L^{2k+2}(h)h^{-1}],\mathcal{O}[L^{k+1}(h)h^{-1}\sum_{i=1}^{h}i^{-1}L^{k+1}(i)]\big\}$ if $(k+1)(2\beta-1)=1$.

By the calculation in (2.2.2), each of the above terms (i), (ii) and (iii) is less than $\mathrm{cov}(U(\mathcal{F}_n),U(\mathcal{F}_{n+h}))$. Thus,

$$|\mathrm{cov}(S(\mathcal{F}_n),S(\mathcal{F}_{n+h}))|<\mathrm{cov}(U(\mathcal{F}_n),U(\mathcal{F}_{n+h})).$$

With the same arguments as in $|\mathrm{cov}(S(\mathcal{F}_n),S(\mathcal{F}_{n+h}))|$,

$$|\mathrm{cov}(U(\mathcal{F}_n),S(\mathcal{F}_{n+h}))|=\sum_{i=1}^{\infty}\|\mathcal{P}_1U(\mathcal{F}_i)\|\mathcal{O}(\theta_{i+h}).$$

Obviously, $\mathcal{O}(\theta_{i+h}) = \mathcal{O}[(i+h)^{-\beta+\frac{k(1-2\beta)}{2}} L^{k+1}(h+i)]$. Also

$$\begin{aligned}
&\mathcal{P}_1 U(\mathcal{F}_i)/K_\infty^{(k)}(0)\\
&= \mathbb{E}\big(\sum_{0\le j_1<\cdots<j_k<\infty} \prod_{s=1}^{k} a_{j_s}\varepsilon_{i-j_s}|\mathcal{F}_1\big) - \mathbb{E}\big(\sum_{0\le j_1<\cdots<j_k<\infty} \prod_{s=1}^{k} a_{j_s}\varepsilon_{i-j_s}|\mathcal{F}_0\big)\\
&= \sum_{i-1\le j_1<\cdots<j_k<\infty} \prod_{s=1}^{k} a_{j_s}\varepsilon_{i-j_s} - \sum_{i\le j_1<\cdots<j_k<\infty} \prod_{s=1}^{k} a_{j_s}\varepsilon_{i-j_s}\\
&= a_{i-1}\varepsilon_1 \sum_{i\le j_2<\cdots<j_k<\infty} \prod_{s=2}^{k} a_{j_s}\varepsilon_{i-j_s}.
\end{aligned}$$

Then

$$\begin{aligned}
&||\mathcal{P}_1 U(\mathcal{F}_i)||^2 = \mathbb{E}[\mathcal{P}_1 U(\mathcal{F}_i)]^2\\
&= [K_\infty^{(k)}(0)]^2 a_{i-1}^2 \mathbb{E}(\varepsilon_1^2)\mathbb{E}\big(\sum_{i\le j_2<\cdots<j_k<\infty} \prod_{s=2}^{k} a_{j_s}\varepsilon_{i-j_s}\big)^2\\
&\le [K_\infty^{(k)}(0)]^2 a_{i-1}^2 A_i^{k-1}(2)[\mathbb{E}(\varepsilon_1^2)]^k.
\end{aligned}$$

See also [51]. Hence

$$||\mathcal{P}_1 U(\mathcal{F}_i)|| = \mathcal{O}[i^{-\beta+(k-1)(1-2\beta)/2} L^k(i)].$$

In consequence, using Karamata's theorem, we have

$$\begin{aligned}
&\sum_{i=1}^{\infty} ||\mathcal{P}_1 U(\mathcal{F}_i)|| \mathcal{O}(\theta_{i+h}) \\
&= \sum_{i=1}^{\infty} \mathcal{O}\big[(i+h)^{-\beta+\frac{k(1-2\beta)}{2}} L^{k+1}(h+i) i^{-\beta+(k-1)(1-2\beta)/2} L^k(i)\big] \\
&= h^{-\beta+\frac{k(1-2\beta)}{2}} \sum_{i=1}^{h} \mathcal{O}\big[(1+i/h)^{-\beta+\frac{k(1-2\beta)}{2}} L^{k+1}(h+i) i^{-\beta+(k-1)(1-2\beta)/2} L^k(i)\big] \\
&+ \sum_{i=h+1}^{\infty} \mathcal{O}\big[i^{-2\beta+(2k-1)(1-2\beta)/2}(1+h/i)^{-\beta+\frac{k(1-2\beta)}{2}} L^{k+1}(h+i) L^k(i)\big] \\
&= \mathcal{O}[h^{(1-2\beta)(k+1/2)} L^{2k+1}(h)],
\end{aligned}$$

which is less than $\text{cov}(U(\mathcal{F}_n), U(\mathcal{F}_{n+h}))$. So

$$|\text{cov}(U(\mathcal{F}_n), S(\mathcal{F}_{n+h}))| < \text{cov}(U(\mathcal{F}_n), U(\mathcal{F}_{n+h})).$$

Similarly, $|\text{cov}(U(\mathcal{F}_{n+h}), S(\mathcal{F}_n))| = \sum_{i=1}^{\infty} \|\mathcal{P}_1 U(\mathcal{F}_{i+h})\| \mathcal{O}(\theta_i)$, which is (i) $\mathcal{O}\big[h^{(k+\frac{1}{2})(1-2\beta)} L^{2k+1}(h)\big]$ if $(k+1)(2\beta-1) < 1$; (ii) $\mathcal{O}\big[h^{-\beta+\frac{k-1}{2}(1-2\beta)} L^k(h)\big]$ if $(k+1)(2\beta-1) > 1$; (iii) $\max\big\{\mathcal{O}[h^{-\frac{2k+1}{2(k+1)}} L^k(h) \sum_{i=1}^{h} i^{-1} L^{k+1}(i)], \mathcal{O}[h^{-\frac{2k+1}{2(k+1)}} L^{2k+1}(h)]\big\}$ if $(k+1)(2\beta-1) = 1$. Each of the terms (i), (ii) and (iii) is less than $\text{cov}(U(\mathcal{F}_n), U(\mathcal{F}_{n+h}))$ by the analysis in (2.2.2). Hence $|\text{cov}(U(\mathcal{F}_{n+h}), S(\mathcal{F}_n))| < \text{cov}(U(\mathcal{F}_n), U(\mathcal{F}_{n+h}))$.

So under the condition $k < (2\beta-1)^{-1}$,

$$\text{cov}(K(X_n), K(X_{n+h})) \simeq cov(U(\mathcal{F}_n), U(\mathcal{F}_{n+h})).$$

But by (2.2.2),

$$\text{cov}(U(\mathcal{F}_n), U(\mathcal{F}_{n+h})) \geq (k!)^{-1}[K_{\infty}^{(k)}(0)]^2 h^{k(1-2\beta)} L^k(h) \min_{1\leq s\leq h} L^k(s),$$

which is not summable. Therefore $K(X_n)$ has long memory in the covariance sense if $k(2\beta - 1) < 1$.

Now, we consider the case $k > (2\beta - 1)^{-1}$. With similar arguments as the case $k(2\beta - 1) < 1$, we have

$$
\begin{aligned}
|\mathrm{cov}(U(\mathcal{F}_n), S(\mathcal{F}_{n+h}))| &= \mathcal{O}[h^{-2\beta} L^2(h)] \quad \text{if } (k-1)(2\beta-1) > 1, \\
|\mathrm{cov}(U(\mathcal{F}_n), S(\mathcal{F}_{n+h}))| &= \mathcal{O}\big\{h^{-2\beta} \max[L^2(h), L^{k+1}(h)]\big\} \quad \text{if } (k-1)(2\beta-1) = 1, \\
|\mathrm{cov}(U(\mathcal{F}_n), S(\mathcal{F}_{n+h}))| &= \mathcal{O}[h^{-\beta+k(1-2\beta)/2} L^{k+1}(h)] \quad \text{if } (k-1)(2\beta-1) < 1; \qquad (2.2.6)
\end{aligned}
$$

$$
\begin{aligned}
|\mathrm{cov}(S(\mathcal{F}_n), S(\mathcal{F}_{n+h}))| &= \mathcal{O}[h^{-2\beta} L^2(h)] \quad \text{if } (k-1)(2\beta-1) > 1, \\
|\mathrm{cov}(S(\mathcal{F}_n), S(\mathcal{F}_{n+h}))| &= \mathcal{O}\{h^{-2\beta} \max[L^2(h), L^{k+1}(h)]\} \quad \text{if } (k-1)(2\beta-1) = 1, \\
|\mathrm{cov}(S(\mathcal{F}_n), S(\mathcal{F}_{n+h})| &= \mathcal{O}[h^{-\beta+(1-2\beta)k/2} L^{k+1}(h)] \quad \text{if } (k-1)(2\beta-1) < 1; \qquad (2.2.7)
\end{aligned}
$$

and

$$
|\mathrm{cov}(S(\mathcal{F}_n), U(\mathcal{F}_{n+h}))| = \mathcal{O}[h^{-\beta+(k-1)(1-2\beta)/2} L^k(h)], \qquad (2.2.8)
$$

which are all summable. Additionally, (2.2.3) is also summable in this case $k(2\beta - 1) > 1$. Therefore $K(X_n)$ has short memory in the covariance sense if $k(2\beta - 1) > 1$.

□

This theorem shows that $K(X_n)$ has long memory as long as the power rank of $K(\cdot)$ satisfies $k < (2\beta - 1)^{-1}$. Hence $K(X_n)$ keeps the long memory property for a wide range (in terms of the power rank k) of transformations if the parameter β of the original series X_n is close to 1/2 and therefore X_n has very strong long memory. Nevertheless, $K(X_n)$ losses the long memory property for a wide range of transformations if β is not close to 1/2. For example, if $3/4 < \beta < 1$, only X_n and other transformations with power rank $k = 1$ keep the long memory property.

Remark 2.2.9. [19, 20] studied the limit theorems by assuming $\mathbb{E}K^2(X_n) < \infty$, $\mathbb{E}\varepsilon^8 < \infty$ and the condition $C(t, \tau, \lambda)$ there. [51] studied the functional limit theorems under the improved condition $\mathbb{E}\varepsilon^4 < \infty$ and the Condition 2.1.24. The memory property in Theorem 2.2.1 is consistent with the above limit theorems via the order of normalization. Theorem 2.2.1 only requires the Condition 2.1.24 with $q = 4$.

Remark 2.2.10. It is well known that both long memory and heavy tail parameters play roles in the asymptotic behaviors of partial sums of time series. See e.g.,[31]. As in [51], for the major results in this chapter, we assume that the innovation of the linear process has fourth moment, although not necessary Gaussian. Therefore, the memory parameter dominates the growth of the partial sum, which is the case in the upper right hand region of Figure 1 in [31].

The next corollary shows that if the slowly varying function $L(x)$ is a constant asymptotically, $K(X_n)$ has long memory also in the case that $(2\beta-1)^{-1}$ is an integer and the power rank $k = (2\beta - 1)^{-1}$.

Corollary 2.2.2. *In the case that* $\lim_{n\to\infty} L(n) = L$ *for some constant* $L > 0$*, under the same conditions as in Theorem 2.2.1,* $K(X_n)$ *has long memory in the covariance sense if* $k \leq (2\beta - 1)^{-1}$*.* $K(X_n)$ *has short memory in the covariance sense if* $k > (2\beta - 1)^{-1}$*.*

Proof. We just need consider the case that $(2\beta - 1)^{-1}$ is an integer and the power rank $k = (2\beta - 1)^{-1}$. In this case, since $\lim_{n\to\infty} L(n) = L$, by (2.2.2), we have

$$\mathrm{cov}(U(\mathcal{F}_n), U(\mathcal{F}_{n+h})) \gg (k!)^{-1}[K_\infty^{(k)}(0)]^2 h^{-1}.$$

Hence $K(X_n)$ has long memory in the covariance sense in this case.

□

Corollary 2.2.2 is applicable to FARIMA(p, d, q) process. Recall that in this case, $\beta = 1 - d$ and $\lim_{n\to\infty} L(n) = \frac{\theta(1)}{\phi(1)\Gamma(d)}$. Furthermore, we have detailed knowledge on the memory parameter of $K(X_n)$ from the following Theorem 2.2.3 if the linear process is a FARIMA(p, d, q) process.

Theorem 2.2.3. *Let X_n be a stationary FARIMA(p, d, q) process (2.1.17) with $0 < d < 1/2$ and Condition 2.1.24 holds with $q = 4$. $K(\cdot)$ has power rank k with respect to the FARIMA(p,d,q) process. Then $K(X_n)$ is a long-memory process LM$(\tilde{d})$ with $\tilde{d} = (d - 1/2)k + 1/2$ when $k(1 - 2d) < 1$, and a short-memory process LM(0) if $k(1 - 2d) > 1$ but $(k - 1)(1 - 2d) < 1$.*

Proof. Since $\lim_{j\to\infty} \frac{a_j}{j^{d-1}} = \frac{\theta(1)}{\phi(1)\Gamma(d)}$, we have $\beta = 1 - d$ and the slowly varying function is a constant asymptotically. Hence, applying (2.2.2) and (2.2.3) yields

$$\operatorname{cov}(U(\mathcal{F}_n), U(\mathcal{F}_{n+h})) \simeq h^{k(1-2\beta)} = h^{k(2d-1)} = h^{2\bar{d}-1} \tag{2.2.11}$$

with $\bar{d} = (d - 1/2)k + 1/2 > 0$.

We first consider the case $k < (2\beta - 1)^{-1}$. By (2.2.11) and the proof of Theorem 2.2.1, $\operatorname{cov}(K(X_n), K(X_{n+h})) \simeq \operatorname{cov}(U(\mathcal{F}_n), U(\mathcal{F}_{n+h})) \simeq h^{k(1-2\beta)}$. Then by the same argument as in Proposition 1 of [9], $K(X_n)$ is a long-memory process LM$(\tilde{d})$ when $k(2\beta - 1) < 1$.

In the case that $k(2\beta - 1) > 1$ and $(k - 1)(2\beta - 1) < 1$, from (2.2.11), (2.2.6), (2.2.7) and (2.2.8), $\operatorname{cov}(K(X_n), K(X_{n+h}))$ is dominated by $\operatorname{cov}(U(\mathcal{F}_n), U(\mathcal{F}_{n+h}))$. So,

$$\operatorname{cov}(K(X_n), K(X_{n+h})) \simeq h^{2\bar{d}-1}$$

with $\bar{d} = (d - 1/2)k + 1/2 < 0$. Therefore the process $K(X_n)$ has the same autocorrelation delay pattern as an FARIMA$(0, \bar{d}, 0)$ process. But we shall show that it is a short-memory LM(0) process. Denote $f_K(\lambda)$ as the spectral density of $K(X_n)$. Since $\operatorname{cov}(K(X_n), K(X_{n+h}))$

is dominated by $\operatorname{cov}(U(\mathcal{F}_n), U(\mathcal{F}_{n+h}))$, which is positive and summable, then

$$0 < f_K(0) = \operatorname{var}(K(X_n)) + 2\sum_{h=1}^{\infty} \operatorname{cov}(K(X_n), K(X_{n+h})) < \infty.$$

Therefore $K(X_n)$ is a LM(0) process. □

This theorem shows that $K(X_n)$ can never have stronger long range dependence than the original process since $\tilde{d} \leq d$ and $\tilde{d} = d$ if and only if $k = 1$.

2.2 Transformations of Short-memory Processes

Now we study the transformations of short memory linear processes in the form of (1.1.3). The following theorem provides a result in the general setting $\sum_{i=0}^{\infty} |a_i| < \infty$.

Theorem 2.2.4. *Assume $\sum_{i=0}^{\infty} |a_i| < \infty$ in the model (1.1.3) and*

$$\|K_{n-1}(X_{n,1}) - K_{n-1}(X_{n,0})\| = \mathcal{O}(|a_{n-1}|). \tag{2.2.12}$$

Then $K(X_n)$ has short memory in the covariance sense for any transformation $K(\cdot)$ with $\mathbb{E}K^2(X_n) < \infty$.

Proof. Again, using the projection operator and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\operatorname{cov}(K(X_n), K(X_{n+h})) = \operatorname{cov}(\sum_{i=-\infty}^{n-1} \mathcal{P}_{i+1}K(X_n), \sum_{j=-\infty}^{n+h-1} \mathcal{P}_{j+1}K(X_{n+h})) \\
&= \sum_{i=-\infty}^{n-1} \mathbb{E}[\mathcal{P}_{i+1}K(X_n)\mathcal{P}_{i+1}K(X_{n+h})] \leq \sum_{i=-\infty}^{n-1} \|\mathcal{P}_{i+1}K(X_n)\|\|\mathcal{P}_{i+1}K(X_{n+h})\| \\
&= \sum_{i=-\infty}^{n-1} \|\mathcal{P}_1K(X_{n-i})\|\|\mathcal{P}_1K(X_{n+h-i})\| \\
&= \sum_{i=-\infty}^{n-1} \mathcal{O}(|a_{n-i-1}a_{n+h-i-1}|) \qquad (2.2.13) \\
&= \sum_{i=0}^{\infty} \mathcal{O}(|a_i a_{i+h}|),
\end{aligned}$$

where equality (2.2.13) is obtained from [51]: $\|\mathcal{P}_1 K(X_n)\| = \mathcal{O}(|a_{n-1}|)$, if condition (2.2.12) holds. Thus

$$\sum_{h=1}^{\infty} \operatorname{cov}(K(X_n), K(X_{n+h})) = \sum_{h=1}^{\infty}\sum_{i=0}^{\infty} \mathcal{O}(|a_i a_{i+h}|),$$

which is finite. This finishes the proof.

□

As Condition 2.1.24, the condition (2.2.12) is proposed by [51] and only requires certain smoothness requirements on K_{n-1}. Theorem 2.2.4 shows that we can never get long memory process from transformations if the original process has short memory.

It is an open question that whether $K(X_n)$ is a short memory LM(0) process if X_n is a stationary FARIMA(p, d, q) process with $-1 < d < 0$. The following Theorem 2.2.5 gives a confirmative answer to this question for the special case, the FARIMA$(0, d, 0)$ process with $-1 < d < 0$, if $K(x) = x^2$. The FARIMA$(0, d, 0)$ is not necessarily Gaussian.

Theorem 2.2.5. *Let X_n be a stationary FARIMA$(0, d, 0)$ process, $-1 < d < 0$, defined as in (1.1.1). Then X_n^2 is a short-memory process LM(0).*

Proof. Denote $f_K(\lambda)$ as the spectral density of $K(X_n) = X_n^2$, then

$$\begin{aligned}
f_K(0) &= \operatorname{var}(K(X_n)) + 2\sum_{h=1}^{\infty} \operatorname{cov}(K(X_n), K(X_{n+h})) \\
&= \sum_{i=0}^{\infty} a_i^4 \operatorname{var}(\varepsilon_1^2) + 4\sum_{0\le i<j<\infty} a_i^2 a_j^2 + 2\sum_{h=1}^{\infty}\sum_{i=0}^{\infty} a_i^2 a_{h+i}^2 \operatorname{var}(\varepsilon_1^2) \\
&+ 8\sum_{h=1}^{\infty}\sum_{0\le i<j<\infty} a_i a_{h+i} a_j a_{h+j}. \\
&= \sum_{i=0}^{\infty} a_i^4 \operatorname{var}(\varepsilon_1^2) + 4\sum_{1\le i<j<\infty} a_i^2 a_j^2 + 2\sum_{h=1}^{\infty}\sum_{i=0}^{\infty} a_i^2 a_{h+i}^2 \operatorname{Var}(\varepsilon_1^2) \\
&+ 8\sum_{h=1}^{\infty}\sum_{1\le i<j<\infty} a_i a_{h+i} a_j a_{h+j} + 4\sum_{i=1}^{\infty} a_i^2 + 8\sum_{h=1}^{\infty} a_h \sum_{i=1}^{\infty} a_i a_{i+h}.
\end{aligned}$$

We shall show that $f_K(0) > 0$. The condition $-1 < d < 0$ implies $a_i < 0$ for all $i > 0$. Therefore only the last term of the above decomposition of $f_K(0)$ is negative. To prove $f_K(0) > 0$, it suffices to show that

$$\sum_{i=1}^{\infty} a_i^2 + 2\sum_{h=1}^{\infty} a_h \sum_{i=1}^{\infty} a_i a_{i+h} \tag{2.2.14}$$

is positive. In fact,

$$\begin{aligned}
(2.2.14) &= \sum_{i=1}^{\infty} a_i^2 + 2\sum_{h=1}^{\infty} a_h \sum_{i=1}^{\infty} a_i a_{i+h} \\
&= \sum_{i=1}^{\infty} \frac{\Gamma^2(i+d)}{\Gamma^2(d)\Gamma^2(i+1)} + 2\sum_{h=1}^{\infty} \frac{\Gamma(h+d)}{\Gamma(d)\Gamma(h+1)} \sum_{i=1}^{\infty} \frac{\Gamma(i+d)\Gamma(i+h+d)}{\Gamma^2(d)\Gamma(i+1)\Gamma(i+h+1)} \\
&= \sum_{i=1}^{\infty} \frac{\Gamma^2(i+d)}{\Gamma^2(d)\Gamma^2(i+1)} + 2\sum_{h=1}^{\infty} \frac{\Gamma^2(h+d)}{\Gamma^2(d)\Gamma^2(h+1)}[F(d, h+d; h+1; 1) - 1] \qquad (2.2.15) \\
&= \sum_{h=1}^{\infty} \frac{\Gamma^2(h+d)}{\Gamma^2(d)\Gamma^2(h+1)}[2F(d, h+d; h+1; 1) - 1] \\
&= \sum_{h=1}^{\infty} \frac{\Gamma^2(h+d)}{\Gamma^2(d)\Gamma^2(h+1)} \frac{2\Gamma(h+1)\Gamma(1-2d)}{\Gamma(h+1-d)\Gamma(1-d)} - \sum_{h=1}^{\infty} \frac{\Gamma^2(h+d)}{\Gamma^2(d)\Gamma^2(h+1)} \qquad (2.2.16) \\
&= \frac{2\Gamma(1-2d)}{\Gamma^2(1-d)}[F(d, d; 1-d; 1) - 1] - [F(d, d; 1; 1) - 1] \\
&= \frac{2\Gamma(1-2d)}{\Gamma^2(1-d)}[\frac{\Gamma(1-d)\Gamma(1-3d)}{\Gamma^2(1-2d)} - 1] - \frac{\Gamma(1-2d)}{\Gamma^2(1-d)} + 1 \qquad (2.2.17) \\
&= \frac{3\Gamma(-3d)\Gamma(-d) - d\Gamma^2(-d)\Gamma(-2d) - 6\Gamma^2(-2d)}{-d\Gamma(-2d)\Gamma^2(-d)}. \qquad (2.2.18)
\end{aligned}$$

The notation $F(a, b; c; z)$ from (2.2.15) and thereafter is the hypergeometric series. (2.2.16) and (2.2.17) are obtained by applying the Gauss's theorem for hypergeometric series ([13]), see also page 2 of [1]. The denominate of the last equation (2.2.18) is positive since $-1 < d < 0$. Hence it suffices to prove that

$$3\Gamma(-3d)\Gamma(-d) - d\Gamma^2(-d)\Gamma(-2d) - 6\Gamma^2(-2d) > 0.$$

Define $f(x) = 3\Gamma(3x)\Gamma(x) + x\Gamma^2(x)\Gamma(2x) - 6\Gamma^2(2x)$, $0 < x < 1$. The function $f(x)$ is continuous for $x > 0$. Straight forward numerical calculation shows that $f(x) > 1/4 > 0$ for all $0 < x < 1$. Thus, (2.2.14) is positive. Hence, X_n^2 is a LM(0) process. □

Similar to the Gaussian case, Theorem 2.2.5 shows that antipersistence is a much more fragile property than long memory property. The antipersistence is immediately lost for the square transformation.

2.2 Simulation Study

To verify the main results in this section, in particular Theorem 2.2.3 and Theorem 2.2.5, we conduct simulation study for the memory of some common transformations of FARIMA(p, d, q) processes. These transformations include $K(x) = x^2$, x^3, x^4, $x^3 - 3x$, $x^4 - 6x^2$, $\sin x$, e^x and the non-continuous indicator function $I(x \leq c)$ for some constant c. First of all, we calculate the power rank of $K(\cdot)$ with respect to X_n and then find the theoretical memory parameter of each transformed process from Theorem 2.2.3. Although the power rank of $K(X)$ is identical to its Hermite rank if X has standard normal distribution ([20]), it may be different under different distributions. Nevertheless, one can easily find the power rank of a specific transformation under different distributions by the Definition 2.1.23. For example, provided that $\int \cos y dF(y) \neq 0$ or $\int e^y dF(y) < \infty$, the power rank of $K(x) = \sin x$ or $K(x) = e^x$ is 1 since

$$K_\infty(x) = \sin x \int \cos y dF(y) + \cos x \int \sin y dF(y)$$

or $K_\infty(x) = e^x \int e^y dF(y)$ satisfies $K'_\infty(0) \neq 0$. By similar analysis as above, the transformations $K(x) = x^2$, x^3, x^4, $x^3 - 3x$, $x^4 - 6x^2$, $\sin x$ and e^x have power rank 2, 1, 2, 3, 4, 1 and 1 respectively under some regular conditions on X_n (the conditions for different transformations may be different). For the indicator function $K(x) = I(x \leq c)$, the power rank depends on the value of the constant c by the following argument: Let $F(x)$ be the distribution function of X and assume that the density function $f(x)$ of X exists. Then

$$K_\infty(x) = \int I(x + y \le c) dF(y) = F(c - x)$$

and $K'_\infty(x) = -f(c-x)$. We then have $K'_\infty(0) = -f(c) \neq 0$ if $f(c)$ is quite far away from 0. Under this condition, the power rank of the indicator function is 1. If $f(c)$ is very close to 0 but $f'(c)$ exists and is quite far away from 0, we can say that the power rank of this indicator function $K(\cdot)$ is 2. Under certain smoothness condition, we can continue this procedure to find the power rank of $K(\cdot)$ for c in different ranges.

Secondly, to compare with the theoretical memory parameters, we perform simulation study for these transformations of FARIMA(p, d, q) processes X_n with memory parameters $d = 0.2$ and 0.4. The three processes in the simulation study are FARIMA$(0, d, 0)$ process (when $K(x) = x^2$, we also consider the cases that $d = -0.8$, -0.4, -0.2), FARIMA$(1, d, 0)$ process with the AR coefficient $\phi_1 = -0.3$, and FARIMA$(1, d, 1)$ process with the AR coefficient $\phi_1 = -0.4$ and the MA coefficient $\theta_1 = 0.7$.

Since our results require $\mathbb{E}K^2(X) < \infty$, $\mathbb{E}\varepsilon^4 < \infty$ and some transformations involve x^4, we take the Student t distribution with degree freedom 10 as the innovations of the FARIMA(p, d, q) processes for all transformations in our study except the last one $K(x) = e^x$. We choose the Gaussian FARIMA(p, d, q) processes in the transformation $K(X) = e^X$ since $\mathbb{E}e^{2X} < \infty$ is required. For each of these three processes and for each d, we conduct $N = 2,000$ simulations with $n = 2,000$ observations in each process by applying the algorithm in [12]. The memory parameters of each process and their transformations are estimated by the Fourier regression method proposed in [14]. As studied in [26], we choose the bandwidth $[n^{4/5}]$ for each estimation. The theoretical memory parameters and the estimated values are listed in Tables 2.2.1 and 2.2.2 respectively for each of these three processes. We also report the empirical standard error of the $N = 2,000$ estimates for each process in these tables. When d is negative, $d = -0.8, -0.4$ and -0.2, the theoretical memory parameters of all transformations of FARIMA$(0, d, 0)$ except the square of the FARIMA$(0, d, 0)$, are left in blank since we do not have theoretical results for these cases. When $d = 0.2$, we need $k \leq 2$ by the condition on d in Theorem 2.2, so the theoretical memory parameters of transformations with rank greater than 2 are left blank.

Table 2.2.1: Average estimated memory parameters of some transformations of 2,000 simulated stationary FARIMA$(0, d, 0)$ processes with 2,000 observations in each process and $t(10)$ innovations (except the transformation e^x, for which we use Gaussian innovations since $\mathbb{E}e^{2X} < \infty$ is required).

$K(X)$ and its power rank		Memory parameter of the original series X				
		$d = -0.8$	$d = -0.4$	$d = -0.2$	$d = 0.2$	$d = 0.4$
X	Theory	-0.8	-0.4	-0.2	0.2	0.4
(rank 1)	Simulation	-0.7674	-0.4008	-0.2005	0.2042	0.4075
	Std error	0.0496	0.0335	0.0333	0.0319	0.0332
X^2	Theory	0	0	0	0	0.3
(rank 2)	Simulation	0.0387	0.0250	0.0094	0.0405	0.2755
	Std error	0.0330	0.0323	0.0322	0.0364	0.0648
X^3	Theory				0.2	0.4
(rank 1)	Simulation	-0.1501	-0.0895	-0.0540	0.0960	0.2824
	Std error	0.0462	0.0383	0.0353	0.0372	0.0561
X^4	Theory				0	0.3
(rank 2)	Simulation	0.0330	0.0144	0.0038	0.0157	0.1855
	Std error	0.0329	0.0329	0.0292	0.0360	0.0790
$X^3 - 3X$	Theory					0.2
(rank 3)	Simulation	-0.0757	-0.0160	-0.0029	0.0087	0.2049
	Std error	0.0481	0.0345	0.0321	0.0347	0.0800
$X^4 - 6X^2$	Theory					0.1
(rank 4)	Simulation	0.0257	0.0051	0.0020	0.0008	0.1138
	Std error	0.0301	0.0294	0.0317	0.0322	0.0882
$\sin X$	Theory				0.2	0.4
(rank 1)	Simulation	-0.1651	-0.1863	-0.1365	0.1841	0.3167
	Std error	0.0349	0.0347	0.0334	0.0320	0.0439
e^X	Theory				0.2	0.4
(rank 1)	Simulation	-0.0486	-0.0919	-0.0796	0.1432	0.2952
	Std error	0.0339	0.0348	0.0321	0.0385	0.0603
$I(X \leq 0.1)$	Theory				0.2	0.4
(rank 1)	Simulation	-0.1408	-0.1342	-0.0961	0.1579	0.3124
	Std error	0.0316	0.0319	0.0326	0.0325	0.0371

Table 2.2.2: Average estimated memory parameters of some transformations of $2,000$ simulated FARIMA$(1, d, 0)$ and FARIMA$(1, d, 1)$ processes with $2,000$ observations in each process and $t(10)$ innovations (except the transformation e^x, for which we use Gaussian innovations since $\mathbb{E}e^{2X} < \infty$ is required). The FARIMA$(1, d, 0)$ processes have $\phi_1 = -0.3$ and the FARIMA$(1, d, 1)$ processes have $\phi_1 = -0.4,\ \theta_1 = 0.7$.

$K(X)$ and its power rank		Memory parameter of the original series X			
		FARIMA$(1, d, 0)$		FARIMA$(1, d, 1)$	
		$d = 0.2$	$d = 0.4$	$d = 0.2$	$d = 0.4$
X	Theory	0.2	0.4	0.2	0.4
(rank 1)	Simulation	0.1624	0.3663	0.2136	0.4188
	Std error	0.0325	0.0332	0.0329	0.0317
X^2	Theory	0	0.3	0	0.3
(rank 2)	Simulation	0.0212	0.2107	0.0646	0.3007
	Std error	0.0329	0.0635	0.0376	0.0615
X^3	Theory	0.2	0.4	0.2	0.4
(rank 1)	Simulation	0.0640	0.2173	0.1237	0.3083
	Std error	0.0350	0.0542	0.0374	0.0563
X^4	Theory	0	0.3	0	0.3
(rank 2)	Simulation	0.0084	0.1119	0.0356	0.2214
	Std error	0.0308	0.0696	0.0372	0.0778
$X^3 - 3X$	Theory		0.2		0.2
(rank 3)	Simulation	0.0087	0.2049	0.0406	0.2582
	Std error	0.0347	0.0800	0.0384	0.0695
$X^4 - 6X^2$	Theory		0.1		0.1
(rank 4)	Simulation	0.0004	0.0425	0.0144	0.1800
	Std error	0.0302	0.0633	0.0378	0.0905
$\sin X$	Theory	0.2	0.4	0.2	0.4
(rank 1)	Simulation	0.1417	0.2996	0.1868	0.2896
	Std error	0.0323	0.0433	0.0328	0.0412
e^X	Theory	0.2	0.4	0.2	0.4
(rank 1)	Simulation	0.1042	0.2672	0.1482	0.2856
	Std error	0.0364	0.0504	0.0392	0.0696
$I(X \le 0.1)$	Theory	0.2	0.4	0.2	0.4
(rank 1)	Simulation	0.1145	0.2728	0.1712	0.3228
	Std error	0.0327	0.0366	0.0331	0.0374

The simulation study with these polynomial or non-polynomial transformations clearly confirms the theoretical results in Theorem 2.2.3 and Theorem 2.2.5 for FARIMA$(0, d, 0)$ processes with $-1 < d < 1/2$ or in general FARIMA(p, d, q) processes with $0 < d < 1/2$.

One can also compare the result in Table 2.2.1 with the simulation study performed in [9]. They obtained the theoretical results for the memory parameters of the polynomial transformations of stationary Gaussian FARIMA$(0, d, 0)$ processes. But there were no theoretical results for the FARIMA(p, d, q) processes with p or q not zero or for the non-polynomial transformations, such as $K(x) = \sin x$, e^x, even in the Gaussian case. They performed simulation study for all the transformations in the Table 2.2.1 of Gaussian FARIMA$(0, d, 0)$ processes. As expected, due to the heavy tail innovation, the result in Table 2.2.1 is slightly worse than the one in[9]. The innovation $t(10)$ used here has heavier tail than Gaussian innovation.

2.3 TRANSFORMATIONS OF NON-STATIONARY PROCESSES

In this section, we explore the memory properties of polynomial transformations of one type non-stationary processes. In the case $1/2 < d < 3/2$, a non-stationary process X_n can be defined as the sum of a FARIMA$(0, d-1, 0)$ processes, i.e.,

$$X_n = X_0 + \sum_{j=1}^{n} Y_j, \tag{2.3.1}$$

where the distribution of the random variable X_0 does not depend on n,

$$Y_j = \sum_{i=0}^{\infty} a_i \varepsilon_{j-i}, \tag{2.3.2}$$

$a_i = \frac{\Gamma(i+d-1)}{\Gamma(i+1)\Gamma(d-1)}$ and ε_t are i.i.d. random variables with mean 0 and variance 1. $a_i \sim i^{d-2}/\Gamma(d-1)$ for large $i \in \mathbb{N}$. As in [44, 45], one can define X_n analogously in the case $d \geq 3/2$. In the literature, X_n defined in this way is called Type I process.

In the following theorem we obtain the memory property of X_n^2 for type I processes X_n. The memory property of $K(X_n)$ with a general transformation $K(\cdot)$ of type I processes X_n is complicate and we leave it as an open question.

Notice that $X_n - X_{n-1} = Y_n$ is a FARIMA$(0, d-1, 0)$ process. Thus $X_n \sim LM(d)$. In the following theorem, we show that this is also true asymptotically for X_n^2 in the case $1/2 < d < 1$.

Theorem 2.3.1. *Let X_n be a type I non-stationary process with $1/2 < d < 1$. Assume that $X_0 = 0$ and $\mathbb{E}\varepsilon^4 < \infty$. Then X_n^2 is asymptotically LM(d).*

Proof. By (2.3.1) and (2.3.2), X_n can be written in the form

$$X_n = \sum_{j=0}^{\infty} b_n(j)\varepsilon_{n-j},$$

where $b_n(j) = \sum_{i=0}^{j} a_i$, if $0 \le j \le n$, and $b_n(j) = \sum_{i=j-n+1}^{j} a_i$, if $j > n$. By convention, define $b_n(j) = 0$ for $j < 0$. Let $Z_n = X_n + X_{n-1}$. Then $X_n^2 - X_{n-1}^2 = Y_n Z_n$ and

$$Z_n = \sum_{j=0}^{\infty}[b_n(j) + b_{n-1}(j-1)]\varepsilon_{n-j}. \tag{2.3.3}$$

i). Denote $\gamma_y(h) = \text{cov}(Y_n, Y_{n+h})$ as the autocovariance function of the stationary process Y_n. We first show that, in the case $d < 5/4$,

$$\begin{aligned}&\text{cov}(Y_n Z_n, Y_{n+h}Z_{n+h})\\ &= \gamma_y(h)\text{cov}(Z_n, Z_{n+h}) + \text{cov}(Y_n, Z_{n+h})\text{cov}(Z_n, Y_{n+h}) + C(n,h)\end{aligned} \tag{2.3.4}$$

as $n \to \infty$ for some constant $C(n,h)$ with uniform bound $0 < C < \infty$. In fact, by (2.3.3),

$$Y_n Z_n = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} a_i[b_n(j) + b_{n-1}(j-1)]\varepsilon_{n-i}\varepsilon_{n-j}$$

and by the change of variables,

$$Y_{n+h}Z_{n+h} = \sum_{i=-h}^{\infty}\sum_{j=-h}^{\infty} a_{i+h}[b_{n+h}(j+h) + b_{n+h-1}(j+h-1)]\varepsilon_{n-i}\varepsilon_{n-j}.$$

Hence by the independence of the innovations ε_i, $i \in \mathbb{Z}$,

$$\begin{aligned}
&\mathrm{cov}(Y_n Z_n, Y_{n+h}Z_{n+h}) \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \{a_i a_{h+i}[b_n(j) + b_{n-1}(j-1)] \\
&\times [b_{n+h}(h+j) + b_{n+h-1}(h+j-1)] var(\varepsilon_{n-i}\varepsilon_{n-j})\} \\
&+ \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \{a_i a_{h+j}[b_n(j) + b_{n-1}(j-1)] \\
&\times [b_{n+h}(h+i) + b_{n+h-1}(h+i-1)] var(\varepsilon_{n-i}\varepsilon_{n-j})\}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\gamma_y(h)\mathrm{cov}(Z_n, Z_{n+h}) = \mathrm{cov}(Y_n, Y_{n+h})\mathrm{cov}(Z_n, Z_{n+h}) \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} a_i a_{h+i}[b_n(j) + b_{n-1}(j-1)][b_{n+h}(h+j) + b_{n+h-1}(h+j-1)]
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{cov}(Y_n, Z_{n+h})\mathrm{cov}(Z_n, Y_{n+h}) \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} a_i a_{h+j}[b_n(j) + b_{n-1}(j-1)][b_{n+h}(h+i) + b_{n+h-1}(h+i-1)].
\end{aligned}$$

Since $\text{var}(\varepsilon_1\varepsilon_2) = 1$, we have

$$
\begin{aligned}
&\text{cov}(Y_n Z_n, Y_{n+h} Z_{n+h}) - \gamma_y(h)\text{cov}(Z_n, Z_{n+h}) - \text{cov}(Y_n, Z_{n+h})\text{cov}(Z_n, Y_{n+h}) \\
&= 2\sum_{j=0}^{\infty} a_j a_{h+j}[b_n(j) + b_{n-1}(j-1)][b_{n+h}(h+j) + b_{n+h-1}(h+j-1)] \\
&\times [\text{var}(\varepsilon^2) - 1] \\
&= 2\sum_{j=0}^{n} a_j a_{j+h}\left[\sum_{i=0}^{j} a_i + \sum_{i=0}^{j-1} a_i\right]\left[\sum_{i=0}^{h+j} a_i + \sum_{i=0}^{h+j-1} a_i\right][\text{var}(\varepsilon^2) - 1] \\
&+ 2\sum_{j=n+1}^{\infty} a_j a_{j+h}\left[\sum_{i=j-n+1}^{j} a_i + \sum_{i=j-n+1}^{j-1} a_i\right]\left[\sum_{i=j-n+1}^{h+j} a_i + \sum_{i=j-n+1}^{h+j-1} a_i\right] \\
&\times [\text{var}(\varepsilon^2) - 1]. \qquad (2.3.5)
\end{aligned}
$$

In the above equations, since $a_i \sim \frac{i^{d-2}}{\Gamma(d-1)}$,

$$
\left| a_j a_{j+h}\left[\sum_{i=0}^{j} a_i + \sum_{i=0}^{j-1} a_i\right]\left[\sum_{i=0}^{h+j} a_i + \sum_{i=0}^{h+j-1} a_i\right]\right| < Cj^{2d-4}
$$

for some $0 < C < \infty$. Similarly,

$$
\left| a_j a_{j+h}\left[\sum_{i=j-n+1}^{j} a_i + \sum_{i=j-n+1}^{j-1} a_i\right]\left[\sum_{i=j-n+1}^{h+j} a_i + \sum_{i=j-n+1}^{h+j-1} a_i\right]\right| < Cj^{4d-6}
$$

for some $0 < C < \infty$. So that (2.3.5) converges as $n \to \infty$ if $d < 5/4$. Consequently, in this case $d < 5/4$, (2.3.4) holds as $n \to \infty$.

ii). Now we consider the second term of (2.3.4), $\mathrm{cov}(Y_n, Z_{n+h})\mathrm{cov}(Z_n, Y_{n+h})$.

$$
\begin{aligned}
|\mathrm{cov}(Y_n, Z_{n+h})| &= \left|\sum_{j=0}^{\infty} a_j[b_{n+h}(h+j) + b_{n+h-1}(h+j-1)]\right| \\
&= \left|\sum_{j=0}^{n} a_j[\sum_{i=0}^{h+j} a_i + \sum_{i=0}^{h+j-1} a_i] + \sum_{j=n+1}^{\infty} a_j[\sum_{i=j-n+1}^{h+j} a_i + \sum_{i=j-n+1}^{h+j-1} a_i]\right| \\
&< C\left(\sum_{j=1}^{n} j^{d-2} + \sum_{j=n+1}^{\infty} j^{2d-3}\right)
\end{aligned}
$$

for some constant $C > 0$. Therefore in the case $d < 1$, the series for $\mathrm{cov}(Y_n, Z_{n+h})$ converges as $n \to \infty$. By the same argument, the series for $\mathrm{cov}(Z_n, Y_{n+h})$ converges and hence the series for the product $\mathrm{cov}(Y_n, Z_{n+h})\mathrm{cov}(Z_n, Y_{n+h})$ converges as $n \to \infty$ if $d < 1$. So in the case $d < 1$,

$$
\mathrm{cov}(Y_n Z_n, Y_{n+h} Z_{n+h}) = \gamma_y(h)\mathrm{cov}(Z_n, Z_{n+h}) + C(n, h) \qquad (2.3.6)
$$

as $n \to \infty$ for some constant $C(n, h)$ with uniform bound $0 < C < \infty$. As a particular case of (2.3.6),

$$
\mathrm{var}(Y_n Z_n) = \mathrm{var}(Y_n)\mathrm{var}(Z_n) + C(n) \quad \text{as } n \to \infty \qquad (2.3.7)
$$

for some constant $C(n)$ with uniform bound $0 < C < \infty$.

iii). Next we prove that the non-stationary process Z_n satisfies:

$$
\mathrm{corr}(Z_n, Z_{n+h}) \to 1 \quad \text{as } n \to \infty \qquad (2.3.8)
$$

and

$$
\mathrm{var}(Z_n) \to \infty \quad \text{as } n \to \infty. \qquad (2.3.9)
$$

We first show that (2.3.8) holds under the condition (2.3.9). In fact,

$$\begin{aligned}\operatorname{cov}(Z_n, Z_{n+h}) &= \operatorname{cov}(X_n + X_{n-1}, X_{n+h} + X_{n+h-1}) \\ &= \operatorname{cov}(2\sum_{j=1}^{n-1} Y_j + Y_n, 2\sum_{i=1}^{n+h-1} Y_i + Y_{n+h}) \\ &= 4\sum_{i=1}^{n+h-1}\sum_{j=1}^{n-1}\gamma_y(i-j) + 2\sum_{j=1}^{n-1}\gamma_y(n+h-j) \\ &+ 2\sum_{i=1}^{n+h-1}\gamma_y(n-i) + \gamma_y(h). \end{aligned} \tag{2.3.10}$$

Let $h = 0$ in (2.3.10), we have

$$\operatorname{var}(Z_n) = 4\sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\gamma_y(i-j) + 4\sum_{j=1}^{n-1}\gamma_y(n-j) + \gamma_y(0) \tag{2.3.11}$$

By replacing the n in (2.3.11),

$$\begin{aligned}\operatorname{var}(Z_{n+h}) &= 4\sum_{i=1}^{n+h-1}\sum_{j=1}^{n+h-1}\gamma_y(i-j) + 4\sum_{j=1}^{n+h-1}\gamma_y(n+h-j) + \gamma_y(0) \\ &= 4\sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\gamma_y(i-j) + 4\sum_{j=1}^{n-1}\gamma_y(n-j) + \gamma_y(0) \\ &+ 8\sum_{i=1}^{n-1}\sum_{j=n}^{n+h-1}\gamma_y(i-j) + 4\sum_{i=n}^{n+h-1}\sum_{j=n}^{n+h-1}\gamma_y(i-j) + 4\sum_{j=0}^{h-1}\gamma_y(n+j) \\ &= \operatorname{var}(Z_n) + C(n,h)\end{aligned}$$

for some constant $C(n,h)$ with uniform bound $0 < C < \infty$ since Y_n is a short memory process. Therefore,

$$\operatorname{var}(Z_n)\operatorname{var}(Z_{n+h}) = \operatorname{var}^2(Z_n)\left[1 + \frac{C(n,h)}{\operatorname{var}(Z_n)}\right]. \tag{2.3.12}$$

On the other hand, by (2.3.10) and (2.3.11),

$$\begin{aligned}
&\operatorname{cov}(Z_n, Z_{n+h}) \\
&= 4\sum_{i=1}^{n-1}\sum_{j=1}^{n-1}\gamma_y(i-j) + 4\sum_{j=1}^{n-1}\gamma_y(n-j) + \gamma_y(0) \\
&+ 4\sum_{i=n}^{n+h-1}\sum_{j=1}^{n-1}\gamma_y(i-j) - 2\sum_{i=n-h}^{n-1}\gamma_y(n-i) \\
&+ 2\sum_{i=0}^{h-1}\gamma_y(n+i) + 2\sum_{i=n}^{n+h-1}\gamma_y(n-i) + \gamma_y(h) - \gamma_y(0) \\
&= \operatorname{var}(Z_n) + 4\sum_{i=n}^{n+h-1}\sum_{j=1}^{n-1}\gamma_y(i-j) + 2\sum_{i=0}^{h-1}\gamma_y(n+i) - \gamma_y(h) + \gamma_y(0) \\
&= \operatorname{var}(Z_n) + C(n,h), \qquad (2.3.13)
\end{aligned}$$

where $C(n,h)$ is bounded uniformly by some constant $C > 0$.

Provided that (2.3.9) holds, i.e., $\operatorname{var}(Z_n) \to \infty$, by (2.3.12) and (2.3.13),

$$\begin{aligned}
\operatorname{corr}(Z_n, Z_{n+h}) &= \frac{\operatorname{cov}(Z_n, Z_{n+h})}{\sqrt{\operatorname{var}(Z_n)\operatorname{var}(Z_{n+h})}} \\
&= \frac{\operatorname{var}(Z_n) + C(n,h)}{\operatorname{var}(Z_n)\sqrt{[1 + \frac{C(n,h)}{\operatorname{var}(Z_n)}]}} \\
&\to 1 \quad \text{as } n \to \infty,
\end{aligned}$$

which proves that (2.3.8) is true.

Now we show that the second property (2.3.9) of the non-stationary process Z_n also holds. Since $Y_n \sim \text{FARIMA}(0, d-1, 0)$ with $1/2 < d < 1$, we have $\gamma_y(h) < 0$ for all $h > 0$ and the spectral density at frequency zero

$$f(0) = \gamma_y(0) + 2\sum_{h=1}^{\infty}\gamma_y(h) = 0. \qquad (2.3.14)$$

For computational convenience, we assume $Y_n \sim \text{FARIMA}(0, d, 0)$ with $-1/2 < d < 0$ in the following process. Brockwell and Davis [4] gave the autocovariance function and autocorrelation function for FARIMA$(0, d, 0)$ processes with $-1/2 < d < 1/2$,

$$\gamma_y(0) = \frac{\sigma^2 \Gamma(1-2d)}{\Gamma^2(1-d)}$$

and

$$\rho_y(h) = \frac{\Gamma(h+d)\Gamma(1-d)}{\Gamma(h-d+1)\Gamma(d)}, \quad h \in \mathbb{N},$$

where $\sigma^2 = 1$ is the variance of innovation ε. Thus,

$$\gamma_y(h) = \frac{\Gamma(1-2d)\sigma^2}{\Gamma(1-d)\Gamma(d)} \frac{\Gamma(h+d)}{\Gamma(h+1-d)}. \tag{2.3.15}$$

To prove (2.3.9), it suffices to prove that the first quantity in (2.3.11) goes to infinity as $n \to \infty$ since Y_n is a short memory process in the covariance sense and therefore the second and third terms in (2.3.11) are bounded. By collecting terms, we get

$$\sum_{i=1}^{n-1}\sum_{j=1}^{n-1} \gamma_y(i-j) = (n-1)\gamma_y(0) + 2\sum_{h=2}^{n-1}(n-h)\gamma_y(h-1)$$

$$= (n-1)[\gamma_y(0) + 2\sum_{h=1}^{n-2} \frac{n-1-h}{n-1}\gamma_y(h)]$$

$$= (n-1)[\gamma_y(0) + 2\sum_{h=1}^{n-2} \gamma_y(h) - 2\sum_{h=1}^{n-2} \frac{h}{n-1}\gamma_y(h)]$$

$$= (n-1)[-2\sum_{h=n-1}^{\infty} \gamma_y(h) - 2\sum_{h=1}^{n-2} \frac{h}{n-1}\gamma_y(h)] \tag{2.3.16}$$

$$= -2(n-1)\sum_{h=n-1}^{\infty} \gamma_y(h) - 2\sum_{h=1}^{n-2} h\gamma_y(h). \tag{2.3.17}$$

The equality (2.3.16) is from (2.3.14). Both of the two terms in equation (2.3.17) are positive since $\gamma_y(h) < 0$ if $h > 0$. So we prove that (2.3.17) goes to infinity by showing that the first term of (2.3.17) goes to infinity as $n \to \infty$.

$$\begin{aligned}
& -2(n-1)\sum_{h=n-1}^{\infty}\gamma_y(h) \\
&= -2\sigma^2\frac{\Gamma(1-2d)}{\Gamma(1-d)\Gamma(d)}(n-1)\sum_{h=n-1}^{\infty}\frac{\Gamma(h+d)}{\Gamma(h+1-d)} && (2.3.18)\\
&\simeq (n-1)\alpha\sum_{h=n-1}^{\infty}\frac{e^{-h-d}(h+d)^{h+d-1/2}}{e^{-h+d-1}(h+1-d)^{h-d+1/2}} && (2.3.19)\\
&= \alpha e^{1-2d}(n-1)\sum_{h=n-1}^{\infty}\left(1+\frac{1-2d}{h+d}\right)^{-(h+d-1/2)}(h+1-d)^{2d-1}\\
&\simeq \alpha e^{1-2d}(n-1)e^{2d-1}\sum_{h=n-1}^{\infty}(h+1-d)^{2d-1} && (2.3.20)\\
&\to \infty \quad \text{as } n\to\infty \text{ since } 2d+1>0.
\end{aligned}$$

In the above equations, $\alpha = -2\sigma^2\frac{\Gamma(1-2d)}{\Gamma(1-d)\Gamma(d)}$; (2.3.18) is obtained by using (2.3.15); (2.3.19) is from the Stirling's approximation; and (2.3.20) is obtained from the fact that $\lim_{n\to\infty}(1+\frac{1}{n})^n = e$. Consequently, (2.3.9) is true.

From the above analysis, in particular (2.3.6), (2.3.7), (2.3.8) and (2.3.9),

$$\begin{aligned}
& \operatorname{corr}(Y_nZ_n, Y_{n+h}Z_{n+h})\\
=\ & \frac{\operatorname{cov}(Y_nZ_n, Y_{n+h}Z_{n+h})}{\sqrt{\operatorname{var}(Y_nZ_n)\operatorname{var}(Y_{n+h}Z_{n+h})}}\\
=\ & \frac{\operatorname{cov}(Y_n, Y_{n+h})\operatorname{cov}(Z_n, Z_{n+h}) + C(n,h)}{\sqrt{\operatorname{var}(Y_n)\operatorname{var}(Z_n) + C(n)}\sqrt{\operatorname{var}(Y_{n+h})\operatorname{var}(Z_{n+h}) + C(n,h)}}\\
\simeq\ & \frac{\gamma_y(h)\operatorname{cov}(Z_n, Z_{n+h})}{\sqrt{\gamma_y(0)\operatorname{var}(Z_n)}\sqrt{\gamma_y(0)\operatorname{var}(Z_{n+h})}} \quad \text{since } \operatorname{var}(Z_n), \operatorname{var}(Z_{n+h}) \to \infty\\
=\ & \frac{\gamma_y(h)}{\gamma_y(0)}\frac{\operatorname{cov}(Z_n, Z_{n+h})}{\sqrt{\operatorname{var}(Z_n)\operatorname{var}(Z_{n+h})}}\\
\to\ & \operatorname{corr}(Y_n, Y_{n+h}) \quad \text{as } n\to\infty.
\end{aligned}$$

As $Y_n \sim \text{FARIMA}(0, d-1, 0)$, $X_n^2 \sim LM(d)$ when $1/2 < d < 1$.

□

Theorem 2.3.1 shows that taking the square of a non-stationary long memory process does not change the size of the long memory parameter, which is contrast to the result of stationary FARIMA(p, d, q) processes.

The simulation study in Table 2.3.1 is to confirm the result in Theorem 2.3.1. We simulate FARIMA$(0, d-1, 0)$ processes for each of the d values, $d = 0.55$, 0.65, 0.75, 0.85, 0.95. The i.i.d. innovations have Student t distribution with degree of freedom 5. By the Definition 2.3.1, the partial sum gives a type I process. The method to produce FARIMA$(0, d-1, 0)$ processes and the method to estimate the memory parameters are same as the ones in Section 2.2.1. It is clear from Table 2.3.1 that X_t^2 is asymptotically LM(d) process. Also the rest part of Table 2.3.1 seems to confirm one conjecture: under suitable moment conditions, any polynomial transformations of non-stationary FARIMA$(0, d, 0)$ processes are LM(d) processes, $1/2 < d < 1$. The memory property of polynomial transformations of FARIMA$(0, d, 0)$ processes is not related to the power ranks of these transformations.

Table 2.3.1: Average estimated parameters of some polynomial transformations of 2,000 simulated FARIMA$(0, d, 0)$ processes with 2,000 observations in each process. For the transformation $K(x) = x^2$, the innovation of the original process X_n has Student t distribution with degree of freedom 5. For other transformations, the innovation of the original process X_n has Student t distribution with degree of freedom 10.

$K(X)$		Memory parameter d of the original series X				
		0.55	0.65	0.75	0.85	0.95
X^2	Theory	0.55	0.65	0.75	0.85	0.95
	Simulation	0.4826	0.6170	0.7433	0.8629	0.9647
	Std error	0.0673	0.0592	0.0530	0.0531	0.0456
X^3	Simulation	0.4845	0.6102	0.7308	0.8479	0.9598
	Std error	0.0637	0.0639	0.0620	0.0597	0.0554
X^4	Simulation	0.4265	0.5702	0.7066	0.8087	0.9441
	Std error	0.0913	0.0873	0.0821	0.0757	0.0678
$X^3 - 3X$	Simulation	0.4667	0.6029	0.7298	0.8502	0.9564
	Std error	0.0781	0.0671	0.0637	0.0628	0.0554
$X^4 - 6X^2$	Simulation	0.4037	0.5585	0.7024	0.8280	0.9452
	Std error	0.1053	0.0969	0.0854	0.0785	0.0675

Remark 2.3.21. The decomposition in Section 2.2.1 is not applicable for the non-stationary process with $1/2 < d < 1$. In fact, the result in Theorem 2.3.1 is not related to the ranks of the transformations.

2.4 APPLICATION IN OPTION PROCESSES

The transformation $K(x) = (x - C)^+$ itself has independent interest. It is $x - C$ if $x \geq C > 0$. Otherwise it is 0. Notice that this $K(x)$ is not differentiable at C. For the reason to be clear later, let $X \geq 0$ be a random variable with mean μ. Then $Y = X - \mu$ has mean 0 and

$$K(X) = (X - C)^+ = (Y - (C - \mu))^+ := H(Y).$$

Let $G(y)$ be the distribution function of Y. Assume that the density function of Y exists and let it be $g(y)$. Then

$$\begin{aligned} H_\infty(y) &= \int (y + z - (C-\mu))^+ dG(z) \\ &= \int_{C-\mu-y}^{\infty} (y + z - (C-\mu)) dG(z) \\ &= (y - C + \mu)[1 - G(C-\mu-y)] + \int_{C-\mu-y}^{\infty} z dG(z) \end{aligned}$$

and

$$\begin{aligned} &H'_\infty(y) \\ &= 1 - G(C-\mu-y) + (y - C + \mu) g(C-\mu-y) + (C-\mu-y) g(C-\mu-y) \\ &= 1 - G(C-\mu-y). \end{aligned}$$

We then have $H'_\infty(0) = 1 - G(C-\mu) \neq 0$ if $C-\mu$ is small enough. Therefore in this case the power rank of this $H(\cdot)$ is 1. If a larger $C - \mu > 0$ is in the range such that $H'_\infty(0) = 1 - G(C-\mu) \approx 0$ and at the same time $g(C-\mu)$ is quite far away from 0, we can say that the power rank of this $H(\cdot)$ is 2 since $H''_\infty(y) = g(C-\mu-y)$ and therefore $H''_\infty(0) = g(C-\mu) > 0$. If $G(y)$ is smooth enough (measured by the order of differentiability), we can continue this procedure to find the power rank of $H(\cdot)$ for $C-\mu$ in different ranges. In particular, if $G^{(r)}(y)$ exists for any $r \in \mathbb{N}$, and for $C-\mu > 0$ large enough, $G^{(r)}(C-\mu) \approx 0$, then we say the power rank is ∞ since $H^{(k)}_\infty(y) = (-1)^k g^{(k-2)}(C-\mu-y)$ and therefore $H^{(k)}_\infty(0) = (-1)^k g^{(k-2)}(C-\mu) \approx 0$.

We conduct simulation study for transformations $(X_n - C)^+$ with $C = 0.3$, 1.5, 5, 9, 44.8 and 45.5 where $Y_n = X_n - \mu$ are FARIMA$(0, d, 0)$ processes with $d = 0.2$ and 0.4 as in Section 2.2.1. The innovations of X_n are the absolute values of Student t random variables with degree of freedom 10. The way to estimate the memory parameters of the transformations including the selection of the bandwidth $[n^{4/5}]$ is same as the one in Section

2.2.1. For each $d = 0.2$ and 0.4, we conduct $N = 2,000$ simulations with $n = 2^{20}$ observations in each process. Notice that the mean μ of X_n changes for different memory parameters d. Therefore the power rank of $(X_n - C)^+ = (Y_n - (C - \mu))^+$ also varies with μ for each fixed $C > 0$. The result is listed in Table 2.4.1. Again, If $d = 0.2$, the theoretical memory parameters of transformations of the FARIMA$(0, d, 0)$ processes with rank greater than 2 are left in blank. In the table, there are no estimates if $C - \mu > 5$ since the length of each simulated process X_n is finite and therefore the transformed values are all zeros if $C - \mu$ is too large. We use NA to denote them. Theoretically, the memory of a degenerate time series is zero. The simulation study confirms the results in Theorem 2.2.3 and the above analysis.

The study of the memory parameter of $K(x) = (x - C)^+$ has direct application to call option time series in finance. Suppose X_n is the price process of the underlying asset and C is the strike price, then $K(X_n)$ is the value of the call option. Our result shows that the memory parameter of $(X_n - C)^+$ is same as the memory parameter of the underlying asset X_n if $C - \mu$ is small. The power rank of $(X_n - C)^+ = (Y_n - (C - \mu))^+$ is 2 approximately if $C - \mu$ is in some moderate range. In this case, according to Theorem 2.2.3 (which is confirmed by the simulation study), the memory parameter of $K(X_n)$ is $2d - 1/2$ if the memory parameter d of the original mean adjusted asset price process $X_n - \mu \sim$ FARIMA$(0, d, 0)$ satisfies $1/4 < d < 1/2$.

Similar analysis can be conducted for the truncation function $K(x) = (C - x)^+$ and the put option time series $(C - X_n)^+$ at different $C > 0$.

Table 2.4.1: Average estimated memory parameter of some transformations $(X_n - C)^+$ of $2,000$ simulated FARIMA$(0, d, 0)$ processes $X_n - \mu$ with $|t(10)|$ innovations and 2^{20} observations in each process X_n. $C = 0.3$, 1.5, 5, 9, 44.8 and 45.5 and $Y_n = X_n - \mu$.

	Memory parameter of the original series X	
	$d = 0.2$	$d = 0.4$
$(Y - (0.3 - \mu))^+$	$(Y + 6.89)^+$	$(Y + 44.74)^+$
Rank	1	1
Theory	0.2	0.4
Simulation	0.1999	0.3998
Std error	0.0024	0.0025
$(Y - (1.5 - \mu))^+$	$(Y + 5.69)^+$	$(Y + 43.24)^+$
Rank	1	1
Theory	0.2	0.4
Simulation	0.1999	0.3998
Std error	0.0025	0.0024
$(Y - (5 - \mu))^+$	$(Y + 2.19)^+$	$(Y + 39.74)^+$
Rank	1	1
Theory	0.2	0.4
Simulation	0.1999	0.3999
Std error	0.0025	0.0024
$(Y - (9 - \mu))^+$	$(Y - 1.81)^+$	$(Y + 35.74)^+$
Rank	2	1
Theory	0	0.4
Simulation	0.0246	0.3998
Std error	0.0027	0.0025
$(Y - (44.8 - \mu))^+$	$(Y - 37.61)^+$	$(Y - 0.06)^+$
Rank	∞	1
Theory		0.4
Simulation	NA	0.3332
Std error	NA	0.0049
$(Y - (45.5 - \mu))^+$	$(Y - 38.31)^+$	$(Y - 0.76)^+$
Rank	∞	2
Theory		0.3
Simulation	NA	0.2541
Std error	NA	0.0074

2.5 CONCLUSION

Under the condition proposed by [51], we have obtained memory properties of transformation of stationary processes both in time domain and frequency domain. The results in time domain is consistent with the limit theorems in [20] and [51] via the order of normalization. The results are applicable not only to FARIMA$(0, d, 0)$ processes as studied in [9] for Gaussian case, but also to general FARIMA(p, d, q) processes for some special transformation. The results hold not only for smooth transformations, they also hold for functions which are not differentiable. In particular, we study the memory properties of option time series $(X_n - C)^+$ in finance for different strike price $C > 0$. We have also derived the memories of the square transformation of an non-stationary process.

3 SYMMETRIC GINI COVARIANCE AND CORRELATION

3.1 PRELIMINARIES

When we develop the memory properties for a time series process, we always apply the autocorrelation to measure those memories. Correlation analysis is another popular topic in statistics, and the standard Gini correlation is one of the important type of correlations which are used to measure the dependence of random variables or processes. Standard Gini correlation is developed based on the Gini's mean difference.

3.1 Gini Mean Difference

Gini's mean difference (GMD) was introduced by Gini [15] as an alternative measure of variability to the standard deviation. The original definition of the GMD is the expected absolute difference between two realizations of i.i.d. random variables. That is,

Definition 3.1.1 (Gini Mean Difference)**.** *For a random variable X from a univariate distribution F, the GMD of X (or F) is*

$$\sigma_g = \sigma_g(X) = \sigma_g(F) = \mathbb{E}|X_1 - X_2|,$$

where X_1 and X_2 are independent random variables from F.

The variance of X (or F) is $\sigma_v^2(F) = \text{var}(X) = \frac{1}{2}\mathbb{E}(X_1 - X_2)^2$. Apparently, GMD only needs the existence of the first moment of the distribution F, so it is more robust than the variance and is often used for heavy-tailed asymmetric distributions. GMD has more than 14 different alternative representations, and there are four types of formulas for

GMD. Among all the formulations of GMD (L_1 metric, integrals of cumulative distribution functions, covariance and Lorenz curves), we are interested in the covariance formulations. One representation of GMD based on the covariance is

$$\sigma_g(F) = 4\text{cov}(X, F(X)).$$

While the variance is the covariance of X with itself, the GMD is (four times) the covariance of X with $F(X)$. In this spirit, two Gini-type alternatives to the usual covariance for measuring the dependence of a random variable X and another random variable Y with distribution function G are

$$\text{covg}(X, Y) = 4\text{cov}(X, G(Y)), \quad \text{covg}(Y, X) = 4\text{cov}(Y, F(X)). \tag{3.1.2}$$

Based on (3.1.2), the Gini correlations defined by (1.2.1) are not equal in general. An even worse part is that $\gamma(X, Y)$ and $\gamma(Y, X)$ may have different signs in some cases ([54]), which brings substantial difficulty in interpretation. The asymmetry stems from the usage of $F(X)$ or $G(Y)$, which can be thought as a standardized marginal rank. A symmetry one calls for a joint rank of X and Y. The other covariance type formulation for GMD is

$$\sigma_g(F) = 2\text{cov}(X, 2F(X) - 1),$$

allowing an insightful interpretation: $\sigma_g(X)$ is twice the covariance of X and the centered rank function $r(X) = 2F(X) - 1$. $r(X)$ is centered because $\mathbb{E}r(X) = 0$ if F is continuous. So

$$\sigma_g(F) = 2\text{cov}(X, r(X)) = 2\mathbb{E}(Xr(X)). \tag{3.1.3}$$

A nice generalization of the centered rank in high dimensions provides a joint rank, and along with the representation of GMD in (3.1.3) yields a natural extension of GMD for a multivariate distribution F.

3.1 Spatial Rank Function

Let $\boldsymbol{X}$ be a d-variate random vector from continuous distribution F with a finite first moment and the expected Euclidean distance from $\boldsymbol{x}$ to $\boldsymbol{X}$ be $D(\boldsymbol{x}, F) = \mathbb{E}_F\|\mathbf{x} - \mathbf{X}\|$. Then the gradient of D is denoted as the centered *spatial rank function* ([32]), that is,

$$\mathbf{r}(\mathbf{x}) = \nabla_{\boldsymbol{x}} D(\boldsymbol{x}, F) = \mathbb{E}\frac{\mathbf{x} - \mathbf{X}}{\|\mathbf{x} - \mathbf{X}\|} = \mathbb{E}\{\mathbf{s}(\mathbf{x} - \mathbf{X})\},$$

where $\mathbf{s}(\mathbf{x} - \mathbf{X}) = \mathbf{x}/\|\mathbf{x}\|$ ($\mathbf{s}(\mathbf{0}) = \mathbf{0}$) is the *spatial sign function* in $\mathbb{R}^d$. The spatial rank function is the expected direction from $\mathbf{X}$ to $\mathbf{x}$. We call it centered because a random rank is centered at $\mathbf{0}$, that is, $\mathbb{E}\mathbf{r}(\mathbf{X}) = \mathbf{0}$. The solution of $\mathbf{x}$ in $\mathbf{r}(\mathbf{x}) = \mathbf{0}$ is called the spatial median of F, which minimizes D. In the univariate case, the derivative of $D(x, F) = \mathbb{E}|x - X|$ with respect to x leads to the univariate centered rank function $r(x) = \mathbb{E}\text{sign}(x - X) = 2F(x) - 1 \in [-1, 1]$ if F is continuous. Clearly, the median of F has a center rank 0.

3.2 SYMMETRIC GINI CORRELATION

Given a random vector $\mathbf{Z}$ in $\mathbb{R}^d$ with distribution H, the *spatial rank* of $\mathbf{z}$ with respect to the distribution H is defined as

$$\mathbf{r}(\mathbf{z}, H) := \mathbb{E}\mathbf{s}(\mathbf{z}\text{-}\mathbf{Z}) = \mathbb{E}\frac{\mathbf{z}\text{-}\ \mathbf{Z}}{||\mathbf{z}\text{-}\ \mathbf{Z}||}.$$

For a more comprehensive account of the spatial rank, see [33].

In particular, for $d = 2$ with $\mathbf{Z} = (X, Y)^T$, the bivariate spatial rank function of $\mathbf{z} = (x, y)^T$ is

$$\mathbf{r}(\mathbf{z}, H) = \mathbb{E}\frac{(x - X, y - Y)^T}{\|\mathbf{z} - \mathbf{Z}\|} := (R_1(\mathbf{z}), R_2(\mathbf{z}))^T,$$

where $R_1(\mathbf{z}) = \mathbb{E}(x - X)/\|\mathbf{z} - \mathbf{Z}\|$ and $R_2(\mathbf{z}) = \mathbb{E}(y - Y)/\|\mathbf{z} - \mathbf{Z}\|$ are two components of the joint rank function $\mathbf{r}(\mathbf{z}, H)$.

3.2 Symmetric Gini Covariance

Our new symmetric covariance and correlation are defined based on the bivariate spatial rank function. Replacing the univariate centered rank in (3.1.3) with $R_2(\mathbf{z})$, we define the *symmetric Gini covariance* as

$$\text{cov}_g(X, Y) := 2\mathbb{E}XR_2(\mathbf{Z}). \tag{3.2.1}$$

Note that $\text{cov}_g(X, Y) = 2\text{cov}(X, R_2(\mathbf{Z}))$ if H is continuous. Dually, $\text{cov}_g(Y, X) = 2\mathbb{E}YR_1(\mathbf{Z})$ can also be taken as the definition of the symmetric Gini covariance between X and Y. Indeed,

$$\begin{aligned}\text{cov}_g(X, Y) &= 2\mathbb{E}XR_2(\mathbf{Z}) = 2\mathbb{E}(X_1\mathbb{E}[\frac{Y_1 - Y_2}{||\mathbf{Z}_1 - \mathbf{Z}_2||}|\mathbf{Z}_1]) = 2\mathbb{E}X_1\frac{Y_1 - Y_2}{||\mathbf{Z}_1 - \mathbf{Z}_2||}\\ &= -2\mathbb{E}X_2\frac{Y_1 - Y_2}{||\mathbf{Z}_1 - \mathbf{Z}_2||} = \mathbb{E}[\frac{(X_1 - X_2)(Y_1 - Y_2)}{||\mathbf{Z}_1 - \mathbf{Z}_2||}] = \text{cov}_g(Y, X),\end{aligned} \tag{3.2.2}$$

where $\mathbf{Z}_1 = (X_1, Y_1)^T$ and $\mathbf{Z}_2 = (X_2, Y_2)^T$ are independent copies of $\mathbf{Z} = (X, Y)^T$ from H. In addition, we define

$$\text{cov}_g(X, X) := 2\mathbb{E}XR_1(\mathbf{Z}) = \mathbb{E}\frac{(X_1 - X_2)^2}{\|\mathbf{Z}_1 - \mathbf{Z}_2\|}; \tag{3.2.3}$$

$$\text{cov}_g(Y, Y) := 2\mathbb{E}YR_2(\mathbf{Z}) = \mathbb{E}\frac{(Y_1 - Y_2)^2}{\|\mathbf{Z}_1 - \mathbf{Z}_2\|}. \tag{3.2.4}$$

We see that not only the Gini covariance between X and Y but also Gini variances of X and of Y are defined jointly through the spatial rank. [7] considered the Gini covariance matrix $\mathbf{\Sigma}_g = 2\mathbb{E}\mathbf{Z}\mathbf{r}^T(\mathbf{Z})$. The covariances defined in (3.2.1), (3.2.3) and (3.2.4) are elements of $\mathbf{\Sigma}_g$ for two dimensional random vectors. Rather than the assumption of a finite second moment in the usual covariance and variance, the Gini counterparts assume only the first moment, hence being more suitable for heavy-tailed distributions. A related covariance matrix is the spatial sign covariance matrix (SSCM), which requires a location parameter to be known but no assumption on moments ([46]).

Particularly if Z is a one dimensional random variable, we have $\text{cov}_g(Z, Z) = \mathbb{E}|Z_1 - Z_2|$, which reduces to GMD. In this sense, we may view the symmetric Gini covariance as a direct generalization of GMD to two variables.

3.2 Symmetric Gini Correlation

Using the symmetric Gini covariance defined by (3.2.1), we propose a symmetric *Gini correlation coefficient* as follows.

Definition 3.2.5. *$\boldsymbol{Z} = (X, Y)^T$ is a bivariate random vector from the distribution H with finite first moment and non-degenerate marginal distributions, then the symmetric Gini correlation between X and Y is*

$$\rho_g(X,Y) := \frac{cov_g(X,Y)}{\sqrt{cov_g(X,X)}\sqrt{cov_g(Y,Y)}} = \frac{\mathbb{E}XR_2(\boldsymbol{Z})}{\sqrt{\mathbb{E}XR_1(\boldsymbol{Z})}\sqrt{\mathbb{E}YR_2(\boldsymbol{Z})}}. \tag{3.2.6}$$

Theorem 3.2.1. *For a bivariate random vector $(X, Y)^T$ from H with finite first moment, ρ_g has the following properties:*

12. $\rho_g(X, Y) = \rho_g(Y, X)$.

22. $-1 \leq \rho_g(X, Y) \leq 1$.

32. If X, Y are independent, then $\rho_g(X, Y) = 0$.

42. *If* $Y = aX + b$ *and* $a \neq 0$*, then* $\rho_g = sgn(a)$*.*

52. $\rho_g(aX + b, aY + d) = \rho_g(X, Y)$ *for any constants* b*,* d *and* $a \neq 0$*. Measure* ρ_g *is sensitive to a heterogeneous change, i.e.,* $\rho_g(aX, cY) \neq \rho_g(X, Y)$ *for* $a \neq c$*. In particular,* $\rho_g(X, Y) = -\rho_g(aX, -aY) = -\rho_g(-aX, aY)$*.*

Proof. The first property is obvious. Hölder's inequality implies $\left|\mathbb{E}\frac{(X_1-X_2)(Y_1-Y_2)}{\|\mathbf{Z}_1-\mathbf{Z}_2\|}\right| \leq \sqrt{\mathbb{E}\frac{(X_1-X_2)^2}{\|\mathbf{Z}_1-\mathbf{Z}_2\|}\mathbb{E}\frac{(Y_1-Y_2)^2}{\|\mathbf{Z}_1-\mathbf{Z}_2\|}}$ and hence $|\rho_g(X, Y)| \leq 1$.

Let $(X_i, Y_i), i = 1, 2$, be independent copies of (X, Y), then $\text{cov}_g(X, Y) = 2\mathbb{E}\frac{X_1(Y_1-Y_2)}{\|\mathbf{Z}_1-\mathbf{Z}_2\|} = 2\mathbb{E}\frac{X_1Y_1}{\|\mathbf{Z}_1-\mathbf{Z}_2\|} - 2\mathbb{E}\frac{X_1Y_2}{\|\mathbf{Z}_1-\mathbf{Z}_2\|} = 0$ by symmetry. Hence $\rho_g(X, Y) = 0$.

If $Y = aX + b$, then

$$\rho_g(X, Y) = \frac{\mathbb{E}\frac{(X_1-X_2)(Y_1-Y_2)}{\|\mathbf{Z}_1-\mathbf{Z}_2\|}}{\sqrt{\mathbb{E}\frac{(X_1-X_2)^2}{\|\mathbf{Z}_1-\mathbf{Z}_2\|}\mathbb{E}\frac{(Y_1-Y_2)^2}{\|\mathbf{Z}_1-\mathbf{Z}_2\|}}} = \frac{\frac{a}{\sqrt{a^2+1}}\mathbb{E}|X_1 - X_2|}{\frac{|a|}{\sqrt{a^2+1}}\mathbb{E}|X_1 - X_2|} = \text{sgn}(a).$$

$\rho_g(aX + b, aY + d) = \rho_g(X, Y)$ can be obtained from

$$\text{cov}_g(aX + b, aY + d) = \mathbb{E}\frac{a^2(X_1 - X_2)(Y_1 - Y_2)}{|a|\|\mathbf{Z}_1 - \mathbf{Z}_2\|} = |a|\text{cov}_g(X, Y),$$

$$\text{cov}_g(aX + b, aX + b) = |a|\text{cov}_g(X, X),$$

$$\text{cov}_g(aY + d, aY + d) = |a|\text{cov}_g(Y, Y).$$

By (3.2.2), (3.2.3) and (3.2.4), it is easy to see the remainder of property 5. □

Theorem 3.2.1 shows that the symmetric Gini correlation has all of the properties of the Pearson correlation coefficient except Property 5. It loses the invariance property under heterogeneous changes because of the Euclidean norm in the spatial rank function. To overcome this drawback, we give the affine invariant version of the ρ_g in Section 3.5. Compared with the Pearson correlation, as we will see in Section 3.3, the Gini correlation is more robust in terms of its influence function.

3.2 Symmetric Gini Correlation for Elliptical Distributions

The relationship between Kendall's τ and the linear correlation coefficient ρ, $\tau = 2/\pi \arcsin(\rho)$, holds for all elliptical distributions. So $\rho = \sin(\pi\tau/2)$ provides a robust estimation method for ρ by estimating τ ([29]). This motivates us to explore the relationship between the symmetric Gini correlation ρ_g and the linear correlation coefficient ρ under elliptical distributions.

Definition 3.2.7 (Elliptical Distribution)**.** *A d-dimensional continuous random vector* $\boldsymbol{Z}$ *has an elliptical distribution if its density function is of the form*

$$f(\boldsymbol{z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-1/2} g\{(\boldsymbol{z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{z} - \boldsymbol{\mu})\}, \tag{3.2.8}$$

where $\boldsymbol{\Sigma}$ *is the scatter matrix,* μ *is the location parameter and the nonnegative function* g *is the density generating function.*

An important property for the elliptical distribution is that the nonnegative random variable $R = ||\boldsymbol{\Sigma}^{-1/2}(\mathbf{Z} - \boldsymbol{\mu})||$ is independent of $\boldsymbol{U} = \{\boldsymbol{\Sigma}^{-1/2}(\mathbf{Z} - \boldsymbol{\mu})\}/R$, which is uniformly distributed on the unit sphere. When $d = 1$, the class of elliptical distributions coincides with the location-scale class. For $d = 2$, let $\mathbf{Z} = (X, Y)^T$ and Σ_{ij} be the (i, j) element of $\boldsymbol{\Sigma}$, then the *linear correlation coefficient* of X and Y is $\rho = \rho(X, Y) := \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}$. If the second moment of $\mathbf{Z}$ exists, then the scatter parameter $\boldsymbol{\Sigma}$ is proportional to the covariance matrix. Thus the Pearson correlation ρ_p is well defined and is equal to the parameter ρ in the elliptical distributions. If $\Sigma_{11} = \Sigma_{22} = \sigma^2$, we say X and Y are homogeneous, and $\boldsymbol{\Sigma}$ can then be written as $\boldsymbol{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. In this case, if $\rho = \pm 1$, $\boldsymbol{\Sigma}$ is singular and the distribution reduces to an one-dimensional distribution.

The following theorem states the relationship between ρ_g and ρ under elliptical distributions.

Theorem 3.2.2. *If* $\boldsymbol{Z} = (X,Y)^T$ *has an elliptical distribution* H *with finite first moment and the scatter matrix* $\boldsymbol{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, *then we have*

$$\rho_g = k(\rho) = \begin{cases} \rho & \rho = 0, \pm 1, \\ \dfrac{1}{\rho} + \dfrac{\rho - 1}{\rho} \dfrac{EK(\frac{2\rho}{\rho+1})}{EE(\frac{2\rho}{\rho+1})}, & \textit{otherwise}, \end{cases} \tag{3.2.9}$$

where

$$EK(x) = \int_0^{\pi/2} \frac{1}{\sqrt{1 - x^2 \sin^2 \theta}}\, d\theta \quad \textit{and} \quad EE(x) = \int_0^{\pi/2} \sqrt{1 - x^2 \sin^2 \theta}\, d\theta$$

are the complete elliptic integral of the first kind and the second kind, respectively.

Proof. To prove the theorem, we need a result from [7]. They consider the Gini covariance matrix $\boldsymbol{\Sigma}_g = 2\mathbb{E}\mathbf{Z}\mathbf{r}^T(\mathbf{Z})$. Their Theorem 2.1 states that if the scatter matrix $\boldsymbol{\Sigma}$ has the spectral decomposition $V\Lambda V^T$ with $\Lambda = \text{diag}(\lambda_1, \lambda_2)$, then $\boldsymbol{\Sigma}_g = V\Lambda_g V^T$ with $\Lambda_g = \text{diag}(\lambda_{g,1}, \lambda_{g,2})$ and

$$\lambda_{g,i} = c(H)\mathbb{E}\left[\frac{\lambda_i u_i^2}{\sqrt{\lambda_1 u_1^2 + \lambda_2 u_2^2}}\right], \qquad i = 1, 2 \tag{A1}$$

where $\boldsymbol{u} = (u_1, u_2)^T$ is uniformly distributed on the unit circle, λ_i's are the eigenvalues of $\boldsymbol{\Sigma}$ and $c(H)$ is a constant depending on the distribution H. Here the eigenvalues of $\boldsymbol{\Sigma}$ are $\lambda_1 = \sigma^2(1-\rho)$ and $\lambda_2 = \sigma^2(1+\rho)$, and the corresponding eigenvectors are $(1,-1)^T$ and $(1,1)^T$. Consequently, $\rho_g = \frac{\lambda_{g,2} - \lambda_{g,1}}{\lambda_{g,2} + \lambda_{g,1}}$. Obviously, if $\rho = \pm 1$, either λ_1 or λ_2 is zero. With (A1), we have $\rho_g = \pm 1 = \rho$. If $\rho = 0$, then $\lambda_1 = \lambda_2$, and hence we have $\lambda_{g,1} = \lambda_{g,2}$ and $\rho_g = 0 = \rho$. When $|\rho| < 1$ and $\rho \neq 0$, let $u_1 = \cos\theta$ and $u_2 = \sin\theta$, then θ is uniformly distributed in $[0, 2\pi]$. With $(A1)$, we have

$$\rho_g = \frac{\int_0^{2\pi} \frac{1}{2\pi} \frac{(1-\rho)\cos^2\theta - (1+\rho)\sin^2\theta}{\sqrt{(1-\rho)\cos^2\theta + (1+\rho)\sin^2\theta}} d\theta}{\int_0^{2\pi} \frac{1}{2\pi} \sqrt{(1-\rho)\cos^2\theta + (1+\rho)\sin^2\theta} d\theta} = \frac{\int_0^{\pi/2} \frac{\rho - \cos 2\theta}{\sqrt{1-\rho\cos 2\theta}} d\theta}{\int_0^{\pi/2} \sqrt{1-\rho\cos 2\theta} d\theta}$$
$$= \frac{1}{\rho} + \frac{\rho - 1}{\rho} \frac{EK(2\rho/(\rho+1))}{EE(2\rho/(\rho+1))}.$$

□

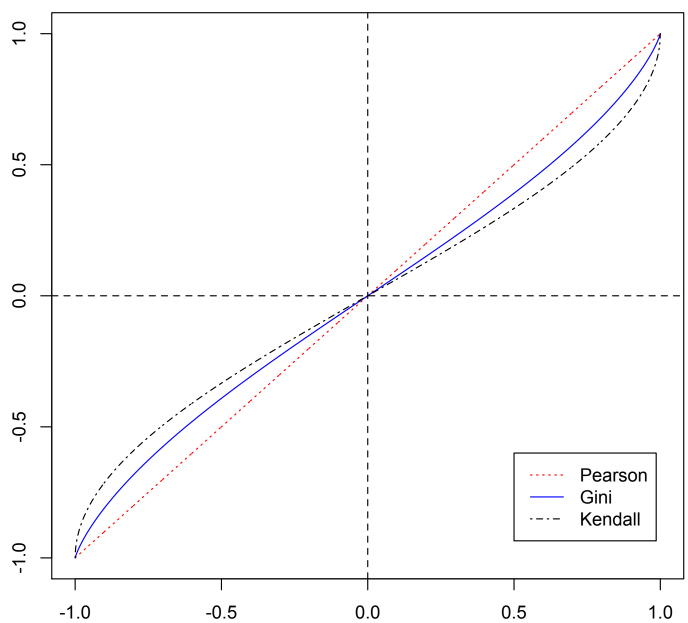

Figure 3.2.1: Pearson ρ_p, Kendall's τ and symmetric Gini ρ_g correlation coefficients versus ρ, the correlation parameter of homogeneous elliptical distributions with finite second moment.

The relationship (3.2.9) holds only for $\mathbf{\Sigma}$ with $\Sigma_{11} = \Sigma_{22}$ because of the loss of invariance property of ρ_g under the heterogeneous changes (Theorem 3.2.1). Note that for any elliptical distribution, the regular Gini correlations are equal to ρ. [36] proved that $\gamma(X, Y) = \gamma(Y, X) = \rho$ for bivariate normal distributions, but their proof can be modified for all elliptical distributions. Based on the spatial sign covariance matrix, [10] considered a spatial sign correlation coefficient, which equals to ρ for elliptical distributions.

Figure 3.2.1 plots the proposed symmetric Gini correlation ρ_g as a function of ρ under homogeneous elliptical distributions with finite second moment. In comparison, we also plot Pearson ρ_p and Kendall's τ against ρ. All correlations are increasing in ρ. It is clear that $|\tau| < |\rho_g| < |\rho_p| = |\rho|$.

With (3.2.9), the estimate $\hat{\rho}_g$ of ρ_g can be corrected to ensure Fisher consistency by using the inversion transformation $k^{-1}(\hat{\rho}_g)$, denoted as $\hat{\rho}^g$. In the next section, we study the influence function of ρ_g, which can be used to evaluate robustness and efficiency of the estimators $\hat{\rho}_g$ in any distribution and that of $\hat{\rho}^g$ under elliptical distributions.

3.3 INFLUENCE FUNCTION

The influence function (IF) introduced by Hampel ([17]) is now a standard tool in robust statistics for measuring effects on estimators due to infinitesimal perturbations of sample distribution functions ([18]).

Definition 3.3.1 (Influence Function). *For a cdf H on $\mathbb{R}^d$ and a functional $T : H \mapsto T(H) \in \mathbb{R}^m$ with $m \geq 1$, the IF of T at H is defined as*

$$IF(\boldsymbol{z}; T, H) = \lim_{\varepsilon \downarrow 0} \frac{T((1-\varepsilon)H + \varepsilon\delta_{\mathcal{B}z}) - T(H)}{\varepsilon}, \quad \boldsymbol{z} \in \mathbb{R}^d,$$

where $\delta_{\boldsymbol{z}}$ denotes the point mass distribution at $\boldsymbol{z}$.

Under regularity conditions on T ([18, 41]), we have $\mathbb{E}_H\{\text{IF}(\mathbf{Z}; T, H)\} = \mathbf{0}$ and the von Mises expansion

$$T(H_n) - T(H) = \frac{1}{n}\sum_{i=1}^{n} \text{IF}(\mathbf{z}_i; T, H) + o_p(n^{-1/2}), \tag{3.3.2}$$

where H_n denotes the empirical distribution based on sample $\mathbf{z}_1,...,\mathbf{z}_n$. This representation shows the connection of the IF with robustness of T, observation by observation. Furthermore, (3.3.2) yields asymptotic m-variate normality of $T(H_n)$,

$$\sqrt{n}(T(H_n) - T(H)) \xrightarrow{d} N(\mathbf{0}, \mathbb{E}_H(\mathrm{IF}(\mathbf{Z};T,H)\mathrm{IF}(\mathbf{Z};T,H)^T) \text{ as } n \to \infty. \tag{3.3.3}$$

To find the influence function of the symmetric Gini correlation defined in (3.2.6), let $T_1(H) = 2\mathbb{E}XR_1(\mathbf{Z})$, $T_2(H) = 2\mathbb{E}XR_2(\mathbf{Z})$, $T_3(H) = 2\mathbb{E}YR_2(\mathbf{Z})$ and $h(t_1,t_2,t_3) = t_2/\sqrt{t_1t_3}$. Then $\rho_g = T(H) = h(T_1,T_2,T_3)$. Denote the influence function of T_i as $L_i(x,y) = \mathrm{IF}((x,y)^T;T_i,H)$ for $i = 1,2,3$.

Theorem 3.3.1. *For any distribution H with finite first moment, the influence function of $\rho_g = T(H)$ is given by*

$$\begin{aligned} IF((x,y)^T;\rho_g,H) =& -\frac{\rho_g}{2}\left(\frac{L_1(x,y)}{T_1} - \frac{2L_2(x,y)}{T_2} + \frac{L_3(x,y)}{T_3}\right) \\ =& -\frac{\rho_g}{2}\left(\frac{1}{T_1}\int \frac{2(x-x_1)^2}{\sqrt{(x-x_1)^2+(y-y_1)^2}}dH(x_1,y_1)\right. \\ & -\frac{1}{T_2}\int \frac{4(x-x_1)(y-y_1)}{\sqrt{(x-x_1)^2+(y-y_1)^2}}dH(x_1,y_1) \\ & \left. +\frac{1}{T_3}\int \frac{2(y-y_1)^2}{\sqrt{(x-x_1)^2+(y-y_1)^2}}dH(x_1,y_1)\right). \end{aligned}$$

Proof. Let $\tilde{H} = (1-\varepsilon)H + \varepsilon\delta_{(x,y)}$, then

$$\begin{aligned} T_1(\tilde{H}) &= 2\iint \frac{x_1(x_1-x_2)}{\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}}d\tilde{H}(x_2,y_2)d\tilde{H}(x_1,y_1) \\ &= 2(1-\varepsilon)^2T_1(H) + 2\varepsilon(1-\varepsilon)\int \frac{(x-x_2)^2}{\sqrt{(x-x_2)^2+(y-y_2)^2}}dH(x_2,y_2). \end{aligned}$$

We have

$$L_1(x,y) = 2\int \frac{(x-x_2)^2}{\sqrt{(x-x_2)^2+(y-y_2)^2}} dH(x_2,y_2) - 4T_1(H),$$
$$L_2(x,y) = 2\int \frac{(x-x_2)(y-y_2)}{\sqrt{(x-x_2)^2+(y-y_2)^2}} dH(x_2,y_2) - 4T_2(H),$$
$$L_3(x,y) = 2\int \frac{(y-y_2)^2}{\sqrt{(x-x_2)^2+(y-y_2)^2}} dH(x_2,y_2) - 4T_3(H).$$

Hence,

$$\begin{aligned}\mathrm{IF}((x,y)^T;\rho_g,H) &= \sum_{i=1}^{3} \frac{\partial h}{\partial t_i}\Big|_T L_i(x,y) \\ &= -\frac{T_2}{2\sqrt{T_1^3 T_3}} L_1(x,y) - \frac{T_2}{2\sqrt{T_1 T_3^3}} L_3(x,y) + \frac{1}{\sqrt{T_1 T_3}} L_2(x,y).\end{aligned}$$

Replacing $T_2/\sqrt{T_1 T_3}$ with ρ_g completes the proof. □

Note that each of $L_i(x,y)$ is approximately linear in x or y. Comparing with the quadratic effects in the Pearson's correlation coefficient ([8]),

$$\mathrm{IF}((x,y)^T;\rho_p,H) = \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} - \frac{1}{2}\rho\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right],$$

ρ_g is more robust than the Pearson correlation. However, ρ_g is not as robust as Kendall's τ correlation since the influence function of ρ_g is unbounded. Kendall's τ correlation has a bounded influence function ([6]), which is $\mathrm{IF}((x,y)^T;\tau,H) = 2\{2P_H[(x-X)(y-Y) > 0] - 1 - \tau\}$. In this sense, ρ_g is more robust than ρ_p but less robust than τ.

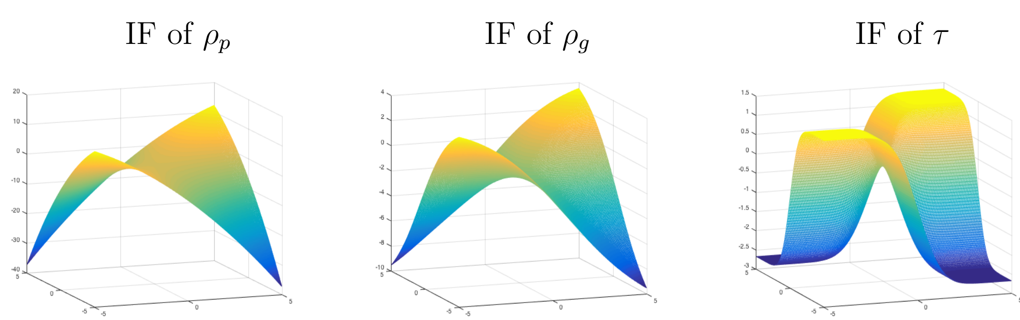

Figure 3.3.1: Influence functions of correlation correlations ρ_p, ρ_g and τ for the bivariate normal distribution with $\mu_x = \mu_y = 0$, $\sigma_x = \sigma_y = 1$ and $\rho = 0.5$.

Figure 3.3.1 displays the influence function of each correlation coefficient for the bivariate normal distribution with $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$ and $\rho = 0.5$. Note that scales of the value of the influence functions in the three plots are quite different.

3.4 ESTIMATION

Let $\mathbf{z}_i = (x_i, y_i)^T$, and $\mathcal{Z} = (\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_n)$ be a random sample from a continuous distribution H with an empirical distribution H_n. Replacing H in (3.2.6) with H_n, we have the sample counterpart of the symmetric Gini correlation coefficient $\rho_g(H_n) = \hat{\rho}_g$:

$$\hat{\rho}_g = \frac{\sum_{1\le i<j\le n} \frac{(x_i-x_j)(y_i-y_j)}{\sqrt{(x_i-x_j)^2+(y_i-y_j)^2}}}{\sqrt{\sum_{1\le i<j\le n} \frac{(x_i-x_j)^2}{\sqrt{(x_i-x_j)^2+(y_i-y_j)^2}}}\sqrt{\sum_{1\le i<j\le n} \frac{(y_i-y_j)^2}{\sqrt{(x_i-x_j)^2+(y_i-y_j)^2}}}}.$$

Using the same notation as in Section 3.3, we have the following central limit theorem of the sample symmetric Gini correlation $\hat{\rho}_g$.

Theorem 3.4.1. *Let $\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_n$ be a random sample from 2-dimensional distribution H with finite second moment. Then $\hat{\rho}_g$ is an unbiased, $\sqrt{n}$-consistent estimator of ρ_g. Furthermore, $\sqrt{n}(\hat{\rho}_g - \rho_g) \xrightarrow{d} N(0, v_g)$ as $n \to \infty$, where*

$$\begin{aligned} v_g = \mathbb{E}[IF((X,Y)^T, \rho_g, H)]^2 = \frac{\rho_g^2}{4}\bigg(&\frac{1}{T_1^2}\mathbb{E}[L_1^2(X,Y)] + \frac{4}{T_2^2}\mathbb{E}[L_2^2(X,Y)] \\ &+ \frac{1}{T_3^2}\mathbb{E}[L_3^2(X,Y)] - \frac{4}{T_1T_2}\mathbb{E}L_1(X,Y)L_2(X,Y) + \frac{2}{T_1T_3}\mathbb{E}L_1(X,Y)L_3(X,Y) \\ &- \frac{4}{T_2T_3}\mathbb{E}L_2(X,Y)L_3(X,Y)\bigg). \end{aligned}$$

Proof. Let $\boldsymbol{\Sigma}_g$ be the Gini covariance matrix of $\mathbf{Z} = (X, Y)^T$ and $\hat{\boldsymbol{\Sigma}}_g = \begin{pmatrix} G_x^2 & G_{xy} \\ G_{xy} & G_y^2 \end{pmatrix}$ be the sample Gini covariance matrix for sample $\{\mathbf{Z}_i\}_{i=1}^n$. Let $\text{vec}(M)$ be the operator that stacks the columns of M to form a vector. According to Theorem 4.1 of [7], we get

$$\sqrt{n}(vec(\hat{\mathbf{\Sigma}}_g) - vec(\mathbf{\Sigma}_g)) \xrightarrow{d} N_4(\mathbf{0}, \mathbf{V}),$$

where $\mathbf{V} = 4\mathbb{E}[\boldsymbol{\psi}(\mathbf{Z})\boldsymbol{\psi}^T(\mathbf{Z})]$, $\boldsymbol{\psi}(\mathbf{z}) = \text{vec}\left(\mathbb{E}[\frac{(\mathbf{z}-\mathbf{Z})(\mathbf{z}-\mathbf{Z})^T}{\|\mathbf{z}-\mathbf{Z}\|}] - \mathbf{\Sigma}_g\right)$. Then $\sqrt{n}\left\{(G_x^2, G_{xy}, G_y^2)^T - (\text{cov}_g(X,X), \text{cov}_g(X,Y), \text{cov}_g(Y,Y))^T\right\} \xrightarrow{d} N_3(\mathbf{0}, \boldsymbol{V}^*)$ with $\boldsymbol{V}^*$ being the matrix of $\mathbf{V}$ deleting the third row and third column. Now, since $\hat{\rho}_g = h(G_x^2, G_{xy}, G_y^2) = G_{xy}/\sqrt{G_x^2 G_y^2}$, and the derivative of h is $\dot{h}(a,b,c) = -b/(2\sqrt{ac})(1/a, -2/b, 1/c)$, we have

$$\dot{h}(\text{cov}_g(X,X), \text{cov}_g(X,Y), \text{cov}_g(Y,Y)) = \frac{-\rho_g}{2}\left(\frac{1}{\text{cov}_g(X,X)}, \frac{-2}{\text{cov}_g(X,Y)}, \frac{1}{\text{cov}_g(Y,Y)}\right),$$

which is denoted as B. Applying the delta method yields the asymptotic normality of $\hat{\rho}_g$ with the asymptotic variance $v_g = B\mathbf{V}^*B^T$. Working out the explicit form of v_g completes the proof. □

Although (3.3.3) implies Theorem 3.4.1, it is hard to check regularity conditions for the von Mises expansion (3.3.2). Instead, we prove it using the multivariate delta method and the asymptotic normality of the sample Gini covariance matrix, which is based on the U-statistics theory ([7]).

For an elliptical distribution H, Theorem 3.2.2 shows that $\hat{\rho}_g$ is not a Fisher consistent estimator of ρ. We need to consider the inverse transformation $\hat{\rho}^g = k^{-1}(\hat{\rho}_g)$, where the function k is given in (3.2.9). Applying the delta method, we obtain the $\sqrt{n}$-consistency of estimator $\hat{\rho}^g$ for ρ.

Theorem 3.4.2. *Let $\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_n$ be a sample from elliptical distribution H with finite second moment and $\mathbf{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. Then $\hat{\rho}^g = k^{-1}(\hat{\rho}_g)$ is unbiased and a $\sqrt{n}$-consistent estimator of ρ. Moreover, $\sqrt{n}(\hat{\rho}^g - \rho) \xrightarrow{d} N(0, [1/k'(\rho)]^2 v_g)$ as $n \to \infty$, where the function k*

is given in (3.2.9), v_g is given in Theorem 3.4.1, and $k'(\rho)$ is

$$k'(\rho) = \frac{-3(\rho+1)EE^2(\frac{2\rho}{\rho+1}) + 4EE(\frac{2\rho}{\rho+1})EK(\frac{2\rho}{\rho+1}) + (\rho-1)EK^2(\frac{2\rho}{\rho+1}))}{2(\rho+1)\rho^2 EE^2(\frac{2\rho}{\rho+1})}.$$

Theorem 3.4.2 provides an estimator based on $\hat{\rho}_g$ for the correlation parameter for elliptical distributions. The asymptotic variance $[k'(\rho)]^{-2}v_g$ can be used to evaluate the asymptotic efficiency of $\hat{\rho}^g$.

3.4 Asymptotic Efficiency

To compare relative efficiency, we present the asymptotic variances (ASV) of the other three estimators of ρ including Pearson's estimator $\hat{\rho}_p$, the regular Gini correlation estimator $\hat{\rho}_\gamma$, and Kendall's estimator $\hat{\rho}_\tau$.

Witting and Müller-Funk [47] established asymptotic normality for the regular sample Pearson correlation coefficient $\hat{\rho}_p$:

$$\sqrt{n}(\hat{\rho}_p - \rho) \xrightarrow{d} N(0, v_p) \ \ as \ \ n \to \infty,$$

where

$$v_p = (1 + \frac{\rho^2}{2})\frac{\sigma_{22}}{\sigma_{20}\sigma_{02}} + \frac{\rho^2}{4}(\frac{\sigma_{40}}{\sigma_{20}^2} + \frac{\sigma_{04}}{\sigma_{02}^2} - \frac{4\sigma_{31}}{\sigma_{11}\sigma_{20}} - \frac{4\sigma_{13}}{\sigma_{11}\sigma_{02}}),$$

and $\sigma_{kl} = \mathbb{E}[(X - \mathbb{E}X)^k(Y - \mathbb{E}Y)^l]$. The Pearson correlation estimator requires a finite fourth moment on the distribution to evaluate its asymptotic variance. For bivariate normal distributions, the asymptotic variance v_p simplifies to $(1-\rho^2)^2$.

An estimator $\hat{\rho}_\gamma$ of the regular Gini correlation $\gamma(X,Y)$ is

$$\hat{\rho}_\gamma = \frac{\binom{n}{2}^{-1}\sum_{1\le i<j\le n} h_1(\mathbf{z}_i, \mathbf{z}_j)}{\binom{n}{2}^{-1}\sum_{1\le i<j\le n} h_2(\mathbf{z}_i, \mathbf{z}_j)},$$

where

$$h_1(\mathbf{z}_1, \mathbf{z}_2) = [(x_1 - x_2)I(y_1 > y_2) + (x_2 - x_1)I(y_2 > y_1)]/4,$$

$$h_2(\mathbf{z}_1, \mathbf{z}_2) = |x_1 - x_2|/4.$$

Using U-statistic theory, Schechtman and Yitzhaki [36] provided the asymptotic normality:

$$\sqrt{n}(\hat{\rho}_\gamma - \rho_\gamma) \xrightarrow{d} N(0, v_\gamma) \quad as \quad n \to \infty,$$

with

$$v_\gamma = (4/\theta_2^2)\zeta_1(\theta_1) + (4\theta_1^2/\theta_2^4)\zeta_2(\theta_2) - (8\theta_1/\theta_2^3)\zeta_3(\theta_1, \theta_2),$$

where

$$\theta_1 = \text{cov}(X, G(Y)), \quad \theta_2 = \text{cov}(X, F(X)),$$

$$\zeta_1(\theta_1) = \mathbb{E}_{\mathbf{Z}_1}\{\mathbb{E}_{\mathbf{Z}_2}[h_1(\mathbf{Z}_1, \mathbf{Z}_2)]\}^2 - \theta_1^2,$$

$$\zeta_2(\theta_2) = \mathbb{E}_{\mathbf{Z}_1}\{\mathbb{E}_{\mathbf{Z}_2}[h_2(\mathbf{Z}_1, \mathbf{Z}_2)]\}^2 - \theta_2^2,$$

$$\zeta_3(\theta_1, \theta_2) = \mathbb{E}_{\mathbf{Z}_1}\{\mathbb{E}_{\mathbf{Z}_2}[h_1(\mathbf{Z}_1, \mathbf{Z}_2)]\mathbb{E}_{\mathbf{Z}_2}[h_2(\mathbf{Z}_1, \mathbf{Z}_2)]\} - \theta_1\theta_2.$$

Under elliptical distributions, $\gamma(X, Y) = \gamma(Y, X) = \rho$, hence the asymptotic variance of $\hat{\rho}_\gamma$ is v_γ. For a normal distribution, [53] provided an explicit formula for v_γ, given by $v_\gamma = \pi/3 + (\pi/3 + 4\sqrt{3})\rho^2 - 4\rho \arcsin(\rho/2) - 4\rho^2\sqrt{4 - \rho^2}$.

Borovskikh [3] presented the asymptotic normality of the estimator $\hat{\tau}$:

$$\sqrt{n}(\hat{\tau} - \tau) \xrightarrow{d} N(0, v_\tau) \;\; as \;\; n \to \infty,$$

with

$$v_\tau = 4\mathbb{E}\{\mathbb{E}^2_{\mathbf{z}_1}\{\text{sgn}[(X_2 - X_1)(Y_2 - Y_1)]\}\} - 4\mathbb{E}^2\{\text{sgn}[(X_2 - X_1)(Y_2 - Y_1)]\}.$$

Applying the delta method to $\hat{\rho}_\tau = \sin(\pi\hat{\tau}/2)$, we obtain the asymptotic variance of $\hat{\rho}_\tau$ to be $\frac{\pi^2}{4}(1-\rho^2)v_\tau$. Under a normal distribution, the asymptotic variance of $\hat{\rho}_\tau$ is $\pi^2(1-\rho^2)[\frac{1}{9} - \frac{4}{\pi^2}\arcsin^2(\frac{\rho}{2})]$ ([6]).

We compare the asymptotic efficiency of the four estimators $\hat{\rho}^g$, $\hat{\rho}_\gamma$, $\hat{\rho}_\tau$ and $\hat{\rho}_p$ under three bivariate elliptical distributions (3.2.8): the normal distributions with $g(t) = \frac{1}{2\pi}e^{-t/2}$; the t-distributions with $g(t) = \frac{1}{2\pi}(1 + t/\nu)^{-\nu/2-1}$, where ν is the degrees of freedom; and the Kotz type distribution with $g(t) = \frac{1}{2\pi}e^{-\sqrt{t}}$. The normal distribution is the limiting distribution of the t-distributions as $\nu \to \infty$. The Kotz type distribution is a bivariate generalization of the Laplace distribution with the tail region fatness between that of the normal and t distributions ([11]). We consider only elliptical distributions because all four estimators $\hat{\rho}^g$, $\hat{\rho}_\gamma$, $\hat{\rho}_\tau$ and $\hat{\rho}_p$ are Fisher consistent for parameter ρ. The estimators for non-elliptical distributions may estimate different quantities, resulting in their asymptotic variances being incomparable.

Table 3.4.1: Asymptotic relative efficiencies (ARE) of estimators $\hat{\rho}^g$, $\hat{\rho}_\gamma$ and $\hat{\rho}_\tau$ relative to $\hat{\rho}_p$ for different distributions, with asymptotic variance (ASV($\hat{\rho}_p$)) of Pearson estimator $\hat{\rho}_p$.

Distribution		ARE($\hat{\rho}^g, \hat{\rho}_p$)	ARE($\hat{\rho}_\gamma, \hat{\rho}_p$)	ARE($\hat{\rho}_\tau, \hat{\rho}_p$)	ASV($\hat{\rho}_p$)
	$\rho = 0.1$	0.9321	0.9558	0.9125	0.9816
Normal	$\rho = 0.5$	0.9769	0.9398	0.8925	0.5631
	$\rho = 0.9$	0.9601	0.9004	0.8439	0.0361
	$\rho = 0.1$	1.0182	1.0304	1.0146	1.1558
$t(15)$	$\rho = 0.5$	1.0560	0.9852	0.9896	0.6643
	$\rho = 0.9$	1.0289	0.9468	0.8804	0.0427
	$\rho = 0.1$	2.0095	1.9502	2.2586	2.8800
$t(5)$	$\rho = 0.5$	1.9795	1.7666	2.1060	1.5961
	$\rho = 0.9$	1.8629	1.5346	1.7940	0.1019
	$\rho = 0.1$	1.2081	1.1385	1.2171	1.6382
Kotz	$\rho = 0.5$	1.1850	1.0854	1.1510	0.9378
	$\rho = 0.9$	1.1599	0.9789	1.0256	0.0602

Without loss of generality, we consider only cases with $\rho > 0$. Listed in Table 3.4.1 are asymptotic variances (ASV) of Pearson estimator $\hat{\rho}_p$, and asymptotic relative efficiencies (ARE) of estimators $\hat{\rho}^g$, $\hat{\rho}_\gamma$ and $\hat{\rho}_\tau$ relative to $\hat{\rho}_p$ for different elliptical distributions under the homogeneous assumption, where the asymptotic relative efficiency of one estimator with respect to another is defined as $\text{ARE}(\hat{\rho}_1, \hat{\rho}_2) = \text{ASV}(\hat{\rho}_2)/\text{ASV}(\hat{\rho}_1)$. The asymptotic variance of each estimator is obtained using a combination of numeric integration and Monte Carlo simulation.

Table 3.4.1 shows that the asymptotic variances of $\hat{\rho}_p$, $\hat{\rho}^g$, $\hat{\rho}_\gamma$ and $\hat{\rho}_\tau$ all decrease as ρ increases. When $\rho = 1$, every estimator is equal to 1 without any estimation error. Asymptotic variances increase for t distributions as the degrees of freedom ν decrease. Under normal distributions, the Pearson correlation estimator is the maximum likelihood estimator of ρ, thus is most efficient asymptotically. The symmetric Gini estimator $\hat{\rho}^g$ is high in efficiency with ARE's greater than 93%; it is more efficient than Kendall's estimator $\hat{\rho}_\tau$. For heavy-tailed distributions, the symmetric Gini estimator is more efficient than Pearson's estimator $\hat{\rho}_p$. The AREs of the symmetric Gini estimator are close to those of Kendall's estimator $\hat{\rho}_\tau$ for Kotz samples. Comparing with the regular Gini correlation estimator, the proposed measure has higher efficiency for all cases except for $\rho = 0.1$ under normal and $t(15)$ distributions, in which cases the efficiency is about 2.4% and 1.2% lower, respectively. These results may be explained by the fact that the joint spatial rank used in $\hat{\rho}^g$ takes more dependence information than the marginal rank used in $\hat{\rho}_\gamma$.

In summary, the proposed symmetric Gini estimator has nice asymptotic behavior that well balances between efficiency and robustness. It is more efficient than the regular Gini, which is also symmetric under elliptical distributions.

3.4 Finite Sample Efficiency

We conduct a small simulation to study the finite sample efficiencies of the symmetric Gini, regular Gini, Kendall's τ and Pearson correlation estimators. $M = 3000$ samples of two different sample sizes, $n = 30, 300$, are drawn from t-distributions with $1, 3, 5, 15$ and ∞ degrees of freedom and from the Kotz distribution. We use the R Package "mnormt" to generate samples from multivariate t and normal distributions (referred to as $t(\infty)$ in Table 3.4.2). For the Kotz sample, we first generate uniformly distributed random vectors on the unit circle by $\mathbf{u} = (\cos\theta, \sin\theta)^T$ with θ in $[0, 2\pi]$, and then generate r from a Gamma distribution with $\alpha = 2$ (the shape parameter) and $\beta = 1$ (the scale parameter). Hence $\mathbf{\Sigma}^{1/2} r\mathbf{u} + \boldsymbol{\mu}$ is a sample from a bivariate Kotz($\boldsymbol{\mu}, \mathbf{\Sigma}$) distribution. For additional details, see Dang *et al.* [7].

For each sample m, each estimator $\hat{\rho}^{(m)}$ is calculated and the root mean squared error (RMSE) of the estimator is computed as

$$\text{RMSE}(\hat{\rho}) = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\hat{\rho}^{(m)} - \rho)^2}.$$

The procedure is repeated 100 times for computing the mean and the standard deviation of $\sqrt{n}$RMSE. In Table 3.4.2, we report the mean and standard deviation (in parentheses) of $\sqrt{n}$RMSEs of correlation estimators $\hat{\rho}^g$, $\hat{\rho}_\gamma$, $\hat{\rho}_\tau$ and $\hat{\rho}_p$ when the scatter matrix is homogeneous with $\mathbf{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. The case of $n = \infty$ corresponds to the asymptotic standard deviation of each estimator that can be obtained from Table 3.4.1. Since $\hat{\rho}^g$ cannot be given explicitly due to the inverse transformation involved in $\hat{\rho}^g = k^{-1}(\hat{\rho}_g)$, we numerically obtain $\hat{\rho}^g$ by creating a correspondence between s and t, where $s = k(t)$ and t is a very fine grid on $[0, 1]$. $\hat{\rho}_g$ is computed using the R package "ICSNP" spatial.rank function.

In Table 3.4.2, the $\sqrt{n}$RMSEs demonstrate an increasing trend as ρ decreases or as the degrees of freedom ν decrease for t distributions. For $n = 300$, the behavior of each estimator is similar to its asymptotic efficiency behavior. For example, for $n = 300$ and $\rho = 0.5$ under the normal distribution, the $\sqrt{n}$RMSE of $\hat{\rho}_p$ is 0.7534 which is close to the asymptotic standard deviation 0.7504. We include heavy-tailed $t(1)$ and $t(3)$ distributions in

Table 3.4.2: The mean and standard deviation (in parentheses) of $\sqrt{n}$RMSE of $\hat{\rho}^g$, $\hat{\rho}_\gamma$, $\hat{\rho}_\tau$ and $\hat{\rho}_p$ under different distributions with a homogeneous scatter matrix.

Dist	ρ	n	$\sqrt{n}$RMSE($\hat{\rho}^g$)	$\sqrt{n}$RMSE($\hat{\rho}_\gamma$)	$\sqrt{n}$RMSE($\hat{\rho}_\tau$)	$\sqrt{n}$RMSE($\hat{\rho}_p$)
$t(\infty)$	$\rho = 0.1$	$n = 30$	0.7767 (.0115)	1.0418 (.0115)	1.0785 (.0120)	1.0095 (.0120)
		$n = 300$	*0.9648 (.0104)*	*1.0184 (.0121)*	*1.0427 (.0139)*	*0.9925 (.0121)*
	$\rho = 0.5$	$n = 30$	0.7887 (.0110)	0.8150 (.0115)	0.8517 (.0126)	0.7827 (.0115)
		$n = 300$	*0.7638 (.0087)*	*0.7777 (.0104)*	*0.8002 (.0104)*	*0.7534 (.0104)*
	$\rho = 0.9$	$n = 30$	0.2147 (.0044)	0.2306 (.0044)	0.2541 (.0049)	0.2103 (.0044)
		$n = 300$	*0.1957 (.0017)*	*0.2026 (.0035)*	*0.2113 (.0035)*	*0.1923 (.0017)*
$t(15)$	$\rho = 0.1$	$n = 30$	0.8013 (.0120)	1.0828 (.0120)	1.1026 (.0115)	1.0735 (.0115)
		$n = 300$	*1.0011 (.0104)*	*1.0669 (.0121)*	*1.0721 (.0139)*	*1.0756 (.0121)*
	$\rho = 0.5$	$n = 30$	0.8177 (.0115)	0.8506 (.0126)	0.8731 (.0131)	0.8347 (.0126)
		$n = 300$	*0.7985 (.0104)*	*0.8227 (.0104)*	*0.8279 (.0104)*	*0.8193 (.0104)*
	$\rho = 0.9$	$n = 30$	0.2251 (.0044)	0.2432 (.0044)	0.2635 (.0164)	0.2262 (.0044)
		$n = 300$	*0.2044 (.0035)*	*0.2165 (.0035)*	*0.2200 (.0035)*	*0.2078 (.0035)*
$t(5)$	$\rho = 0.1$	$n = 30$	0.8698 (.0137)	1.2083 (.0126)	1.1562 (.0131)	1.2987 (.0137)
		$n = 300$	*1.1085 (.0121)*	*1.2246 (.0156)*	*1.1310 (.0139)*	*1.5155 (.0242)*
	$\rho = 0.5$	$n = 30$	0.9032 (.0110)	0.9580 (.0126)	0.9202 (.0126)	1.0221 (.0159)
		$n = 300$	*0.9007 (.0121)*	*0.9492 (.0121)*	*0.8764 (.0121)*	*1.1535 (.0208)*
	$\rho = 0.9$	$n = 30$	0.2569 (.0164)	0.2859 (.0066)	0.2832 (.0164)	0.2908 (.0088)
		$n = 300$	*0.2338 (.0069)*	*0.2615 (.0035)*	*0.2408 (.0069)*	*0.2996 (.0087)*
$t(3)$	$\rho = 0.1$	$n = 30$	0.9706 (.0137)	1.3923 (.0170)	1.2050 (.0142)	1.6459 (.0214)
		$n = 300$	*1.2921 (.0156)*	*1.5329 (.0191)*	*1.1865 (.0156)*	*2.7782 (.0554)*
	$\rho = 0.5$	$n = 30$	1.0231 (.0131)	1.1201 (.0170)	0.9651 (.0148)	1.3343 (.0246)
		$n = 300$	*1.1068 (.0173)*	*1.2142 (.0208)*	*0.9284 (.0121)*	*2.1876 (.0675)*
	$\rho = 0.9$	$n = 30$	0.3127 (.0104)	0.3642 (.0131)	0.3051 (.0066)	0.4289 (.0236)
		$n = 300$	*0.2944 (.0104)*	*0.3672 (.0173)*	*0.2615 (.0035)*	*0.6564 (.0658)*
$t(1)$	$\rho = 0.1$	$n = 30$	1.7418 (.0301)	2.7222 (.0285)	1.3704 (.0170)	3.3104 (.0279)
		$n = 300$	*4.3423 (.0814)*	*6.7879 (.0918)*	*1.3735 (.0173)*	*10.256 (.0918)*
	$\rho = 0.5$	$n = 30$	1.6706 (.0153)	2.3892 (.0361)	1.1184 (.0164)	2.9687 (.0466)
		$n = 300$	*4.2574 (.0485)*	*5.9357 (.1057)*	*1.0999 (.0156)*	*9.1781 (.1472)*
	$\rho = 0.9$	$n = 30$	0.9065 (.0361)	1.2083 (.0586)	0.4004 (.0088)	1.5917 (.0728)
		$n = 300$	*2.1616 (.1074)*	*2.9947 (.1784)*	*0.3464 (.0052)*	*4.9589 (.2182)*
Kotz	$\rho = 0.1$	$n = 30$	0.8692 (.0126)	1.2083 (.0148)	1.1842 (.0148)	1.2389 (.0148)
		$n = 300$	*1.0947 (.0139)*	*1.2055 (.0173)*	*1.1639 (.0156)*	*1.2713 (.0173)*
	$\rho = 0.5$	$n = 30$	0.9037 (.0137)	0.9569 (.0148)	0.9465 (.0142)	0.9711 (.0170)
		$n = 300$	*0.8903 (.0121)*	*0.9318 (.0121)*	*0.9059 (.0121)*	*0.9665 (.0121)*
	$\rho = 0.9$	$n = 30$	0.2563 (.0164)	0.2832 (.0164)	0.2952 (.0060)	0.2706 (.0060)
		$n = 300$	*0.2304 (.0035)*	*0.2529 (.0035)*	*0.2494 (.0035)*	*0.2477 (.0035)*

the simulation to demonstrate finite sample behavior of Pearson and Gini estimators when their asymptotic variances may not exist. $\sqrt{n}$RMSE of $\hat{\rho}_p$ is about twice that of $\hat{\rho}^g$ for $n = 300$ with both $t(1)$ and $t(3)$ distributions. For the $t(1)$ distribution, $\hat{\rho}_\tau$ is much better than the others in terms of $\sqrt{n}$RMSE. When the sample size is small ($n = 30$), $\hat{\rho}^g$ performs the best. The $\sqrt{n}$RMSEs of $\hat{\rho}^g$ are smaller than that of $\hat{\rho}_\tau$ even under heavy-tailed $t(3)$ distributions. $\hat{\rho}^g$ has a smaller $\sqrt{n}$RMSE than the Pearson correlation estimator for the normal distribution with $\rho = 0.1$ and all other distributions. The symmetric Gini estimator $\hat{\rho}^g$ has smaller $\sqrt{n}$RMSE than the regular Gini estimator $\hat{\rho}_\gamma$ for all cases we consider. The simulation demonstrates superior finite sample behavior of the proposed estimator.

3.5 THE AFFINE INVARIANT VERSION OF SYMMETRIC GINI CORRELATION

The proposed ρ_g in Section 3.2 is only invariant under translation and homogeneous change. We now provide an affine invariant version of ρ_g, denoted as ρ_G, in order to gain the invariance property under heterogeneous changes. This is based on the affine equivariant (AE) Gini covariance matrix $\mathbf{\Sigma}_G$ proposed by Dang *et al.* [7].

The basic idea of $\mathbf{\Sigma}_G$ is that the Gini covariance matrix for standardized data should be proportional to the identity matrix $\mathbf{I}$. That is, $\mathbb{E}(\mathbf{\Sigma}_G^{-1/2}\mathbf{Z})\mathbf{r}^T(\mathbf{\Sigma}_G^{-1/2}\mathbf{Z}) = c\mathbf{I}$, where c is a positive constant. In other words, the AE version of the Gini covariance matrix is the solution of

$$\mathbb{E}\frac{\mathbf{\Sigma}_G^{-1/2}(\mathbf{Z}_1 - \mathbf{Z}_2)(\mathbf{Z}_1 - \mathbf{Z}_2)^T\mathbf{\Sigma}_G^{-1/2}}{\sqrt{(\mathbf{Z}_1 - \mathbf{Z}_2)^T\mathbf{\Sigma}_G^{-1}(\mathbf{Z}_1 - \mathbf{Z}_2)}} = c(H)\mathbf{I}, \qquad (3.5.1)$$

where $c(H)$ is a constant depending on H. In this way, the matrix valued functional $\mathbf{\Sigma}_G(\cdot)$ is a scatter matrix in the sense that for any nonsingular matrix A and vector $\mathbf{b}$, $\mathbf{\Sigma}_G(A\mathbf{Z}+\mathbf{b}) = A\mathbf{\Sigma}_G(\mathbf{Z})A^T$.

Let $\mathbf{Z} = (X, Y)^T$ be a bivariate random vector with distribution function H and $\mathbf{\Sigma}_G := \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}$ be the solution of (3.5.1). Then the affine invariant version of ρ_g is defined as $\rho_G(X, Y) = \frac{G_{21}}{\sqrt{G_{11}}\sqrt{G_{22}}}$. Since the value of $c(H)$ in (3.5.1) does not change the value of $\rho_G(X, Y)$, without loss of generality, assume $c(H) = 1$.

Theorem 3.5.1. *For any bivariate random vector* $\boldsymbol{Z} = (X, Y)^T$ *having an elliptical distribution* H *with finite first moment,* $\rho_G(aX, bY) = sgn(ab)\rho_G(X, Y)$ *for any* $ab \neq 0$.

Proof. The proof is straightforward. Let A be the diagonal matrix with the diagonal elements being a and b. Since $\mathbf{\Sigma}_G$ is affine equivariant, $\mathbf{\Sigma}_G(A\mathbf{Z}) = A\mathbf{\Sigma}_G(\mathbf{Z})A^T$. As a result, $\rho_G(aX, bY) = \frac{abG_{21}}{\sqrt{a^2G_{11}}\sqrt{b^2G_{22}}} = \text{sgn}(ab)\rho_G(X, Y)$. □

Remark 3.5.2. Under elliptical distributions, $\rho_G = \rho$. This is true since $\mathbf{\Sigma}_G = \mathbf{\Sigma}$ for elliptical distributions.

When a random sample $\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_n$ is available, replacing H with its empirical distribution H_n in (3.5.1) yields the sample counterpart $\hat{\mathbf{\Sigma}}_G$, and hence the sample $\hat{\rho}_G$ is obtained accordingly. We obtain $\hat{\mathbf{\Sigma}}_G$ by a common re-weighted iterative algorithm:

$$\hat{\mathbf{\Sigma}}_G^{(t+1)} \longleftarrow \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \frac{(\mathbf{z}_i - \mathbf{z}_j)(\mathbf{z}_i - \mathbf{z}_j)^T}{\sqrt{(\mathbf{z}_i - \mathbf{z}_j)^T (\hat{\mathbf{\Sigma}}_G^{(t)})^{-1} (\mathbf{z}_i - \mathbf{z}_j)}}.$$

The initial value can take $\hat{\mathbf{\Sigma}}_G^{(0)} = \mathbf{I}_d$. The iteration stops when $\|\hat{\mathbf{\Sigma}}_G^{(t+1)} - \hat{\mathbf{\Sigma}}_G^{(t)}\| < \varepsilon$ for a pre-specified number $\varepsilon > 0$, where $\| \cdot \|$ can take any matrix norm.

Next, we study finite sample efficiency of $\hat{\rho}_G$ under the same simulation setting as in Section 3.4.2 except that the scatter matrix is heterogeneous. The scatter matrix of each elliptical distribution is $\mathbf{\Sigma} = \begin{pmatrix} 1 & 2\rho \\ 2\rho & 4 \end{pmatrix}$. Table 3.5.1 reports $\sqrt{n}$RMSE of correlation estimators $\hat{\rho}_G$, $\hat{\rho}_\gamma$, $\hat{\rho}_\tau$ and $\hat{\rho}_p$. The numbers in the last three columns are very close to those in Table 3.4.2 because $\hat{\rho}_\gamma$, $\hat{\rho}_\tau$ and $\hat{\rho}_p$ are affine invariant. $\sqrt{n}$RMSEs of $\hat{\rho}_G$ are also

Table 3.5.1: The mean and standard deviation (in parentheses) of $\sqrt{n}$RMSE of $\hat{\rho}_G$, $\hat{\rho}_\gamma$, $\hat{\rho}_\tau$ and $\hat{\rho}_p$ under different distributions with a heterogeneous scatter matrix.

Dist	ρ	n	$\sqrt{n}$RMSE($\hat{\rho}_G$)	$\sqrt{n}$RMSE($\hat{\rho}_\gamma$)	$\sqrt{n}$RMSE($\hat{\rho}_\tau$)	$\sqrt{n}$RMSE($\hat{\rho}_p$)
	$\rho = 0.1$	$n = 30$	1.0171 (.0126)	1.0401 (.0126)	1.0768 (.0131)	1.0073 (.0120)
		$n = 300$	*1.0011 (.0139)*	*1.0133 (.0139)*	*1.0392 (.0156)*	*0.9890 (.0139)*
$t(\infty)$	$\rho = 0.5$	$n = 30$	0.7887 (.0120)	0.8123 (.0126)	0.8501 (.0137)	0.7800 (.0120)
		$n = 300$	*0.7621 (.0104)*	*0.7794 (.0104)*	*0.8002 (.0104)*	*0.7534 (.0104)*
	$\rho = 0.9$	$n = 30$	0.2125 (.0022)	0.2306 (.0044)	0.2541 (.0049)	0.2098 (.0044)
		$n = 300$	*0.1940 (.0035)*	*0.2026 (.0035)*	*0.2113 (.0035)*	*0.1923 (.0017)*
	$\rho = 0.1$	$n = 30$	1.0582 (.0126)	1.0839 (.0131)	1.1042 (.0126)	1.0741 (.0126)
		$n = 300$	*1.0496 (.0121)*	*1.0687 (.0121)*	*1.0739 (.0121)*	*1.0756 (.0121)*
$t(15)$	$\rho = 0.5$	$n = 30$	0.8221 (.0099)	0.8506 (.0099)	0.8731 (.0110)	0.8353 (.0099)
		$n = 300$	*0.7967 (.0104)*	*0.8210 (.0121)*	*0.8279 (.0104)*	*0.8175 (.0121)*
	$\rho = 0.9$	$n = 30$	0.2224 (.0049)	0.2437 (.0049)	0.2635 (.0060)	0.2262 (.0049)
		$n = 300$	*0.2026 (.0035)*	*0.2165 (.0035)*	*0.2200 (.0035)*	*0.2078 (.0035)*
	$\rho = 0.1$	$n = 30$	1.1727 (.0164)	1.2072 (.0148)	1.1557 (.0153)	1.2981 (.0192)
		$n = 300$	*1.1847 (.0156)*	*1.2246 (.0156)*	*1.1310 (.0139)*	*1.5155 (.0242)*
$t(5)$	$\rho = 0.5$	$n = 30$	0.9169 (.0120)	0.9585 (.0115)	0.9213 (.0120)	1.0226 (.0137)
		$n = 300$	*0.8989 (.0139)*	*0.9492 (.0139)*	*0.8764 (.0121)*	*1.1553 (.0242)*
	$\rho = 0.9$	$n = 30$	0.2520 (.0060)	0.2865 (.0071)	0.2832 (.0060)	0.2919 (.0110)
		$n = 300$	*0.2304 (.0035)*	*0.2615 (.0035)*	*0.2408 (.0035)*	*0.2979 (.0087)*
	$\rho = 0.1$	$n = 30$	1.3540 (.0519)	1.3918 (.0159)	1.2039 (.0142)	1.6475 (.0203)
		$n = 300$	*1.4497 (.0225)*	*1.5346 (.0225)*	*1.1847 (.0156)*	*2.7782 (.0606)*
$t(3)$	$\rho = 0.5$	$n = 30$	1.0670 (.0159)	1.1190 (.0170)	0.9629 (.0148)	1.3321 (.0219)
		$n = 300$	*1.1033 (.0139)*	*1.2090 (.0173)*	*0.9249 (.0121)*	*2.1910 (.0606)*
	$\rho = 0.9$	$n = 30$	0.3095 (.0099)	0.3681 (.0137)	0.3062 (.0066)	0.4376 (.0230)
		$n = 300$	*0.2841(.0069)*	*0.3655 (.0156)*	*0.2615 (.0035)*	*0.6461 (.0675)*
	$\rho = 0.1$	$n = 30$	2.7622 (.0274)	2.7244 (.0268)	1.3693 (.0192)	3.3148 (.0268)
		$n = 300$	*6.8381 (.0970)*	*6.7879 (.0797)*	*1.3770 (.0173)*	*10.259 (.0901)*
$t(1)$	$\rho = 0.5$	$n = 30$	2.4133 (.0433)	2.3831 (.0372)	1.1206 (.0164)	2.9643 (.0466)
		$n = 300$	*5.8768 (.1386)*	*5.9132 (.1178)*	*1.0947 (.0139)*	*9.1522 (.1455)*
	$\rho = 0.9$	$n = 30$	1.1875 (.0608)	1.2148 (.0537)	0.4009 (.0088)	1.6015 (.0635)
		$n = 300$	*2.7747 (.2148)*	*2.9930 (.1853)*	*0.3481 (.0052)*	*4.9727 (.2113)*
	$\rho = 0.1$	$n = 30$	1.1672 (.0131)	1.2066 (.0131)	1.1831 (.0142)	1.2368 (.0142)
		$n = 300$	*1.1674 (.0139)*	*1.2038 (.0139)*	*1.1605 (.0139)*	*1.2731 (.0156)*
Kotz	$\rho = 0.5$	$n = 30$	0.9136 (.0148)	0.9574 (.0148)	0.9454 (.0153)	0.9706 (.0148)
		$n = 300$	*0.8885 (.0121)*	*0.9336 (.0121)*	*0.9059 (.0121)*	*0.9665 (.0121)*
	$\rho = 0.9$	$n = 30$	0.2503 (.0049)	0.2815 (.0060)	0.2941 (.0060)	0.2684 (.0055)
		$n = 300$	*0.2269 (.0035)*	*0.2546 (.0035)*	*0.2511 (.0035)*	*0.2477 (.0035)*

close to $\sqrt{n}$RMSE of $\hat{\rho}^g$ for $n = 300$, but are larger than those for $n = 30$ and $\rho = 0.1$. The loss of finite sample efficiency of $\hat{\rho}_G$ for a small sample size under low dependence ρ is probably caused by the iterative algorithm in the computation of $\hat{\rho}_G$. The problem is even worse with the $t(1)$ distribution where the first moment does not exist. As the value of ρ increases, the $\sqrt{n}$RMSE of each estimator decreases for all distributions. Under Kotz and $t(15)$ distributions, the affine invariant Gini estimator $\hat{\rho}_G$ is the most efficient; under $t(5)$ distribution, the $\sqrt{n}$RMSE of $\hat{\rho}_G$ is smaller than that of Kendall's $\hat{\rho}_\tau$ when $\rho = 0.9$. For the normal distributions, $\hat{\rho}_G$ is almost as efficient as $\hat{\rho}_p$ when $\rho = 0.9$. The affine invariant Gini correlation estimator shows a good finite sample efficiency. Again, the proposed Gini correlation estimator has smaller $\sqrt{n}$RMSEs than the regular Gini correlation estimator in all cases.

3.6 APPLICATION

For the purposes of illustration, we apply the symmetric Gini correlations to the famous Fisher's Iris data which is available in R. The data set consists of 50 samples from each of three species of Iris (Setosa, Versicolor and Virginica). Four features are measured in centimeters from each sample: sepal length (Sepal L.), sepal width (Sepal W.), petal length (Petal L.), and petal width (Petal W.). The mean and standard deviation of each of the variables for all data and each species data are listed in Table 3.6.1. All the three species have similar sizes in sepals. Setosa has a much smaller petal size than the other two species. We shall study the correlation of the variables for each Iris species.

Table 3.6.1: Summary Statistics of Variables in Iris Data.

	Mean				Standard Deviation			
	All	Setosa	Vesicolor	Virginica	All	Setosa	Vesicolor	Virginica
Sepal L.	5.843	5.006	5.936	6.588	0.828	0.352	0.516	0.636
Sepal W.	3.057	3.428	2.770	2.974	0.436	0.379	0.314	0.322
Petal L.	3.758	1.462	4.260	5.552	1.765	0.174	0.470	0.552
Petal W.	1.199	0.246	1.326	2.026	0.762	0.105	0.198	0.275

For each Iris species, we compute different correlation measures for all pairs of variables. Since standard deviations of four features are quite different, the affine equivariant version of symmetric gini correlation estimator $\hat{\rho}_G$ is used. For each pair of variables X and Y, we also calculate the Pearson correlation, Kendall's τ and the two regular gini correlation estimators, denoted as $\hat{\gamma}_{1,2}$ ($\hat{\gamma}(X, Y)$) and $\hat{\gamma}_{2,1}$ ($\hat{\gamma}(Y, X)$). All correlation estimators are listed in Table 3.6.2.

Table 3.6.2: Pearson correlation, Kendal's τ, Affine equivariant symmetric Gini correlation and Regular Gini correlations of variables for the Iris data set.

Species	Correlations	Sepal L. & Sepal W.	Sepal L. & Petal L.	Sepal L. & Petal W.	Sepal W. & Petal L.	Sepal W. & Petal W.	Petal L. & Petal W.
	$\hat{\rho}_p$	0.743	0.267	0.278	0.178	0.233	0.332
	$\hat{\tau}$	0.597	0.217	0.231	0.143	0.234	0.222
Setosa	$\hat{\rho}_G$	0.742	0.274	0.285	0.182	0.256	0.312
	$\hat{\gamma}_{1,2}$	0.759	0.283	0.261	0.211	0.214	0.280
	$\hat{\gamma}_{2,1}$	0.781	0.295	0.358	0.174	0.350	0.384
	$\hat{\rho}_p$	0.526	0.754	0.546	0.561	0.664	0.787
	$\hat{\tau}$	0.398	0.567	0.403	0.430	0.551	0.646
Versicolor	$\hat{\rho}_G$	0.546	0.756	0.551	0.584	0.687	0.790
	$\hat{\gamma}_{1,2}$	0.533	0.744	0.542	0.580	0.658	0.787
	$\hat{\gamma}_{2,1}$	0.523	0.766	0.559	0.572	0.682	0.809
	$\hat{\rho}_p$	0.457	0.864	0.281	0.401	0.538	0.322
	$\hat{\tau}$	0.307	0.670	0.219	0.291	0.419	0.271
Virginica	$\hat{\rho}_G$	0.687	0.820	0.455	0.621	0.623	0.519
	$\hat{\gamma}_{1,2}$	0.406	0.867	0.278	0.467	0.567	0.304
	$\hat{\gamma}_{2,1}$	0.476	0.832	0.315	0.308	0.548	0.355

From Table 3.6.2 we see that compared with the other two species, Iris Setosa has higher correlation between sepal length and sepal width, but has lower correlation between sepal length and petal length. Versicolor has much larger correlation between petal length and petal width than the other two species. Virginica has the highest correlation between sepal length and petal length among the three species.

Kendall's τ correlation estimate is the smallest among all correlation estimates across all pairs and across all species. Two regular Gini correlation estimates are quite different especially between sepal width and petal length in Iris Virginica species. The difference is as high as 0.159. One might perform a hypothesis test on exchangeability of two variables by testing $\gamma_{1,2} = \gamma_{2,1}$ ([39]). The p-value of the test is 0.0113, which provides strong evidence to reject the hypothesis of exchangeability of two variables sepal width and petal length in Iris Virginica. We also observe that $\hat{\rho}_G$ and $\hat{\rho}_p$ tend to have a similar pattern across variable pairs and across species. For example, for all six pairs of variables in Iris Setosa, $\hat{\rho}_G$ is large or small whenever $\hat{\rho}_p$ is large or small. In other words, the correlation ranking across variable pairs provided by the Pearson correlation is the same as the ranking by the proposed symmetric Gini correlation. However, such a pattern is not shared by any two correlations from $\hat{\rho}_G$, $\hat{\tau}$, $\hat{\gamma}_{1,2}$ and $\hat{\gamma}_{2,1}$. Also, values of $\hat{\rho}_G$ are larger than values of $\hat{\rho}_p$ in most cases.

3.7 CONCLUSION

In this chapter we have proposed a symmetrized Gini correlation ρ_g and have studied its properties. The relationship between ρ_g and ρ is established when the scatter matrix, Σ, is homogeneous. The affine invariant version ρ_G is also proposed to deal with the case when Σ is heterogeneous. Asymptotic normality of the proposed estimators are established. The influence function reveals that ρ_g is more robust than the Pearson correlation while it is less robust than Kendall's τ correlation. Comparing with the Pearson correlation estimator, the regular Gini correlation estimator and the Kendall's τ estimator of ρ, the proposed estimators balance well between efficiency and robustness and provide an attractive option for measuring correlation. Numerical studies demonstrate that the proposed estimators have satisfactory performance under a variety of situations. In particular, the symmetric Gini estimators are more efficient than the regular Gini estimators. This can be explained by the fact that the multivariate spatial rank used in the symmetrized Gini correlations takes more dependence information than the marginal ranks in the traditional ones.

We comment that the symmetric Gini correlation ρ_g is not limited to elliptical distributions. Theorems 3.2.1, 3.3.1 and 3.4.1 hold for any bivariate distribution with a finite first moment. Under elliptical distributions, the linear correlation parameter ρ is well defined and all four estimators are Fisher consistent. Hence their asymptotic variances are comparable and can be used for evaluating relative asymptotic efficiency among the estimators.

The proposed symmetric Gini correlation has some disadvantages. Although its formulation is natural, the symmetric Gini loses an intuitive interpretation. It is more difficult to compute than the Pearson correlation, especially when X and Y are heterogeneous. In this case, an iterative scheme is required to obtain the affine invariant version of symmetric Gini correlation. When applying the proposed measure, one may consider the trade-off among efficiency, robustness, computation and interpretability.

BIBLIOGRAPHY

[1] Bailey, W. N. (1935). *General Hypergeometric Series.* Cambridge Tracts in Mathematics and Mathematical Physics.

[2] Blitz, R.C. and Brittain, J.A. (1964). An extension of the Lorenz diagram to the correlation of two variables. *Metron* **XXIII** (1-4) 137-143.

[3] Borovskikh, Y.V. (1996). *U-statistics in Banach spaces*, VSP, Utrecht.

[4] Brockwell, P. J. and Davis, R. A.(1987). *Time Series: Theory and Methods.* Springer.

[5] Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press, Princeton, NJ.

[6] Croux, C. and Dehon, C. (2010). The influence function of the Spearman and Kendall correlation measures. *Stat. Methods Appl.* **19** (4) 497-515.

[7] Dang, X., Sang, H. and Weatherall, L. (2015). Gini Covariance Matrix and its Affine Equivariant Version. *Statist. Papers* Accepted.

[8] Devlin, S.J., Gnanadesikan, R. and Kettering, J.R. (1975). Robust estimation and outlier detection with correlation coefficients. *Biometrika* **62** 531-545.

[9] Dittmann, I. and Granger, C. (2002). Properties of nonlinear transformations of fractionally integrated processes. *J. Econometrics* **110** 113-133.

[10] Dürre, A., Vogel, D. and Fried, R. (2015). Spatial sign correlation. *J. Multivariate Anal.* **135** 89-105.

[11] Fang, K.T., Kotz, S. and Hg, K.W. (1987). *Symmetric Multivariate and Related Distributions*, Chapman & Hall, London.

[12] Faÿ, G., Moulines, E., Roueff, F. and Taqqu, M. S. (2009). Estimators of long-memory: Fourier versus wavelets. *J. Econometrics* **151** 159-177.

[13] Gauss, F. (1866). Disquisitiones generales circa sericm infinitam, *Ges. Werke* **3** 123-163 and 207-229.

[14] Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *J. Time Series Anal.* **4** 221-238.

[15] Gini, C. (1914). Reprinted: On the measurement of concentration and variability of characters (2005). *Metron* **LXIII** (1) 3-38.

[16] Giraitis, L. and Surgailis, D. (1985). Central limit theorems and other limit theorems for functionals of Gaussian processes. *Probab. Theory Related Fields* **70** 191-212.

[17] Hampel, F.R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383-393.

[18] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. J. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

[19] Ho, H. and Hsing, T. (1996). On the asymptotic expansion of the empirical processes of long-memory moving averages. *Ann. Statist.* **24** 992-1024.

[20] Ho, H. and Hsing, T. (1997). Limit theorems for functionals of moving averages. *Ann. Probab.* **25** 1636-1669.

[21] Ho, H. and Hsing, T. (2003). A decomposition for generalized U-statistics of long-memory linear processes. *Theory and applications of long-range dependence, Birkhäuser Boston*, 143-155, Boston, MA.

[22] Honda, T. (2000). Nonparametric density estimation for a long-range dependent linear process. *Ann. Inst. Statist. Math.* **52** 599-611.

[23] Hsing, T. (1999). On the asymptotic distributions of partial sums of functionals of infinite-variance moving averages. *Ann. Probab.* **27** 1579-1599.

[24] Hsing, T. (2000). Linear processes, long-range dependence and asymptotic expansions. *Stat. Inference Stoch. Process.* **3** 19-29.

[25] Hsing, T. and Wu, W. B. (2004). On weighted U-statistics for stationary processes. *Ann. Statist.* **32** 1600-1631.

[26] Hurvich, C. M., Dro, R. and Brodsky, J. (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time-series. *J. Time Series Anal.* **19** 19-46.

[27] Kendall, M.G. and Gibbons, J.D. (1990). *Rank Correlation Methods*, 5th edition, Griffin, London.

[28] Kulik, R. (2008). Nonparametric deconvolution problem for dependent sequences. *Electron. J. Stat.* **2** 722-740.

[29] Lindskog, F., Mcneil, A. and Schmock, U. (2003). Kendall's tau for elliptical distributions. In Bol, et al. (Eds), *Credit Risk: Measurement, Evaluation and Management*, 149-156, Springer-Verlag, Heidelberg.

[30] Mari, D.D. and Kotz, S. (2001). *Correlation and Dependence*, Imperial College Press, London.

[31] McElroy, T. and Politis, D. (2007). Self-normalization for heavy-tailed time series with long memory. *Statistica Sinica* **17** 199-220.

[32] Möttönen, J., Oja, H. and Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *Ann. Stat.* **25** 542-552.

[33] Oja, H. (2010). *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks*. Springer

[34] Peligrad, M., Sang, H., Zhong, Y. and Wu, W. B. (2014). Exact moderate and large deviations for linear processes. *Statist. Sinica* **24** 957-969.

[35] Ruppert, D. (2011). *Statistics and Data Analysis for Financial Engineering.* Springer.

[36] Schechtman, E. and Yitzhaki, S. (1987). A measure of association based on Gini's mean difference. *Comm. Statist. Theory Methods* **16** (1) 207-231.

[37] Schechtman, E. and Yitzhaki, S. (1999). On the proper bounds of the Gini correlation. *Econom. Lett.* **63** 133-138.

[38] Schechtman, E. and Yitzhaki, S. (2003). A Family of Correlation Coefficients Based on the Extended Gini Index. *J. Econ. Inequal.* **1** (2) 129-146.

[39] Schechtman, E., Yizhaki, S. and Artsev, Y. (2007). The similarity between mean-variance and mean-Gini: Testing for equality of Gini correlations. *Advances in Investment Analysis and Portfolio Management (AIAPM)*, **3** 103-128.

[40] Seneta, E. (1976). *Regularly Varying Functions.* Lecture Notes in Mathematics **508**, Springer.

[41] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

[42] Shevlyakov G.L. and Smirnov P.O. (2011). Robust Estimation of the Correlation Coefficient: An Attempt of Survey. *Aust. J. Stat.* **40** 147-156.

[43] Taqqu, M. S. (1979). Convergence of integrated processes of arbitrary Hermite rank. *Probab. Theory and Related Fields* **50** 53-83.

[44] Velasco, C. (1999 a). Non-stationary log-periodogram regression. *J. Econometrics* **91** 325-371.

[45] Velasco, C. (1999 b). Gaussian semiparametric estimation of non-stationary time series. *J. Econometrics* **20** 87-127.

[46] Visuri, S., Koivunen, V. and Oja, H. (2000). Sign and rank covariance matrices. *J. Stat. Plan. Infer.* **91** 557-575.

[47] Witting, H. and Müller-Funk, U. (1995). *Mathematische Statistik II*, B.G. Teubner, Stuttgart.

[48] Wu, W. B. and Mielniczuk, J. (2002). Kernel density estimation for linear processes. *Ann. Statist* **30** 1441-1459.

[49] Wu, W. B. (2002). Central limit theorems for functionals of linear processes and their applications. *Statist. Sinica* **12** 635 - 649.

[50] Wu, W. B. (2003). Empirical processes of long-memory sequences. *Bernoulli* **9**(5) 809 - 831.

[51] Wu, W. B. (2006). Unit root testing for functionals of linear processes. *Econometric Theory* **22** 1-14.

[52] Wu, W. B. and Zhao, Z. (2008). Moderate deviations for stationary processes. *Statist. Sinica* **18** 769–782.

[53] Xu, W., Huang, Y.S., Niranjan, M. and Shen, M. (2010). Asymptotic mean and variance of Gini correlation for bivariate normal samples. *IEEE Trans. Signal Process.* **58** (2) 522-534.

[54] Yizhaki, S. (2003). Gini's mean difference: a superior measure of variability for non-normal distribution. *Metron. Int. J. Stat.* **61** 285-316.

VITA

Yongli Sang was born in Jining, China, on July 18, 1986. She received a Bachelor of Science degree in 2009 from Shandong Normal University, a Master of Science degree in 2012 from Central China Normal University, and a Master of Science degree in 2014 from The University of Mississippi. She is currently a Ph.D. candidate in Mathematics at The University of Mississippi.