Technical Report RM 707, Department of Statistics and Probability

Michigan State University

A CLASS OF ESTIMATORS: A UNIFYING TOOL TOWARDS THE ESTIMATION OF GINI INDEX AND ITS VARIANT

Sudheesh K Kattumannil[*,†]

[†] Department of Statistics and Probability, Michigan State University, USA.

[*]Indian Statistical Institute, Chennai, India.

Abstract. The Gini index and its variant are widely used as a measure of income inequality. Finding reliable estimators of these measures and studying its asymptotic properties has been an important area of research in the last two decades. Due to the fragmentation of literature among statistician and economist, several results in this direction have been republished often with a clear lack of reference to previous work. In this paper, we propose a simple unique approach to find the estimators of different income inequality measures. Asymptotic properties of these estimators can be proved in an identical way. The method described here provides an explicit formula for finding the asymptotic variance of the proposed estimators. A consistent estimator of the asymptotic variance can also find by plug-in method. We bring several research problems related to the estimation of Gini index and related concepts into our uniform framework. The asymptotic distribution obtained for Gini covariance has far reaching consequence due to its potential application in non-linear time series analysis.

Keywords: Gini mean differences, Gini covariance, Gini index,Gini regression, Extended Gini index, order statistics, time series.

[†]This work was completed while visiting Prof. Hira L Koul, Department of Statistics and Probability, Michigan State University, during the calender year 2014, under IUSSTF Fellowship. E-mail: skkattu@isichennai.res.in.

1. Introduction

Large number of indices of economic inequality, compatible with various axioms of fairness, has been proposed in the literature. Most of these measures are generalizations of the Gini mean difference, placing smaller or greater weights on various portions of the income distribution. We refer readers to Yitzhaki and Schechtman (2005, 2013), Davidson (2009), Peng (2011), Shelef and Schechtman (2011), Ceriani and Verme (2012), Langel and Tille (2013), Kattumannil and Dewan (2013) and Carcea and Serfling (2014) for the discussion of the area and main references. Among these Ceriani and Verme (2012) has given an overview on the origin and development of Gini index and provided a list of different expressions of the Gini index. Langel and Tille (2013) survey a large part of the literature related to the topic and show that the same results, as well as the same errors, have been republished several times, often with a clear lack of reference to previous work. They also reviewed several linearization techniques for approximating the variance of a non-linear statistic to derive the variance estimator of the Gini index.

Yitzhaki and Schechtman (2013) depicted the development and the understanding of the Gini index and its applications in statistics and economics. This book by Yitzhaki and Schechtman (2013) represents a useful primer on the Gini methodology. The first part of the book (Chapters 2-12) presented the development and inference procedure related to a family of parameters based on Gini mean difference: the Gini index, the Gini covariance, the Gini correlation, the Extended Gini index and the Gini parameters of a regression, among others. The second part (Chapters 12-22) is devoted to discuss the applications of these inequality indices in statistics, finance and economics. In Chapter 23, they discuss several open problems related to regression and time series analysis.

Based on Gini autocovariance function (GACF), Serfling (2010) and Carcea and Serfling (2014) have laid a theoretical foundation for analysing time series with heavy tail innovations. Using GACF they developed a general estimation procedure to find the parameters involved in the linear model where they illustrated their technique for the autoregressive, moving average, and ARMA models. They also illustrated the role of the GACF in analyzing the nonlinear autoregressive Pareto process. In parallel to (GACF), Shelef and Schechtman (2011) defined Gini partial autocovariance function (GPACF) and used the Gini-based methodology for identifying and analyzing time series with non-normal innovations. Since the Gini methodology is a rank based methodology, which takes into account both the variate values and the ranks it has great significance in time series analysis. As it relies only on first order moment assumptions it is a valid tool for analysis time series with stable innovations.

Shelef (2014) developed a new Gini based unit root test. This test is based on the well-known Dickey-Fuller test, where the ordinary least squares regression coefficient is replaced by Gini parameter of regression. Shelef (2014) used the bootstrap technique for finding the critical values of the test as it is difficult to find the exact or/and asymptotic null distribution of the test statistic. This motivate us to propose a general method to find the asymptotic properties of the estimators of Gini index and its variant. In due course we propose a class of estimators where the estimators of different inequality indices can be derived from it. Moreover, making use of the asymptotic theory developed here one can solve the most of the open problems involving Gini index and its variant.

The paper is organized as follows. In Section 2 we propose a class of estimators for finding the estimators of several income inequality measures. It enables us to find the asymptotic distribution of the estimators so obtained in a unified fashion. The method also suggest a tool for finding the asymptotic variance of the proposed estimators. We bring several research problems related to the estimation of Gini index and related indices into our uniform framework. In Section 3, we discuss some open problems that can be carried out as further research in this area.

2. Estimation and Asymptotic of income inequality measures

Mainly two methods are available in literature for finding the estimators of Gini index and its variants; one based on U-statistics and another based on empirical distribution function. In the first case, the Gini index is expressed as an expectation of a function of random variables and then one finds a U-statistic which is an unbiased estimator of the Gini index. In this method studying asymptotic properties of the estimators are simple and straight forward. See Xu (2007) and Kattumannil and Dewan (2013) for a detailed discussion of estimation of Gini index based on U-statistics.

In the second case, the Gini index is expressed as an integral of a quantity involving the underlying distribution function, which is then estimated by replacing the distribution function by the empirical distribution function. Studying the asymptotic properties of these estimators is not simple and requires several algebraic manipulations, see Davidson (2009) to get a flavour of it. Since the empirical distribution function is a consistent and sufficient estimator of the cumulative distribution function, this method has its own relevance. We refer to Peng (2011) for a recent discussion on the inference of Gini index based on empirical distribution function.

As pointed out earlier, finding the reliable estimators of the Gini index has been subject to numerous publications. Hence an attempt is made here to propose (rediscover) a class of estimators that can be used to find the estimators of different inequality indices.

Let (X, Y) be a bivariate random vector with joint distribution function F_{XY}. Also let F_X and F_Y be the respective marginal distribution functions. We assume that the first moment of these random variables is finite. Suppose (X_1, Y_1), ...,(X_n, Y_n) are independent and identically distributed as the bivariate random vector (X, Y). Then the i-th ordered X variate is denoted by $X_{i:n}$ and associate Y variate paired with the $X_{i:n}$, the concomitant of the i-th order statistics by $Y_{[i:n]}$. Under the above formulation, consider the statistic of the form

$$T(F_n) = \frac{1}{n}\sum_{i=1}^{n} J(\frac{i}{n})h(X_{i:n}, Y_{[i:n]}), \tag{1}$$

where J is a bounded smooth function, $h(x, y)$ is a real valued function of (x, y) and F_n is the empirical distribution function of F. Clearly $T(F_n)$ is a natural plug-in estimator of the integral of the form

$$T(F) = \int_0^{\infty}\int_0^{\infty} J(F_X(x))h(x, y)dF_{XY}(x, y). \tag{2}$$

Some of the properties of $T(F_n)$ are first discussed by Yang (1981) in the context of non-parametric estimation of a regression function.

The form of the estimator (1) gives a unique way to find the estimators of different income inequality measures. Consequently, while finding the estimators our task is reduces to rewrite these inequality measures in the form of $T(F)$. The consistency of $T(F_n)$ is proved in the following theorem.

Theorem 1. *If $E|h(x, y)| < \infty$, then as $n \to \infty$, $T(F_n)$ converge to $T(F)$ in probability.*

Proof: Note that $J(.)$ is a smooth function bounded by one. It is well-known that the $F_n(x,y)$ converges uniformly to $F_{XY}(x,y)$. Hence the proof of the theorem is trivial.

The asymptotic distributions of $T(F_n)$ have been obtained by Yang (1981) and Sandstrom (1987). Under quite mild conditions Yang (1981) established the asymptotic normality of $\sqrt{n}(T(F_n)-E(T(F_n)))$ using Hajek's projection lemma. Using a stochastic Gateaux differential, Sandstrom (1987) proved the asymptotic normality of $\sqrt{n}(T(F_n) - T(F))$.

Next we introduce some notation. Let

$$\alpha_h(x) = E(h(X,Y)|X = x) \tag{3}$$

and

$$\tau_h^2(x) = V(h(X,Y)|X = x). \tag{4}$$

Also let

$$\sigma^2 = \sigma_{11}^2 + \sigma_{22}^2, \tag{5}$$

where

$$\begin{aligned} \sigma_{11}^2 &= \int_0^\infty \int_0^\infty \left[min\{F_X(x), F_X(z)\} - F_X(X)F_X(z)\right] \\ &\qquad J(F_X(x))J(F_X(z))d\alpha_h(x)d\alpha_h(z) \end{aligned} \tag{6}$$

and

$$\sigma_{22}^2 = \int_0^\infty J^2(F_X(x))\tau_h^2(x)dF_X(x). \tag{7}$$

Theorem 1. *Assume $\alpha_h(x)$ is right continuous and that J is bounded in $[0,1]$ and differentiable. Also assume that $\alpha_h(x)$ and $\tau_h^2(x)$ are finite. Suppose, σ^2 is as defined in (5), then as $n \to \infty$, the distribution of $\sqrt{n}((T(F_n) - T(F))/\sigma$ converges to standard normal distribution.*

Proof: The proof is immediate from Theorem 1 of Sandstrom (1987) by taking uniform weight in weighted empirical distribution function proposed by Koul (1970).

We will use this theorem to derive the asymptotic distribution of the estimators of different inequality measures. This shows the advantages of our task over the work done by Davidson (2009), Peng (2011), among others.

2.1. **Estimation of Gini index.** Before presenting the methodology, we briefly review the concepts of Gini mean difference and Gini index which will be the main focus of the present study.

Definition 1. *The Gini mean difference is defined as*

$$GMD = E|X_1 - X_2|, \tag{8}$$

where X_1 *and* X_2 *are the independent and identical copies of* X.

The Gini mean difference was first introduced by Corrado Gini in 1912 as an alternative measure of variability. Gini mean difference and different parameters which are derived from it have been widely using in the area of income distribution. Note that

$$|X_1 - X_2| = max(X_1, X_2) - min(X_1, X_2).$$

Hence Gini mean difference can be expressed as

$$GMD = E(max(X_1, X_2) - min(X_1, X_2)). \tag{9}$$

Accordingly, Gini mean difference can be interpreted as the expected difference between the maximum and the minimum of two random draws from F_X. The Gini index is defined in connection with the Gini mean difference as below.

Definition 2. *The Gini index is defined as*

$$GI = \frac{E|X_1 - X_2|}{2\mu}, \tag{10}$$

where $\mu = \int x dF_X(x)$.

Clearly, the Gini index is the scaled version of half of the Gini mean difference. The estimation problem of the Gini index mainly concentrated on finding plug-in estimators of the Gini index with reliable standard errors. This is achieved by expressing the Gini index in different forms involving cumulative distribution function (Yitzhaki, 1998) and then finds a plug-in estimator of it.

Even though the expression (9) is useful to connect the Gini index to the Gini mean difference, Gini index is usually defined either through the Lorenz curve or through covariance identity involving cumulative distribution function. The Lorenz curve denoted by $L(.)$ is defined by the equation

$$L(F(x)) = \frac{1}{\mu}\int_0^x y dF(y). \tag{11}$$

If X represents the annual income, $L(p)(p = F_X(x))$ is the proportion of the total income that accrues to the individuals having $100p\%$ lowest income. In terms of Loreze curve the Gini index is defined as

$$\begin{aligned} GI &= 2\int_0^1 (z - L(z)))dz \\ &= 1 - 2\int_0^1 L(z)dz. \end{aligned} \tag{12}$$

In fact, from equation (12), it can be seen that the Gini index is the twice the area between the Lorenz curve and egalitarian line. This interpretation makes the Gini index as the most popular measure of income inequality.

By simple algebra, we can show that the expression given in (12) is same as

$$GI = \frac{2}{\mu}\int_0^\infty yF(y)dF(y) - 1.$$

One can also write the above expression as follows:

$$G = \frac{2}{\mu}Cov(X, F_X(X)). \tag{13}$$

That is, for given F_X, the Gini index is simply the covariance between X and F_X. This expression has great significance while studying the estimation problem related to the regression parameters as one needs to decompose the population Gini index.

Next we derive the estimators of Gini mean difference and Gini index from (1).

By simple algebra we can rewrite the equation (9) as

$$\begin{aligned} GMD &= \int_0^\infty 2xF_X(x)dF_X(x) - \int_0^\infty 2x(1-F_X(x))dF_X(x) \\ &= 4\int_0^\infty xF_X(x)dF_X(x) - 2\int_0^\infty xdF_X(x) \\ &= 2\int_0^\infty (2F_X(x)-1)xdF_X(x) \\ &= 2\int_0^\infty \int_0^\infty (2F_X(x)-1)xdF_{XY}(x,y). \end{aligned} \tag{14}$$

By taking $J = (2F_X(x) - 1)$, $h(x, y) = 2x$, the equation (14) coincides with (2). Hence from (1), we obtain the following estimator of the Gini mean difference

$$\widehat{GMD} = \frac{2}{n^2}\sum_{i=1}^{n}(2i-n)X_{i:n}. \tag{15}$$

Note that $\frac{1}{n}\sum_{i=1}^{n} X_i$ is an unbiased estimator of μ. Hence an estimator of Gini index is given by

$$\widehat{GI} = \frac{\sum_{i=1}^{n}(2i-n)X_{i:n}}{n\sum_{i=1}^{n} X_i}. \tag{16}$$

Next we obtain the asymptotic distribution of the above estimators using Theorem 2.

Corollary 1. *As $n \to \infty$, the distribution of $\sqrt{n}(\widehat{GMD} - GMD)/\sigma_1$ converges to standard normal distribution, where σ_1 is given by*

$$\begin{aligned}\sigma_1^2 &= 4\int_0^\infty\int_0^\infty \left[min\{F_X(x), F_X(z)\} - F_X(x)F_X(z)\right] \\ &\qquad (2F_X(x)-1)(2F_X(z)-1)dxdz.\end{aligned}$$

Proof: The asymptotic normality follows from Theorem 2. Note that $\alpha_h(x) = 2x$, $\tau_h^2(x) = 0$ and $J = (2F_X(x) - 1)$, hence from (5) we have the variance expression given as in Corollary 1.

Corollary 2. *As $n \to \infty$, the distribution of $\sqrt{n}(\widehat{GI} - GI)/\sigma_2$ converges to standard normal, where σ_2 is given by*

$$\sigma_2^2 = \frac{\sigma_1^2}{4\mu^2}.$$

Proof: Clearly $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a consistent estimator of μ. Hence the proof follows by applying the Slutsky's theorem in the Corollary 1.

We realize that several estimators of the Gini mean difference and the Gini index available in literature can be obtained from (1) and we illustrated it for Davidson (2009) estimator. Davidson (2009) obtained a plug-in estimator of the Gini index by expressing it in the form

$$GI = \frac{2}{\mu}\int_0^\infty xF_X(x)dF_X(x) - 1,$$

which is can be rewritten as

$$GI = \frac{1}{\mu}\int_0^\infty \int_0^\infty x(2F_X(x) - 1)dF_{XY}(x, y).$$

Note that the above equation is of the form $T(F)$. Hence the asymptotic normality of Davidson (2009) estimator follows from Theorem 2. In fact it reduces to the mathematical complexity involving in finding the asymptotic variance of Davidson estimator. This shows the advantages of our task over the work done by Davidson (2009).

2.2. **Estimation of Extended Gini index.** The extended Gini index is a family of inequality measures that depends on one parameter, the extended Gini parameter. This measure is widely used in finance. By choosing a member of the family, the investigator can perform a sensitivity analysis and can evaluate the robustness of the result. For more details see Chapter 6 of Yitzhaki and Schechtman (2013). Next we give the definition of extended Gini index and relative extended Gini index.

Definition 3. *The extended Gini index is defined as (Yitzhaki, 1983)*

$$EG(v) = -vCov(X, \bar{F}_X^{v-1}(X)), \quad v > 0 \quad v \neq 1, \tag{17}$$

where $\bar{F}_X(x) = 1 - F_X(x)$.

Definition 4. *The relative extended Gini index is defined as*

$$REG(v) = \frac{-v}{\mu}Cov(X, \bar{F}_X^{v-1}(X)), \quad v > 0 \quad v \neq 1. \tag{18}$$

Similar to the Gini index, the value of the relative extended Gini index for non-negative distributions lies between zero and one where only difference is in the weighting scheme that applied to the vertical distance between the egalitarian line and the Lorenz curve. In view of the equation (13), it can be easily verified that the relative extended Gini index with $v = 2$ reduces to the Gini index.

Next, to find the estimator of the extended Gini index we express it in the form of (2). The expression (17) can be rewritten as

$$EG(v) = \int_0^\infty \int_0^\infty x(1 - v\bar{F}^{v-1}(x))dF_{XY}(x,y).$$

For $J = (1 - v\bar{F}^{v-1})$, $h(x,y) = x$ the above expression coincides with (2). Then from (1) we obtain the following estimator of the extended Gini index

$$\widehat{EG}(v) = \frac{\sum_{i=1}^{n}(n^{v-1} - v(n-i)^{v-1})X_{i:n}}{n^v}. \tag{19}$$

The estimator of the relative extended Gini index is given by

$$\widehat{EG}(v) = \frac{\sum_{i=1}^{n}(n^{v-1} - v(n-i)^{v-1})X_{i:n}}{n^{v-1}\sum_{i=1}^{n} X_i}. \tag{20}$$

Since the above estimators are obtained from (1), the asymptotic normality of these estimators can be established using the Theorem 2.

Corollary 3. *As $n \to \infty$, the distribution of $\sqrt{n}(\widehat{EG}(v) - EG(v))/\sigma_3$ converges to standard normal, where σ_3 is given by*

$$\begin{aligned}\sigma_3^2 &= \int_0^\infty \int_0^\infty \big[min\{F_X(x), F_X(z)\} - F_X(x)F_X(z)\big] \\ &\qquad (1 - v\bar{F}_X^{(v-1)}(x))(1 - v\bar{F}_X^{(v-1)}(z))dxdz.\end{aligned}$$

Proof: The proof is similar to that of the Corollary 1.

Since $\bar{X} = \frac{1}{n}\sum_{i=1}^{n}$ is a consistent estimator of μ we have the following result for the asymptotic normality of the relative extended Gini index.

Corollary 4. *As $n \to \infty$, the distribution of $\sqrt{n}(\widehat{REG}(v) - REG(v))/\sigma_4$ converges to standard normal, where σ_4 is given by*

$$\sigma_4^2 \quad = \quad \frac{\sigma_3^2}{\mu^2}.$$

Next we show that some of the existing result related to the estimation of extended Gini index can be brought in to our uniform framework.

Zitikis and Gastwirth (2002) obtained the asymptotic distribution of the relative extended Gini index by expressing it as

$$REG(v) = 1 - \frac{v}{\mu}\int_0^1 F_X^{-1}(1-t)^{v-1}dt.$$

By substituting $t = \bar{F}(x)$ and rearranging terms we obtain

$$REG(v) = \frac{1}{\mu}\int_0^\infty x(1 - v\bar{F}^{v-1}(x))dF_X(x).$$

Equivalently

$$REG(v) = \frac{1}{\mu}\int_0^\infty\int_0^\infty x(1 - v\bar{F}^{v-1}(x))dF_{XY}(x, y),$$

which is of the form (2). Hence their asymptotic result can be obtained from Theorem 2.

Remark 1. *Using the empirical quantile process approach Barrett and Donald (2009) developed a general large sample asymptotic theory for various indices of inequality, including relative extended Gini index. As their result of relative extended Gini index is based on the expression provided by Zitikis and Gastwirth (2002), their work on relative extended Gini index also comes under our uniform frame work.*

2.3. **Estimation of Gini covariance.** The representation of Gini mean difference in terms of covariance between X and $F_X(X)$ leads to number of parameters similar to covariance, correlation and regression coefficient.

This parameters are very much useful in non-linear regression and time series analysis. The following definitions are essential for our discussion.

Definition 5. *The Gini covariance between Y and X is defined as*

$$C(Y,X) = 4Cov(Y, F_X(X)). \tag{21}$$

Definition 6. *The Gini correlation between Y and X is defined as*

$$\rho_g(Y,X) = \frac{Cov(Y, F_X(X))}{Cov(Y, F_Y(Y))}. \tag{22}$$

Definition 7. *The Gini regression parameter of Y on X is defined as*

$$\beta_g(Y,X) = \frac{Cov(Y, F_X(X))}{Cov(X, F_X(X))}. \tag{23}$$

Similarly, one can define the Gini covariance, Gini correlation between X and Y and regression parameter of X on Y and we denote it as $C(X,Y)$, $\rho_g(X,Y)$ and $\beta_g(X,Y)$, respectively.

Remark 2. *Note that $\rho_g(X,Y) \neq \rho_g(Y,X)$ in general. If the distribution of (X,Y) is exchangeable up to a linear transformation then $\rho_g(X,Y) = \rho_g(Y,X)$. Moreover, if the distribution of (X,Y) is bivariate normal, then $\rho_g(X,Y) = \rho_g(Y,X) = \rho$, where ρ is the Pearson's correlation coefficient. For more details about Gini covariance and Gini correlation we refer to Schechtman and Yitzhaki (1987). Under bivariate normal assumption of the random vector (X,Y), the Gini regression parameter of Y on X reduces to the ordinary least square regression coefficient of Y on X.*

Next we obtain the estimators of these parameters from (1). Rewrite the expression given in (21) as

$$\begin{aligned} C(Y,X) &= 4Cov(Y,F_X(X)) \\ &= 4\int_0^\infty \int_0^\infty yF_X(x)dF_{XY}(x,y) - \frac{4}{2}\int_0^\infty ydF_Y(y) \\ &= 2\int_0^\infty \int_0^\infty y(2F_X(x)-1)dF_{XY}(x,y). \qquad (24) \end{aligned}$$

By taking $J = (2F_X(x) - 1)$ and $h(x,y) = 2y$, the equation (24) takes the form (2). Hence an estimator of $C(Y,X)$ is given by

$$\widehat{C}(Y,X) = \frac{2}{n^2}\sum_{i=1}^{n}(2i-n)Y_{[i:n]}. \qquad (25)$$

From equations (15) and (25) we have the estimator of $\rho_g(Y,X)$ given by

$$\widehat{\rho}_g(Y,X) = \frac{\sum_{i=1}^{n}(2i-n)Y_{[i:n]}}{\sum_{i=1}^{n}(2i-n)Y_{i:n}}, \qquad (26)$$

where $Y_{i:n}$ is the i-th order statistics from the sample $Y_1,...,Y_n$. Similarly an estimator of the $\beta_g(Y,X)$ is given by

$$\widehat{\beta}_g(Y,X) = \frac{\sum_{i=1}^{n}(2i-n)Y_{[i:n]}}{\sum_{i=1}^{n}(2i-n)X_{i:n}}. \qquad (27)$$

Using the above estimators and the asymptotic theory developed here one can solve the most of the open problems related to hypothesis testing involving Gini based parameters. For example, the estimator of $\beta_g(Y,X)$ can be used in the Gini based unit root test introduced by Shelef (2014). The critical values of the respective test can be obtained using the following asymptotic result.

Corollary 5. *As* $n \to \infty$*, the distribution of* $\sqrt{n}(\widehat{C}(Y,X) - C(Y,X))/\sigma_5$ *converges to standard normal, where* σ_5 *is given by*

$$\begin{aligned}\sigma_5^2 &= \int_0^\infty \int_0^\infty \big[min\{F_X(x), F_X(z)\} - F_X(x), F_X(z)\big] \\ &\qquad (2F_X(x)-1)(2F_X(z)-1)d\alpha_h(x)d\alpha_h(z) \\ &\quad + \int_0^\infty (2F_X(x)-1)^2 dF_X(x),\end{aligned}$$

where $\alpha_h(x) = 2E(Y|X=x)$ *and* $\tau_h^2(x) = 4Var(Y|X=x)$*.*

Proof: For $J = (2F_X(x)-1)$ and $h(x,y) = 2y$, we noticed that $\widehat{C}(Y,X)$ is of the form (1). Hence the asymptotic normality follows from Theorem 2. Moreover we have $\alpha_h(x) = 2E(Y|X=x)$ and $\tau_h^2(x) = 4Var(Y|X=x)$. Hence the variance expression stated in the corollary can be easily obtained from (5).

Corollary 6. *As* $n \to \infty$*, the distribution of* $\sqrt{n}(\widehat{\rho_g}(Y,X) - \rho_g(Y,X))/\sigma_6$ *converges to standard normal, where* σ_6 *is given by*

$$\sigma_6^2 = \frac{\sigma_5^2}{16Cov^2(Y, F_Y(Y))}.$$

Proof: By Theorem 1, $\frac{1}{n}\sum_{i=1}^n (2i-n)Y_{i:n}$ is a consistent estimator of $Cov(Y, F_Y(Y))$. Hence using Slutsky's theorem the results follows from Corollary 5.

In similar line of the proof of Corollary 6 we can prove the following result for $\widehat{\beta}_g(Y,X)$.

Corollary 7. *As* $n \to \infty$*, the distribution of* $\sqrt{n}(\widehat{\beta_g}(Y,X) - \beta_g(Y,X))/\sigma_7$ *converges to standard normal, where* σ_7 *is given by*

$$\sigma_7^2 = \frac{\sigma_5^2}{16Cov^2(X, F_X(X))}.$$

Remark 3. *Schechtman and Yitzhaki (2003) defined family of correlation coefficients based on the extended Gini index. The estimators of these family can be derived from (1). Moreover, in similar line of Corollary 6 one can find the asymptotic distributions of these family of correlation coefficients using Theorem 2.*

3. Conclusion

Even though, Gini index is the most widely used indicator of income inequality in a population, it is not a trivial measure to handle. Hence a large number of techniques for finding the reliable estimators and computing an asymptotically valid standard error have been proposed, of varying degrees of complexity. In fact several of them are the reproduction of the existing estimator with a clear lack of references to earlier work as pointed out by Langel and Tille (2013). This motivates us to propose a unique way for finding the estimators of different income inequality measures.

In this paper, we proposed a class of estimators that can be used to obtain the natural plug-in estimators of different inequality indices including Gini index. The proposed method permits us to find the asymptotic distribution of the estimators so obtained in a unified fashion. This method also gave an explicit expression for the variance of the estimators. We derived several existing results on the estimation of Gini mean difference and related parameters as special cases of our general result.

The asymptotic distribution obtained for Gini correlation has great significance due to its potential application in infinite variance time series analysis. Recently, using Gini correlation, Carcea and Serfling (2014) gave a theoretical foundation for analysing time series with heavy tail innovations. Using the asymptotic theory developed here one can propose different test procedure which involves Gini index and its variants. For example, using the asymptotic distribution of Gini correlation, one can develop a test for serial dependence and stationarity in the non-linear non-Gaussian time series setup. We can also develop a Gini based unit root test using the estimator of Gini regression parameter obtained in this paper and the ongoing work related to this will be reported else-where.

Acknowledgment: The author Sudheesh K K is thankful to Indo-US Science and Technology Forum and Department of Science and Technology, Government of India for the financial assistant provided to carry out this research work at Michigan State University under Prof. Hira L Koul.

References

Barretta, G.F. and Donald, S.G. (2009). Statistical inference with generalized Gini indices of inequality, poverty, and welfare, *Journal of Business & Economic Statistics*, 27, 1-17.

Carcea, M. and Serfling, R. (2014). A Gini autocovariance function for heavy tailed time series modeling, preprint, University of Texas at Dallas.

Ceriani, L. and Verme, P. (2012). The origins of the Gini index: extracts from variabilite mutabilit (1912) by Corrado Gini. *Journal Economic Inequality*, 10, 421-443.

Davidson, R. (2009). Reliable inference for the Gini index. *Journal of Econometrics*, 150, 30-40.

Kattumannil, S.K. and Dewan, I. (2013). A Simple non-parmetric estimator of Gini index, *preprint, Indian Statistical Institute,* New Delhi.

Koul, H.L. (1970). Some convergence theorem for Rank and weighted empirical cumulatives, *Annals of mathematical Statistics*, 41, 1768-1773.

Langel, M. and Tille, Y. (2013). Variance estimation of the Gini index: revisiting a result several time published, *Journal of the Royal Statistical Society-Series A*, 176, 521-540.

Peng, L. (2011). Empirical likelihood methods for the Gini index. *Australian and New Zealand Journal of Statistics*, 53, 131-139.

Sandstrom, A. (1987). Asymptotic normality of linear functions of concomitants of order statistics, *Metrika*, 34, 129-142.

Schechtman, E. and Yitzhaki, S. (1987). A measure of association based on Gini mean difference. *Communications in Statistics, Theory and Methods*, 16, 207-231.

Schechtman, E. and Yitzhaki, S. (2003). A family of correlation coefficients based on the extended Gini index. *Journal of Economic Inequality,* 1, 129-146.

Serfling, R. (2010). Fitting autoregressive models via Yule-Walker equations allowing heavy tail innovations. Preprint available at www.utdallas.edu/ serfling.

Shelef, A. (2014). A Gini based unit root test. *Computational Statistics and Data Analysis*, in press.

Shelef, A. and Schechtman, E. (2011). A Gini-based methodology for identifying and analyzing time series with non-normal innovations. Preprint.

Xu, K. (2007). U-statistics and their asymptotic results for some inequality and poverty measures. *Econometric Reviews*, 26, 567-577.

Yang, S. S. (1981). Linear functions of concomitants of order statistics, with application to nonparametric estimation of a regression function. *Journal of the American Statistical Association*, 76,658-662.

Yitzhaki, S. (1983). On an extension of the Gini inequality index, *International Economic Review*, 24, 617-628.

Yitzhaki, S. (1998). More than a dozen alternative ways of spelling Gini. *Research on Economic Inequality*, 8, 1330.

Yitzhaki, S. Schechtman, E. (2005). The properties of the extended Gini measures of variability and inequality, *Metron: International Journal of Statistics*, 63, 401-433.

Yitzhaki, S. and Schechtman, E. (2013). *The Gini Methodology: A Primer on a Statistical Methodology*, Springer.

Zitikis, R. and Gastwirth, J.L. (2002). The Asymptotic Distribution of the S-Gini index, *Australian and New Zealand Journal of Statistics*, 44, 439-446.