On the Asymptotic Distribution of (Generalized) Lorenz Transvariation Measures

Teng Wah Leo*†
St. Francis Xavier University

20 June, 2017

*Department of Economics, St. Francis Xavier University. Email Address: tleo@stfx.ca

†The author would like to thank Gordon Anderson for his many helpful comments and suggestions in improving this paper. Nonetheless, all errors remain the author's.

Abstract

A common problem associated with evaluating *dominance* relationships between distribution functions and their moments is the lack of resolution regarding the direction of dominance as a result of the functions crossing, prevalent in empirical applications. This paper proposes a method of examining the difference between (Generalized) Lorenz curves over the entire support of the variables, an idea first proposed by Anderson and Leo (2017a) and formalized by Anderson et al. (2017) for the case of stochastic dominance. The method provides a way of ordering all the (Generalized) Lorenz curves under consideration. The paper also provides the exact limit distribution of these associated measures, which in consequence of the results due to Politis and Romano (1994), permits inference via subsampling, in lieu of the crossing of empirical (Generalized) Lorenz curves. We show that due to the relationship between the Lorenz curve and the Gini coefficient, the same can be said of the latter. An example is provided to demonstrate its application.

1 Introduction

There have been significant advances made in the use of stochastic dominance techniques with the increased use in both the fields of finance and welfare studies. Nonetheless, the ability of the technique to achieve a complete ordering of distributions under comparison continues to elude us.[1] There has been however significant progress made recently by Anderson and Leo (2017a). The principal idea reframes the problem of ordering, which had been to search for the order of dominance that yields a definitive ordering of the comparisons, to one where the researcher/decision maker chooses the order at which to examine the order amongst the distributions in question. It uses the entire support, and measures the difference in area of the distributions against a synthetic best distribution generated from the hull (lower envelope) created by all the distributions. The idea was formalized, and asymptotic distribution found in Anderson et al. (2017).

Despite the similarities and relationship between stochastic dominance and Lorenz dominance (Atkinson (1970) showed that second order stochastic dominance between two distributions for mean normalized variables is equivalent to Lorenz dominance, while Shorrocks (1983) showed the same relationship with respect to Generalized Lorenz dominance), there has been little work attempting to solve the issue of establishing a complete ordering in applications of Lorenz dominance. Some recent work since Beach and Davidson (1983) established the asymptotic distribution of the Lorenz ordinates are Seth and Yalonetzky (2016), and Zheng (2016), which are closely related to the current work here, and in Anderson and Leo (2017b).

Zheng (2016) extends the work on *Almost Stochastic Dominance* by Leshno and Levy (2002) to Lorenz Dominance by showing, similarly, that by reducing the weight on lower realization outcomes, a Lorenz dominance relationship may be established, particularly when Lorenz curves exhibit crossings. Nonetheless, the proposed solution is not fullproof since frequently the optimal and/or worst states are not separable due to their proximity to each other. In turn, Seth and Yalonetzky (2016) in seeking to establish whether a set of bounded qualitative measures achieves convergence or divergence, establishes the relationship between the induced movements of the measures, and changes in the *Absolute (Generalized) Lorenz ordinates*. Both papers are similar to the current paper here in that they seek to establish the Lorenz dominance relationship, and order the

[1] Indeed, Atkinson (1970) viewed the partial ordering achieved with Lorenz dominance with skepticism in lieu of the frequency with which Lorenz curves crossed.

different distributions under consideration, developing alternate ways to circumvent the lack of resolution when Lorenz curves cross. Nonetheless, the proposed solutions are not fullproof. The method proposed here, and in Anderson and Leo (2017b), and Anderson et al. (2017) can be extended simply to those methods as well.

The primary contribution of this paper is thus the derivation of the limit distribution for the measures proposed by Anderson and Leo (2017b), and to provide a subsampling technique, in lieu of the contact set due to Lorenz curves crossing, thereby allowing for consistent statistical inference. To consolidate the results, a simple example using the Current Population Survey drawn from the Integrated Public Use Microdata Series (IPUMS) is used to examine the evolution of income inequality between 2001 to 2016.

In the following section, we will first establish the definition of the Lorenz Transvariation, extending the idea of Gini Transvariation first proposed by Gini (1916) and Gini (1959) to comparisons between Lorenz curves, then derive a measure of inequality using these ideas, and establish their limit distributions. The relationship with the Gini coefficient is also drawn. This is followed by a discussion of the subsampling procedure in section 3, and section 4 provides a simple example demonstrating the use of the measure, and method of inference. This is finally followed by the conclusion.

2 Lorenz Transvariation & Related Measures

2.1 Lorenz Transvariation

Let there be M income distributions, F_m, $m = \{1, 2, \ldots, M\}$, then for a vector of abscissae $\mathbf{p} = [p_1, p_2, \ldots, p_K, p_{K+1} = 1]'$, and their corresponding quantiles $\mathcal{Y} = [Y_1, Y_2, \ldots, Y_K, Y_{K+1}]'$, where $Y_{K+1} = Y(p_{K+1} = 1) = Y_{\max}$, the Lorenz curve is just,

$$\mathcal{L}(p) = \int_0^p \frac{Y(t)}{\mu} dF(t) \tag{1}$$

$$\Rightarrow \widehat{\mathcal{L}}(p) = \frac{\sum_{j=1}^{N} Y_{(j)} \mathbb{1}\left(Y_{(j)} \leq Y_p\right)}{\sum_{j=1}^{N} Y_{(j)}} \tag{2}$$

while the Generalized Lorenz is in turn,

$$\begin{aligned}\mathcal{G}(p) &= \mu\,\mathcal{L}(p) \\ &= \int_0^p Y(t)dF(t) \quad (3)\\ \Rightarrow \widehat{\mathcal{G}}(p) &= N^{-1}\sum_{j=1}^{N} Y_{(j)}\mathbb{1}\left(Y_{(j)} \leq Y_p\right) \quad (4)\end{aligned}$$

where the subscript denoting the various distribution under consideration has been omitted here.

It will be easier and clearer to develop the limiting distributional characteristics of the *Generalized Lorenz Transvariation* first, and that for the Lorenz follows simply. In addition, for the rest of the discussion here, we make the following assumption,

Assumption 1 *All observations from each of the M distributions are i.i.d.*

If assumption 1 holds, by Theorem 2.1 of Politis and Romano (1994) the subsampling distribution converges to the true distribution if the measures are convergent. The result then validates the use of subsampling in performing inference on the measures, in lieu of the intersections of Lorenz curves under consideration. The method will be elaborated on after the development of the asymptotic distribution of the Lorenz based measures below.

We can define the *Generalized Lorenz Transvariation* between two income distributions, A and B, as follows.

Definition 1 *The Generalized Lorenz Transvariation, $\mathcal{GT}$, is,*

$$\mathcal{GT} := \int_0^1 \left|\mathcal{G}_A(p) - \mathcal{G}_B(p)\right| dp \quad (5)$$

$\mathcal{GT}$ can be expressed in turn as,

$$\begin{aligned}\mathcal{GT} &= \int_0^1 \left[\max\{\mathcal{G}_A(p), \mathcal{G}_B(p)\} - \min\{\mathcal{G}_A(p), \mathcal{G}_B(p)\}\right] dp \\ &= \int_0^1 \max\left\{\mathcal{G}_A(p) - \mathcal{G}_B(p), \mathcal{G}_B(p) - \mathcal{G}_A(p)\right\} dp\end{aligned}$$

In other words, $\mathcal{GT}$ calculates the area between two Generalized Lorenz curves, or more intuitively it examines how different two Generalized Lorenz curves are, noting the difference relative to Lorenz Dominance which essentially examines the distance. Nonetheless, this is still a pairwise comparison that would in turn be cumbersome for a comparison set composed of large number of distributions. More succinctly, since there is no common basis of comparison, pairwise comparisons when there are large number of distributions under consideration will unlikely yield a complete order, not to mention the large number of permutations of comparisons that need to be exhausted.

To examine the asymptotic characteristics of the estimator among the M Generalized Lorenz curves under consideration, define

$$\begin{aligned}\mathcal{P}_k &= \{p \in \mathcal{P} : \mathcal{G}_k(p) > \mathcal{G}_{k'}(p) \quad, \forall k' \neq k\} \\ \mathcal{P}_{k=k'} &= \{p \in \mathcal{P} : \mathcal{G}_k(p) = \mathcal{G}_{k'}(p) \quad, \forall k' \neq k\}\end{aligned}$$

where $m, m' = \{1, \ldots, M\} = \mathcal{Q}$. We will assume that $\mathcal{P}_{k=k'}$ has Lebesgue measure of zero, so that there are no range of the ordinates where the Lorenz curves under consideration are the same. Further, we will assume for simplicity here that the same sample size is drawn from each i.i.d. distribution, N.

In addition, define,

$$\begin{aligned}\widehat{\mathcal{G}}_m(p) &= N^{-1}\sum_{j=1}^{N} Y_{(j)}^m \mathbb{1}\left(Y_{(j)}^m \leq Y(p)\right) \\ \widehat{\mathcal{G}}_m(p_k \wedge p_{k'}) &= N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N} Y_{(i)}^m Y_{(j)}^m \mathbb{1}\left(Y_{(i)}^m \leq Y(p_k \wedge p_{k'}), Y_{(j)}^m \leq Y(p_k \wedge p_{k'})\right) \\ \widehat{\mathcal{L}}_m(p) &= \frac{N^{-1}\sum_{j=1}^{N} Y_{(j)}^m \mathbb{1}\left(Y_{(j)}^m \leq Y(p)\right)}{\widehat{\mu}_m} \\ \widehat{\mu}_m &= N^{-1}\sum_{j=1}^{N} Y_{(j)}^m \\ S_{m,m',j}^q &= \int_{\mathcal{P}_q}\left[Y_{(j)}^m \mathbb{1}\left(Y_{(j)}^m \leq Y(p)\right) - Y_{(j)}^{m'}\mathbb{1}\left(Y_{(j)}^{m'} \leq Y(p)\right)\right] dp \\ K_{m,q,j} &= \sum_{q=1}^{M}\int_{\mathcal{P}_q}\left[Y_{(j)}^q \mathbb{1}\left(Y_{(j)}^q \leq Y(p)\right) - Y_{(j)}^m \mathbb{1}\left(Y_{(j)}^m \leq Y(p)\right)\right] dp = \sum_{q=1}^{M} S_{q,m,j}^q\end{aligned}$$

The following result extends the work by Anderson et al. (2017) and establishes the asymptotic distribution of the Generalized Lorenz Transvariation.

Theorem 1 *For $m \neq m' \in \mathcal{Q}$, the Generalized Lorenz Transvariation , $\widehat{\mathcal{GT}}$, is asymp-*

totically normal. That is, $\sqrt{N}\left(\widehat{\mathcal{GT}}-\mathcal{GT}\right)\xrightarrow{D}\mathcal{N}\left(0,\sigma^2_{\mathcal{GT}}\right)$*, where,*

$$\begin{aligned}
\sigma^2_{\mathcal{GT}} &= \boldsymbol{Var}\left(S^m_{m,m'}\right)+\boldsymbol{Var}\left(S^{m'}_{m,m'}\right)-2\,\boldsymbol{Cov}\left(S^m_{m,m'},S^{m'}_{m,m'}\right) &(6)\\
&= \sigma^2_m+\sigma^2_{m'}-2\sigma_{m,m'}\\
\sigma^2_m &= \int_{\mathcal{P}_m}(\mathcal{G}_m(p)-\mathcal{G}_{m'}(p))^2dp-\left(\int_{\mathcal{P}_m}(\mathcal{G}_m(p)-\mathcal{G}_{m'}(p))dp\right)^2 &(7)\\
\sigma_{m,m'} &= \int_{\mathcal{P}_m}\int_{\mathcal{P}_{m'}}\left[\begin{array}{cc} & \mathcal{G}_m(p\wedge p')-\mathcal{G}_m(p)\,\mathcal{G}_{m'}(p')\\ - & \mathcal{G}_m(p')\,\mathcal{G}_{m'}(p)+\mathcal{G}_{m'}(p\wedge p')\end{array}\right]dpdp' &(8)\\
&\quad - \left[\int_{\mathcal{P}_m}\left[\mathcal{G}_m(p)-\mathcal{G}_{m'}(p)\right]dp\right]\left[\int_{\mathcal{P}_{m'}}\left[\mathcal{G}_{m'}(p')-\mathcal{G}_m(p')\right]dp'\right]
\end{aligned}$$

Proof of Theorem 1

$$\begin{aligned}
\sqrt{N}\left(\widehat{\mathcal{GT}}-\mathcal{GT}\right) &= \sqrt{N}\left\{\begin{array}{cc} & \int_{\mathcal{P}}\max\left[\widehat{\mathcal{G}}_m(p)-\widehat{\mathcal{G}}_{m'}(p),\widehat{\mathcal{G}}_{m'}(p)-\widehat{\mathcal{G}}_m(p)\right]dp\\ - & \int_{\mathcal{P}}\max[\mathcal{G}_m(p)-\mathcal{G}_{m'}(p),\mathcal{G}_{m'}(p)-\mathcal{G}_m(p)]dp\end{array}\right\}\\
&= \sqrt{N}\left\{\begin{array}{cc} & \int_{\mathcal{P}_m}\left[\max\left\{\widehat{\mathcal{G}}_m(p)-\widehat{\mathcal{G}}_{m'}(p),\widehat{\mathcal{G}}_{m'}(p)-\widehat{\mathcal{G}}_m(p)\right\}-(\mathcal{G}_m(p)-\mathcal{G}_{m'}(p))\right]dp\\ + & \int_{\mathcal{P}_{m'}}\left[\max\left\{\widehat{\mathcal{G}}_m(p)-\widehat{\mathcal{G}}_{m'}(p),\widehat{\mathcal{G}}_{m'}(p)-\widehat{\mathcal{G}}_m(p)\right\}-(\mathcal{G}_{m'}(p)-\mathcal{G}_m(p))\right]dp\end{array}\right.\\
&= \sqrt{N}\left\{\begin{array}{cc} & \int_{\mathcal{P}_m}\left[\left(\widehat{\mathcal{G}}_m(p)-\widehat{\mathcal{G}}_{m'}(p)\right)-(\mathcal{G}_m(p)-\mathcal{G}_{m'}(p))\right]dp\\ - & \int_{\mathcal{P}_{m'}}\left[\left(\widehat{\mathcal{G}}_m(p)-\widehat{\mathcal{G}}_{m'}(p)\right)-(\mathcal{G}_m(p)-\mathcal{G}_{m'}(p))\right]dp\end{array}\right\}+o(1)\\
&= \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\left[\left(S^m_{m,m',j}-S^{m'}_{m,m',j}\right)-\boldsymbol{E}\left(S^m_{m,m',j}-S^{m'}_{m,m',j}\right)\right]+o(1)
\end{aligned}$$

where the third inequality holds since $(\mathcal{G}_m-\mathcal{G}_{m'})=o(N^{-1/2})$*. Then by the Central Limit Theorem* $\sqrt{N}\left(\widehat{\mathcal{GT}}-\mathcal{GT}\right)\xrightarrow{D}\mathcal{N}\left(0,\sigma^2_{\mathcal{GT}}\right)$. ■

Note that convergence still holds for the case where the sample sizes are different for the distributions under consideration. To see that, first note that for a constant $r=\frac{N_m}{N_{m'}}$, and defining,

$$s^q_{m,j} = \int_{\mathcal{P}_q}Y^m_{(j)}\mathbb{1}\left(Y^m_{(j)}\le Y(p)\right)dp$$

so that the last line of the proof is,

$$\sqrt{N}_m\left(\widehat{\mathcal{GT}}-\mathcal{GT}\right) = \left\{\begin{array}{cc} & \left[N_m^{-1/2}\sum_{j=1}^{N_m}\left(s^m_{m,j}-\mathbf{E}\left(s^m_{m,j}\right)\right)-\sqrt{r}N_{m'}^{-1/2}\sum_{j=1}^{N_{m'}}\left(s^m_{m',j}-\mathbf{E}\left(s^m_{m',j}\right)\right)\right]\\ - & \left[N_m^{-1/2}\sum_{j=1}^{N_m}\left(s^{m'}_{m,j}-\mathbf{E}\left(s^{m'}_{m,j}\right)\right)-\sqrt{r}N_{m'}^{-1/2}\sum_{j=1}^{N_{m'}}\left(s^{m'}_{m',j}-\mathbf{E}\left(s^{m'}_{m',j}\right)\right)\right]\end{array}\right\}+o(1)$$

and as in the above proof, by the Central Limit Theorem $\sqrt{N}_m\left(\widehat{\mathcal{GT}}-\mathcal{GT}\right)\xrightarrow{D}\mathcal{N}\left(0,\widetilde{\sigma}^2_{\mathcal{GT}}\right)$, where $\widetilde{\sigma}^2_{\mathcal{GT}}$ is a function of r. The critical point here being the fact that the measure will continue to be convergent, validating the use of the subsampling procedure proposed by Politis and Romano (1994) and Politis et al. (1999).

The asymptotic distribution of the *Lorenz Transvariation* can be similarly shown. First define,

$$\mathbb{S}^q_{m,m',j} = \int_{\mathcal{P}_q}\left[\frac{Y^m_{(j)}\mathbb{1}\left(Y^m_{(j)}\le Y(p)\right)}{\widehat{\mu}_m}-\frac{Y^{m'}_{(j)}\mathbb{1}\left(Y^{m'}_{(j)}\le Y(p)\right)}{\widehat{\mu}_{m'}}\right]dp \tag{9}$$

$$\mathbb{K}_{m,q,j} = \sum_{q=1}^{M}\int_{\mathcal{P}_q}\left[\frac{Y^q_{(j)}\mathbb{1}\left(Y^q_{(j)}\le Y(p)\right)}{\widehat{\mu}_q}-\frac{Y^{m}_{(j)}\mathbb{1}\left(Y^{m}_{(j)}\le Y(p)\right)}{\widehat{\mu}_{m}}\right]dp = \sum_{q=1}^{M}\mathbb{S}^q_{q,m,j} \tag{10}$$

Definition 2 *The Lorenz Transvariation between two Lorenz curves A and B is defined as,*

$$\mathcal{T} := \int_0^1 |\,\mathcal{L}_A(p)-\mathcal{L}_B(p)|dp \tag{11}$$

$\mathcal{T}$ can in turn be written as,

$$\mathcal{T} = \int_0^1 \max\{\mathcal{L}_A(p)-\mathcal{L}_B(p),\mathcal{L}_B(p)-\mathcal{L}_A(p)\}dp \tag{12}$$

and the limit distribution of $\mathcal{T}$ is as follows.

Theorem 2 *The Lorenz Transvariation , $\widehat{\mathcal{T}}$, is asymptotically normal. That is, $\sqrt{N}\left(\widehat{\mathcal{T}}-\mathcal{T}\right)\xrightarrow{D}\mathcal{N}\left(0,\sigma^2_{\mathcal{T}}\right)$, where,*

$$\begin{aligned}\lambda^2_{\mathcal{T}} &= \boldsymbol{Var}\left(\mathbb{S}^m_{m,m'}\right)+\boldsymbol{Var}\left(\mathbb{S}^{m'}_{m,m'}\right)-2\,\boldsymbol{Cov}\left(\mathbb{S}^m_{m,m'},\mathbb{S}^{m'}_{m,m'}\right) \\ &= \lambda^2_m+\lambda^2_{m'}-2\lambda_{m,m'}\end{aligned} \tag{13}$$

$$\lambda^2_m = \int_{\mathcal{P}_m}(\mathcal{L}_m(p)-\mathcal{L}_{m'}(p))^2dp-\left(\int_{\mathcal{P}_m}(\mathcal{L}_m(p)-\mathcal{L}_{m'}(p))dp\right)^2 \tag{14}$$

$$\lambda_{m,m'} = \begin{array}{l}\int_{\mathcal{P}_m}\int_{\mathcal{P}_{m'}}\left[\begin{array}{ll} & \mathcal{L}_m(p\wedge p')-\mathcal{L}_m(p)\,\mathcal{L}_{m'}(p') \\ - & \mathcal{L}_m(p')\,\mathcal{L}_{m'}(p)+\mathcal{L}_{m'}(p\wedge p')\end{array}\right]dpdp' \\ -\ \left[\int_{\mathcal{P}_m}\left[\mathcal{L}_m(p)-\mathcal{L}_{m'}(p)\right]dp\right]\left[\int_{\mathcal{P}_{m'}}\left[\mathcal{L}_{m'}(p')-\mathcal{L}_m(p')\right]dp'\right]\end{array} \tag{15}$$

Proof of Theorem 2 *Proof of Theorem 2 is similar to the proof of theorem 1.* ∎

2.2 Relative Potential Lorenz Transvariation

However, the *Generalized Lorenz Transvariation* and the *Lorenz Transvariation* do not permit a complete ranking of the Lorenz curves under consideration since they are just pairwise comparisons, nor is $\mathcal{GT}$ or $\mathcal{T} \in [0,1]$. In other words, there is no common basis against which all the Lorenz curves are compared, so that a consistent ordering can be achieved. What is needed is such a common base of comparison. To that end, this can be achieved by creating synthetically an upper envelope consisting of piecewise combinations of all the Generalized Lorenz curves under consideration. Then the transvariation measure of any Generalized Lorenz $m \in \mathcal{Q}$ is the *Potential Generalized Lorenz Transvariation*, and it will provide a consistent ordering.

To understand the measure, first denote the upper envelope generated by $\max_{\cup_{m=1}^{M} \mathcal{L}_m}\{\mathcal{G}(p)\} = \overline{\mathcal{G}}(p)$, then denote $\overline{\mathcal{G}}(p)$ as the synthetic baseline against which all the Generalized Lorenz curves are compared against. Precisely,

Definition 3 *The Potential Generalized Lorenz Transvariation is defined as the area between the synthetic empirical best case scenario represented by $\overline{\mathcal{G}}(p)$, and the m^{th} generalized Lorenz curve,*

$$\begin{aligned}
\mathcal{GB}_m &:= \int_{\mathcal{P}} \left[\max_{q\in\mathcal{Q}}\{\mathcal{G}_q(p)\} - \mathcal{G}_m(p)\right] dp \qquad (16)\\
&= \int_{\mathcal{P}} \left[\max_{q\in\mathcal{Q}}\{\mathcal{G}_q(p) - \mathcal{G}_m(p)\}\right] dp \\
&= \int_{\mathcal{P}} \left[\overline{\mathcal{G}}(p) - \mathcal{G}_m(p)\right] dp
\end{aligned}$$

The following theorem shows that $\widehat{\mathcal{GB}}_m$ for $m = 1, \ldots, M$ is asymptotically normal.

Theorem 3 *The Potential Generalized Lorenz Transvariation, $\widehat{\mathcal{GB}}_m$, is asymptotically normal. That is, $\sqrt{N}\left(\widehat{\mathcal{GB}}_m - \mathcal{GB}_m\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2_{\mathcal{GB}_m}\right)$, where $\sigma^2_{\mathcal{GB}_m}$ is,*

$$\begin{aligned}
\sigma^2_{\mathcal{GB}_m} &= \sum_{q=1}^{M} \boldsymbol{Var}\left(S^q_{q,m}\right) - 2\sum_{q=1}^{M}\sum_{q'\neq q}^{M} \boldsymbol{Cov}\left(S^q_{q,m}, S^{q'}_{q',m}\right) \\
&= \sum_{q=1}^{M} \sigma_q^2 - 2\sum_{q=1}^{M}\sum_{q'\neq q}^{M} \sigma_{q,q'} \qquad (17)
\end{aligned}$$

Proof of Theorem 3 *The asymptotic distribution of $\mathcal{GB}_m$ is,*

$$\begin{aligned}
\sqrt{N}\left(\widehat{\mathcal{GB}}_m - \mathcal{GB}_m\right) &= \sqrt{N}\left\{\int_{\mathcal{P}} \max_{q\in\mathcal{Q}}\left[\widehat{\mathcal{G}}_q(p) - \widehat{\mathcal{G}}_m(p)\right]dp - \int_{\mathcal{P}} \max_{q\in\mathcal{Q}}[\mathcal{G}_q(p) - \mathcal{G}_m(p)]dp\right\} \\
&= \sqrt{N}\sum_{q=1}^{M}\int_{\mathcal{P}_q}\left[\max_{q\in\mathcal{Q}}\left\{\widehat{\mathcal{G}}_q(p) - \widehat{\mathcal{G}}_m(p)\right\} - \max_{q\in\mathcal{Q}}\left\{\mathcal{G}_q(p) - \mathcal{G}_m(p)\right\}\right]dp \\
&= \sqrt{N}\sum_{q=1}^{M}\int_{\mathcal{P}_q}\max_{q\neq q'\in\mathcal{Q}}\left\{\left(\widehat{\mathcal{G}}_q(p) - \widehat{\mathcal{G}}_m(p)\right) - \left(\mathcal{G}_{q'}(p) - \mathcal{G}_m(p)\right)\right\}dp \\
&= \sqrt{N}\sum_{q=1}^{M}\int_{\mathcal{P}_q}\left[\left(\widehat{\mathcal{G}}_q(p) - \mathcal{G}_q(p)\right) - \left(\widehat{\mathcal{G}}_m(p) - \mathcal{G}_m(p)\right) + \left(\mathcal{G}_q(p) - \mathcal{G}_{q'}(p)\right)\right]dp \\
&= \sqrt{N}\sum_{q=1}^{M}\int_{\mathcal{P}_q}\left[\left(\widehat{\mathcal{G}}_q(p) - \mathcal{G}_q(p)\right) - \left(\widehat{\mathcal{G}}_m(p) - \mathcal{G}_m(p)\right)\right]dp + o(1) \\
&= \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\left(K_{m,q,j} - \boldsymbol{E}(K_{m,q,j})\right) + o(1) \qquad (18)
\end{aligned}$$

The fourth equality holds since $(\mathcal{G}_q(p) - \mathcal{G}_{q'}(p))$ is $o(N^{-1/2})$, so that by the Central Limit Theorem $\sqrt{N}\left(\widehat{\mathcal{GB}}_m - \mathcal{GB}_m\right) \xrightarrow{D} \mathcal{N}(0, \sigma^2_{\mathcal{GB}})$. ■

Although, the measure would provide a complete ordering of all Generalized Lorenz curves under consideration, its support is not $[0, 1]$. To do so, the total area within which $\mathcal{GB}_m$ could possibly vary needs to be defined, and it is this total area of variation against which $\mathcal{GB}_m$ is measured that would yield a measure within the $[0, 1]$ support. That is we define the total area of deviation between the *synthetic worst and best* case for the Generalized Lorenz curves under comparison. To that end, we define here the *Maximum Generalized Lorenz Transvariation*,

Definition 4 *Define $\min_{\cup_{m=1}^{M}\mathcal{G}_m}\{\mathcal{G}(p)\} = \underline{\mathcal{G}}(p)$, then the Maximum Generalized Lorenz Transvariation is,*

$$\begin{aligned}
\mathcal{GM} &= \int_{\mathcal{P}}\left[\max_{q\in Q}\{\mathcal{G}_q(p)\} - \min_{q'\in Q}\{\mathcal{G}_{q'}(p)\}\right]dp \qquad (19) \\
&= \int_{\mathcal{P}}\max_{q\neq q'\in Q}\left\{\mathcal{G}_q(p) - \mathcal{G}_{q'}(p)\right\}dp
\end{aligned}$$

$\mathcal{GM}$ provides the maximal difference any Generalized Lorenz curve can differ from another given the set of generalized lorenz curves under consideration. To examine the distributional characteristics, denote Π_M for the set of all the permutations of the M Generalized

Lorenz curves, where a typical vector denoted by $\pi = [\pi_1, \pi_2, \ldots, \pi_M]' \in \Pi_M$, $\pi_m = \pi(m)$, $m \in \mathcal{Q}$, such that $\mathcal{P}_m = \cup_{\{\pi:\pi_M=m\}}\Gamma_\pi$, where $\Gamma_\pi = \{p \in \mathcal{P} : \mathcal{G}_{\pi_1}(p) < \mathcal{G}_{\pi_2}(p) < \ldots \mathcal{G}_{\pi_M}(p)\}$. In other words the union of all the permutations where $\max_{j\in\mathcal{Q}}\{\mathcal{G}_j(p)\} = \mathcal{G}_m(p)$. To be clear, π_m then is a *catchment* of the M distributions under consideration that is the m^{th} highest.

Then combining both the measures of equations (16) and (19) we can define the *Relative Potential Generalized Lorenz Transvariation.*

Definition 5 *The Relative Potential Generalized Lorenz Transvariation is*

$$\mathcal{GR}_m = \frac{\int_{\mathcal{P}} \left[\max_{q\in\mathcal{Q}} \{\mathcal{G}_q(p) - \mathcal{G}_m(p)\}\right] dp}{\int_{\mathcal{P}} \max_{q\neq q'\in Q} \{\mathcal{G}_q(p) - \mathcal{G}_{q'}(p)\}\, dp} \tag{20}$$

and it has a support of $[0, 1]$.

An inequality measure using $\mathcal{GR}_m$ can be thus defined:

Definition 6 *The Relative Potential Generalized Lorenz Transvariation Inequality Measure is,*

$$\mathcal{GI}_m = 1 - \frac{\int_{\mathcal{P}} \left[\max_{q\in\mathcal{Q}} \{\mathcal{G}_q(p) - \mathcal{G}_m(p)\}\right] dp}{\int_{\mathcal{P}} \max_{q\neq q'\in Q} \{\mathcal{G}_q(p) - \mathcal{G}_{q'}(p)\}\, dp} = 1 - \mathcal{GR}_m \tag{21}$$

and $\mathcal{GI}_m \in [0, 1]$.

Thus for Generalized Lorenz curve $m \in \mathcal{Q}$ which is closest to the upper envelope would have $\mathcal{GI}_m$ closest to 1, and implies that the underlying population exhibits the lowest level of income inequality. On the other hand, for $m' \in \mathcal{Q}$, $m \neq m'$, which is furthest away from the upper envelope, and in consequence, closest to the lower hull would have $\mathcal{GI}_{m'}$ close to 0, and exhibits the highest level of income inequality among the M populations under consideration.

Evaluating $\mathcal{GI}_m$ for all M Generalized Lorenz curves under consideration would yield a complete ranking of differential in income inequality. Further, given the asymptotic distribution, inference on these measures are easily performed. However, this is complicated by the crossing of Generalized Lorenz curves in empirical applications, which requires the estimation of this intersection points, or alternately using subsampling methods. Since subsampling as proposed by Politis and Romano (1994) and Politis et al. (1999) requires only that the statistic be convergent, the technique is adopted here and the following section will discuss this technique in greater detail.

To determine the asymptotic distribution of $\widehat{\mathcal{GI}}_m$ define now,

$$\Gamma_\pi = \{p \in \mathcal{P} : \mathcal{G}_{\pi_1}(p) < \mathcal{G}_{\pi_2}(p) < \ldots \mathcal{G}_{\pi_M}(p)\} \tag{22}$$

$$A_j = \sum_{\pi \in \Pi_M} \int_{\Gamma_\pi} \left[Y_{(j)}^{\pi_M} \mathbb{1}\left(Y_{(j)}^{\pi_M} \leq Y(p) \right) - Y_{(j)}^{\pi_1} \mathbb{1}\left(Y_{(j)}^{\pi_1} \leq Y(p) \right) \right] dp \tag{23}$$

Theorem 4 *The Relative Potential Generalized Lorenz Transvariation, $\widehat{\mathcal{GR}}_m$, is asymptotically normal. That is, $\sqrt{N}\left(\widehat{\mathcal{GR}}_m - \mathcal{GR}_m\right) \xrightarrow{D} \mathcal{N}\left(0, \Delta'_{\mathcal{G},m}\sigma^2_{\mathcal{GB}_m}\Delta_{\mathcal{G},m}\right)$, where $\sigma^2_{\mathcal{GB}_m}$ is as defined in theorem 3, and*

$$\Delta_{\mathcal{G},m} = \begin{bmatrix} \frac{1}{\mathcal{GM}} \\ -\frac{\mathcal{GB}_m}{\mathcal{GM}^2} \end{bmatrix} \tag{24}$$

so that $\sqrt{N}\left(\widehat{\mathcal{GI}}_m - \mathcal{GI}_m\right) \xrightarrow{D} \mathcal{N}\left(0, \Delta'_{\mathcal{G},m}\sigma^2_{\mathcal{GB}_m}\Delta_{\mathcal{G},m}\right)$.

Proof of Theorem 4 *Note that,*

$$\begin{aligned}
\sqrt{N}\left(\widehat{\mathcal{GM}} - \mathcal{GM}\right) &= \sqrt{N} \int_{\mathcal{P}} \left[\max_{q \neq q' \in Q} \left\{ \widehat{\mathcal{G}}_q(p) - \widehat{\mathcal{G}}_{q'}(p) \right\} - \max_{q \neq q' \in Q} \left\{ \mathcal{G}_q(p) - \mathcal{G}_{q'}(p) \right\} \right] dp \\
&= \sqrt{N} \sum_{\pi \in \Pi_M} \int_{\Gamma_\pi} \left[\max_{q \neq q' \in Q} \left\{ \widehat{\mathcal{G}}_q(p) - \widehat{\mathcal{G}}_{q'}(p) \right\} - \max_{q \neq q' \in Q} \left\{ \mathcal{G}_q(p) - \mathcal{G}_{q'}(p) \right\} \right] dp \\
&= \sqrt{N} \sum_{\pi \in \Pi_M} \int_{\Gamma_\pi} \left[\max_{q \neq q' \in Q} \left\{ \widehat{\mathcal{G}}_q(p) - \widehat{\mathcal{G}}_{q'}(p) \right\} - \left(\mathcal{G}_{\pi_M}(p) - \mathcal{G}_{\pi_1}(p) \right) \right] dp \\
&= \sqrt{N} \sum_{\pi \in \Pi_M} \int_{\Gamma_\pi} \left[\max_{q \neq q' \in Q} \left\{ \widehat{\mathcal{G}}_{\pi_M}(p) - \widehat{\mathcal{G}}_{\pi_1}(p) \right\} - \left(\mathcal{G}_{\pi_M}(p) - \mathcal{G}_{\pi_1}(p) \right) \right] dp + o(1) \\
&= \frac{1}{\sqrt{N}} \sum_{j=1}^{N} (A_j - \boldsymbol{E}(A_j)) + o(1) \qquad (25)
\end{aligned}$$

Then by the results from theorem 3, and using the Delta Method,

$$\left(\widehat{\mathcal{GR}}_m - \mathcal{GR}_m\right) = \left(\frac{\widehat{\mathcal{GB}}_m}{\widehat{\mathcal{GM}}} - \frac{\mathcal{GB}_m}{\mathcal{GM}} \right) \to \mathcal{N}\left(0, \Delta'_{\mathcal{G},m}\sigma^2_{\mathcal{GB}_m}\Delta_{\mathcal{G},m}\right)$$

for $m = \{1, \ldots, M\}$, and the rest of the theorem follows. ■

The same idea can be applied to the Lorenz Transvariation, the proofs of which are similar, augmented by the following definition:

Definition 7 *The Potential Lorenz Transvariation is defined as the area between the synthetic empirical best case scenario represented by* $\overline{\mathcal{L}}(p)$*, and the* m^{th} *Lorenz curve,*

$$\begin{aligned}
\mathcal{B}_m &:= \int_{\mathcal{P}} \left[\max_{q\in\mathcal{Q}}\{\mathcal{L}_q(p)\} - \mathcal{L}_m(p) \right] dp \qquad (26)\\
&= \int_{\mathcal{P}} \left[\max_{q\in\mathcal{Q}} \{\mathcal{L}_q(p) - \mathcal{L}_m(p)\} \right] dp \\
&= \int_{\mathcal{P}} \left[\overline{\mathcal{L}}(p) - \mathcal{L}_m(p) \right] dp
\end{aligned}$$

Theorem 5 *The Potential Lorenz Transvariation,* $\widehat{\mathcal{B}}_m$*, is asymptotically normal. That is,* $\sqrt{N}\left(\widehat{\mathcal{B}}_m - \mathcal{B}_m\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma^2_{\mathcal{B}_m}\right)$*, where* $\sigma^2_{\mathcal{B}_m}$ *is,*

$$\begin{aligned}
\lambda^2_{\mathcal{B}_m} &= \sum_{q=1}^{M} \boldsymbol{Var}\left(\$^{q}_{q,m}\right) - 2\sum_{q=1}^{M}\sum_{q'\neq q}^{M} \boldsymbol{Cov}\left(\$^{q}_{q,m}, \$^{q'}_{q',m}\right) \\
&= \sum_{q=1}^{M} \lambda^2_q - 2\sum_{q=1}^{M}\sum_{q'\neq q}^{M} \lambda_{q,q'} \qquad (27)
\end{aligned}$$

Proof of Theorem 5 *Proof of Theorem 5 is similar to the proof of theorem 3.*■

Then an inequality measure based on the Lorenz curve needs the following definitions for the Maximum Lorenz Transvariation, and the Relative Potential Lorenz Transvariation.

Definition 8 *Define* $\min_{\cup_{m=1}^{M}\mathcal{L}_m}\{\mathcal{L}(p)\} = \underline{\mathcal{L}}(p)$*, then the Maximum Lorenz Transvariation is,*

$$\begin{aligned}
\mathcal{M} &= \int_{\mathcal{P}} \left[\max_{q\in Q}\{\mathcal{L}_q(p)\} - \min_{q'\in Q}\{\mathcal{L}_{q'}(p)\} \right] dp \qquad (28)\\
&= \int_{\mathcal{P}} \max_{q\neq q'\in Q} \{\mathcal{L}_q(p) - \mathcal{L}_{q'}(p)\} \, dp
\end{aligned}$$

Then combining both the measures of equations (26) and (28) we can define the *Relative Potential Generalized Lorenz Transvariation.*

Definition 9 *The Relative Potential Generalized Lorenz Transvariation is*

$$\mathcal{R}_m = \frac{\int_{\mathcal{P}} \left[\max_{q\in\mathcal{Q}} \{\mathcal{L}_q(p) - \mathcal{L}_m(p)\}\right] dp}{\int_{\mathcal{P}} \max_{q\neq q'\in Q} \{\mathcal{L}_q(p) - \mathcal{L}_{q'}(p)\} \, dp} \qquad (29)$$

and it has a support of $[0, 1]$*.*

The inequality measure based on the Lorenz curve using $\mathcal{R}_m$ can be thus defined:

Definition 10 *The Relative Potential Lorenz Transvariation Inequality Measure is,*

$$\mathcal{I}_m = 1 - \frac{\int_{\mathcal{P}} \left[\max_{q\in\mathcal{Q}} \left\{\mathcal{L}_q(p) - \mathcal{L}_m(p)\right\}\right] dp}{\int_{\mathcal{P}} \max_{q\neq q'\in Q} \left\{\mathcal{L}_q(p) - \mathcal{L}_{q'}(p)\right\} dp} = 1 - \mathcal{R}_m \tag{30}$$

and $\mathcal{I}_m \in [0,1]$.

Let Ψ_M denote the set of all the permutations of the M Lorenz curves, where the typical vector is $\psi = [\psi_1, \psi_2, \ldots, \psi_M]' \in \Psi_M$, such that $\mathcal{P}_m = \cup_{\psi:\psi_M=m}\Theta_\psi$, where

$$\Theta_\psi = \{p \in \mathcal{P} : \mathcal{L}_{\psi_1}(p) < \mathcal{L}_{\psi_2}(p) < \ldots \mathcal{L}_{\psi_M}(p)\} \tag{31}$$

$$\mathbb{A}_j = \sum_{\psi\in\Psi_M} \int_{\Theta_\pi} \left[\frac{Y_{(j)}^{\psi_M} \mathbb{1}\left(Y_{(j)}^{\psi_M} \leq Y(p)\right)}{\widehat{\mu}_{\psi_M}} - \frac{Y_{(j)}^{\psi_1} \mathbb{1}\left(Y_{(j)}^{\psi_1} \leq Y(p)\right)}{\widehat{\mu}_{\psi_1}}\right] dp \tag{32}$$

Theorem 6 *The Relative Potential Lorenz Transvariation,* $\widehat{\mathcal{R}}_m$*, is asymptotically normal. That is,* $\sqrt{N}\left(\widehat{\mathcal{R}}_m - \mathcal{R}_m\right) \xrightarrow{D} \mathcal{N}\left(0, \Delta'_{\mathcal{L},m}\sigma^2_{\mathcal{B}}\Delta_{\mathcal{L},m}\right)$*, where* $\sigma^2_{\mathcal{B}}$ *is as defined in theorem 3, and*

$$\Delta_{\mathcal{L},m} = \begin{bmatrix} \frac{1}{\mathcal{M}_m} \\ -\frac{\mathcal{B}_m}{\mathcal{M}_m^2} \end{bmatrix} \tag{33}$$

so that $\sqrt{N}\left(\widehat{\mathcal{I}}_m - \mathcal{I}_m\right) \xrightarrow{D} \mathcal{N}\left(0, \Delta'_{\mathcal{L},m}\sigma^2_{\mathcal{B}}\Delta_{\mathcal{L},m}\right)$.

Proof of Theorem 6 *Proof of theorem 6 is similar to the proof of theorem 4.*■

2.3 Relationship of Lorenz Transvariation to Gini

Observe that since the Gini coefficient may be written as,

$$G \equiv G(\mathcal{L}) = \frac{\int_0^1 [p - \mathcal{L}(p)]dp}{\int_0^1 p dp}$$

we can think of measuring the distance between the Lorenz curves with respect to the 45° line instead of $\overline{\mathcal{L}}$ in the development of the transvariation measures. Specifically, we can write $\mathcal{B}_m$, $\mathcal{R}_m$, and $\mathcal{M}$ as,

$$\begin{aligned}
\mathcal{B}_m &= \int_{\mathcal{P}} \left[\overline{\mathcal{L}}(p) - \mathcal{L}_m(p)\right] dp \\
&= \frac{\int_{\mathcal{P}} \left\{[p - \mathcal{L}_m(p)] - [p - \overline{\mathcal{L}}(p)]\right\} dp}{\int_0^1 pdp} \\
&= G(\mathcal{L}_m) - G\left(\overline{\mathcal{L}}\right) \qquad (34)
\end{aligned}$$

and

$$\mathcal{M} = G\left(\underline{\mathcal{L}}\right) - G\left(\overline{\mathcal{L}}\right) \qquad (35)$$

so that $\mathcal{R}_m$ is,

$$\mathcal{R}_m = \frac{G(\mathcal{L}_m) - G\left(\overline{\mathcal{L}}\right)}{G\left(\underline{\mathcal{L}}\right) - G\left(\overline{\mathcal{L}}\right)} \qquad (36)$$

and $\mathcal{I}_m$ can be similarly written.

This suggests that the same inequality measure discussed thus far can be expressed in terms of the Gini coefficient. First note that,

$$\begin{aligned}
\max_{\cup_{m=1}^M} \mathfrak{G}_m \equiv \overline{\mathfrak{G}} &= \max_{\cup_{m=1}^M} \left\{\frac{\int_0^1 [p - \mathcal{L}_m(p)]dp}{\int_0^1 pdp}\right\} \\
&\leq \frac{\int_{\mathcal{P}} \left[p - \underline{\mathcal{L}}(p)\right] dp}{\int_0^1 pdp} := G(\underline{\mathcal{L}})
\end{aligned}$$

where the last equality follows since,

$$\int_{\mathcal{P}} \left[p - \underline{\mathcal{L}}(p)\right] dp \geq \max_{\cup_{m=1}^M} \left\{\int_0^1 [p - \mathcal{L}_m(p)]dp\right\}$$

because $\underline{\mathcal{L}} \leq \mathcal{L}_m(p)$ for all p, $m = \{1, \ldots, M\}$. By a similar argument, $\min_{\cup_{m=1}^M} \mathfrak{G}_m = \underline{\mathfrak{G}} \geq G(\overline{\mathcal{L}})$. Then we can define parallel concepts of Potential, Maximum, and Relative Transvariation in terms of the Gini coefficient as,

$$\mathfrak{B} = \mathfrak{G} - \underline{\mathfrak{G}} \qquad (37)$$

$$\mathfrak{M} = \overline{\mathfrak{G}} - \underline{\mathfrak{G}} \qquad (38)$$

$$\mathfrak{R} = \frac{\mathfrak{G} - \underline{\mathfrak{G}}}{\overline{\mathfrak{G}} - \underline{\mathfrak{G}}} \qquad (39)$$

where $\mathfrak{G} := G(\mathcal{L})$, and the formulae are for the *Potential Gini Transvariation*, *Maximum Gini Transvariation*, and *Relative Gini Transvariation* respectively. Note further that $\mathfrak{R} \in [0,1]$. Next, observe that since,

$$\begin{aligned} G\left(\underline{\mathcal{L}}\right) - G\left(\overline{\mathcal{L}}\right) &\geq \overline{\mathfrak{G}} - \underline{\mathfrak{G}} \\ \text{and } \mathfrak{G} - G\left(\overline{\mathcal{L}}\right) &\geq \mathfrak{G} - \underline{\mathfrak{G}} \end{aligned}$$

we can bound the Relative Gini Transvariation. First, define,

$$\begin{aligned} G(\underline{\mathcal{L}}) - \overline{\mathfrak{G}} &= \overline{\Delta} \geq 0 \\ G(\overline{\mathcal{L}}) - \underline{\mathfrak{G}} &= \underline{\Delta} \leq 0 \\ \overline{\Delta} - \underline{\Delta} &= \Delta \geq 0 \end{aligned}$$

Then,

$$\begin{aligned} \mathfrak{R}_m &= \frac{\frac{\int_0^1 [p - \mathcal{L}_m(p)] dp}{\int_0^1 p dp} - \underline{\mathfrak{G}}}{\overline{\mathfrak{G}} - \underline{\mathfrak{G}}} \\ &= \frac{\frac{\int_0^1 [p - \mathcal{L}_m(p)] dp}{\int_0^1 p dp} - G(\overline{\mathcal{L}}) + \underline{\Delta}}{G(\underline{\mathcal{L}}) - G(\overline{\mathcal{L}}) - \Delta} \\ &\geq \frac{\frac{\int_0^1 [p - \mathcal{L}_m(p)] dp}{\int_0^1 p dp} - G(\overline{\mathcal{L}})}{G(\underline{\mathcal{L}}) - G(\overline{\mathcal{L}})} = \mathcal{R}_m \end{aligned}$$

Further, since

$$\overline{\mathfrak{G}} = \mathfrak{G}(\underline{\mathcal{L}}) = G(\underline{\mathcal{L}})$$

then

$$1 = \overline{\mathfrak{R}} \geq \overline{\mathcal{R}}$$

In other words, the upper bound of the Relative Gini Transvariation is higher than the upper bound of the Relative Lorenz Transvariation. Similarly for the lower bound,

$$\begin{aligned}
\underline{\mathcal{R}} = \mathcal{R}(\overline{\mathcal{L}}) &= \frac{\frac{\int_{\mathcal{P}}[p-\overline{\mathcal{L}}]dp}{\int_0^1 pdp} - G(\overline{\mathcal{L}})}{G(\underline{\mathcal{L}}) - G(\overline{\mathcal{L}})} \\
&\geq \frac{\frac{\int_{\mathcal{P}}[p-\overline{\mathcal{L}}]dp}{\int_0^1 pdp} - G(\overline{\mathcal{L}}) + \underline{\Delta}}{G(\underline{\mathcal{L}}) - G(\overline{\mathcal{L}}) - \Delta} \\
&= \frac{\frac{\int_{\mathcal{P}}[p-\overline{\mathcal{L}}]dp}{\int_0^1 pdp} - \underline{\mathfrak{G}}}{\overline{\mathfrak{G}} - \underline{\mathfrak{G}}} \\
&= \frac{\frac{\min_{\cup_{m=1}^{M}}\left\{\int_{\mathcal{P}}[p-\overline{\mathcal{L}}]dp\right\}}{\int_0^1 pdp} - \underline{\mathfrak{G}}}{\overline{\mathfrak{G}} - \underline{\mathfrak{G}}} = \underline{\mathfrak{R}} = 0
\end{aligned}$$

Finally, an appropriate inequality measure with respect to the Relative Gini Transvariation would be,

$$\mathfrak{I}_m \;=\; 1 - \mathfrak{G}(\mathcal{L}_m) \tag{40}$$

Insofar as the Gini Transvariation can be written in terms of the Lorenz curve, the prior proofs of convergence can be adapted for that purpose. Further, in lieu of the complications due to intersections between Lorenz curves under consideration, the subsampling method may be adapted here as well for inference. A full discussion is omitted here as it is outside the scope of the current paper.

3 Estimation & Inference

The Lorenz Transvariation, Maximal Lorenz Transvariation, and Relative Potential Lorenz Transvariation and their Generalized counterparts can be estimated by recognizing the the integrals over the indicator functions are essentially step functions over the chosen quantiles, so that the transvariation measures are differences between these steps.

The intersection points can be appropriately estimated to minimize power loss via subsampling techniques as prescribed by Politis and Romano (1994), and extended and summarized in Politis et al. (1999). On a related procedure to stochastic dominance, see Linton et al. (2005) and Linton et al. (2014). The steps are,

1. For each of the populations under consideraion, obtain the relevant statistics of $\widehat{\mathcal{GI}}$ or $\widehat{\mathcal{I}}$ using the full sample, $W_N := \{Y_j : j = 1, \ldots, N\}$.

2. Generate subsamples of length b, $W_{n,b} := \{Y_n, \ldots, Y_{n+b-1}\}$ for $n = 1, \ldots, n-b+1$.

3. Calculate the relevant statistics $\widehat{\mathcal{GI}}_{n,b}$ or $\widehat{\mathcal{I}}_{n,b}$, using the subsamples $W_{n,b}$.

4. The sampling distribution of $\sqrt{N}\left(\widehat{\mathcal{GI}} - \mathcal{GI}\right)$ or $\sqrt{N}\left(\widehat{\mathcal{I}} - \mathcal{I}\right)$, is approximated by,

$$\widehat{\Phi}_{N,b}(\omega) = \frac{1}{N-b+1}\sum_{n=1}^{N-b+1} \mathbb{1}\left(\sqrt{b}\left(\widehat{\mathcal{GI}}_{n,b} - \widehat{\mathcal{GI}}\right) \leq \omega\right) \tag{41}$$

$$\widehat{\Psi}_{N,b}(\omega) = \frac{1}{N-b+1}\sum_{n=1}^{N-b+1} \mathbb{1}\left(\sqrt{b}\left(\widehat{\mathcal{I}}_{n,b} - \widehat{\mathcal{I}}\right) \leq \omega\right) \tag{42}$$

respectively.

5. Find the requisite α^{th} sample quantile of $\widehat{\Phi}_{N,b}(\omega)$ or $\widehat{\Psi}_{N,b}(\omega)$,

$$\phi_{N,b}(\alpha) := \inf\{\omega : \widehat{\Phi}_{N,b}(\omega) \geq \alpha\} \tag{43}$$

$$\psi_{N,b}(\alpha) := \inf\{\omega : \widehat{\Psi}_{N,b}(\omega) \geq \alpha\} \tag{44}$$

respectively.

6. The confidence interval for $\widehat{\mathcal{GI}}$ or $\widehat{\mathcal{I}}$ can be calculated using the respective formulae:

$$CI^{\phi}_{N,b} := \left[\widehat{\mathcal{GI}} - N^{-1/2}\phi_{N,b}\left(1 - \frac{\alpha}{2}\right), \widehat{\mathcal{GI}} - N^{-1/2}\phi_{N,b}\left(\frac{\alpha}{2}\right)\right] \tag{45}$$

$$CI^{\psi}_{N,b} := \left[\widehat{\mathcal{I}} - N^{-1/2}\psi_{N,b}\left(1 - \frac{\alpha}{2}\right), \widehat{\mathcal{I}} - N^{-1/2}\psi_{N,b}\left(\frac{\alpha}{2}\right)\right] \tag{46}$$

Following Politis and Romano (1994), the following Theorem shows that the confidence interval has an asymptotically correct coverage.

Theorem 7 *Given assumption 1 is true, then if* $b \to \infty$ *and* $\frac{b}{N} \to 0$, *then*

$$\Pr(\mathcal{GI} \in CI^{\phi}_{N,b}) \to (1-\alpha) \text{ as } N \to \infty$$

$$\Pr(\mathcal{I} \in CI^{\psi}_{N,b}) \to (1-\alpha) \text{ as } N \to \infty$$

Proof of Theorem 7 *Given that the data is i.i.d., and the convergence of the transvariation measures discussed, the proof follows directly from theorem 2.1 of Politis and Romano (1994)* ■

There remains the second order concern of the choice of the subsample size b, of which some guidance is provided in Politis and Romano (1994), and Linton et al. (2005). Although for the example in the following section, we used only a single subsample size where $b = \frac{N}{\ln(\ln(N))}$, Linton et al. (2005) provides some guidance on automating the process for inference. Linton et al. (2005) proposed the calculation of the mean and median critical values, and/or p-values of the measure. First denote

$$B_N = \left[b_{N_1} < b_{N_2} < \cdots < b_{N_{r_N}}\right]$$

as the sequence of subsample sizes, where $b_{N_j} < N$ are integers, and r_N is the number of elements in B_N. The choice of the subsample size are then integer values between $\ln(\ln(N))$ and $\frac{N}{\ln(\ln(N))}$. These subsample sizes can be applied consistently for all the Lorenz curves under comparison. In turn, the mean and median choice for the statistic for inference are used. For instance, suppose the null hypothesis is for $H_0 : \mathcal{GI} = 1$ or $H_0 : \mathcal{I} = 1$. Then from the appropriately generated sampling distribution under H_0, denote the sample quantitle as $\phi^1_{N,b_{N_i}}(\omega)$ and $\psi^1_{N,b_{N_i}}(\omega)$ respectively for $\mathcal{GI}$ and $\mathcal{I}$. Denote the critical values for each subsample i, $i = \{1, \ldots, r_N\}$, $\mathcal{C}^{\phi}_{N,b_{N_i}}(1-\alpha)$ and $\mathcal{C}^{\psi}_{N,b_{N_i}}(1-\alpha)$, for the Generalized Lorenz and Lorenz based inequality measures respectively. Then the decision is based on the mean and median values of the critical values at α respectively,

$$\overline{\mathcal{C}}^{\phi}_N(1-\alpha) = \frac{1}{r_N}\sum_{i=1}^{r_N} \mathcal{C}^{\phi}_{N,b_{N_i}}(1-\alpha) \tag{47}$$

$$\mathcal{C}^{\phi,\text{Median}}_N(1-\alpha) = \text{median}\{\mathcal{C}^{\phi}_{N,b_{N_i}}(1-\alpha) : i = 1, \ldots, r_N\} \tag{48}$$

and

$$\overline{\mathcal{C}}^{\psi}_N(1-\alpha) = \frac{1}{r_N}\sum_{i=1}^{r_N} \mathcal{C}^{\psi}_{N,b_{N_i}}(1-\alpha) \tag{49}$$

$$\mathcal{C}^{\psi,\text{Median}}_N(1-\alpha) = \text{median}\{\mathcal{C}^{\psi}_{N,b_{N_i}}(1-\alpha) : i = 1, \ldots, r_N\} \tag{50}$$

where the mean critical value/criterion is equivalent to basing inference on the average opinion, while the latter is the majority opinion. In the case on hand, the researcher rejects the null if $\widehat{\mathcal{GI}} < \overline{\mathcal{C}}^{\phi}_N(1-\alpha)$, or $\widehat{\mathcal{GI}} < \mathcal{C}^{\phi,\text{Median}}_N(1-\alpha)$.

4 An Example

We demonstrate here the use of the *Relative Potential Lorenz Transvariation* in providing a complete ordering, as well as the use of the subsampling technique for inference, using U.S. Current Population Survey drawn from the Integrated Public Use Microdata Series (IPUMS), for the years 2001 to 2016. Table 1 presents the summary statistics for log income for the 16 years. It is clear that the weighted average income seems to be higher between 2001-2008, and falling there after, returning to similar levels by 2016. The weighted standard deviations in turn were lower between 2001-2008 compared with 2009-2016, which corresponds with the 43$^{\text{rd}}$ and 44$^{\text{rd}}$ Presidency of George W. Bush and Barack H. Obama respectively.

Table 1: Summary Statistics

Year	2001	2002	2003	2004	2005	2006	2007	2008
Mean	10.3362	10.3262	10.3030	10.2889	10.2816	10.2859	10.3142	10.3237
S.D.	1.0759	1.0838	1.1286	1.1681	1.1584	1.1560	1.1006	1.0499
Maximum	13.0772	13.0961	13.2849	13.2374	13.1628	13.3285	13.2597	13.0810
Minimum	-3.1368	-3.1793	-3.9396	-2.4744	-3.7119	-3.4669	-3.1858	-2.7989
# of Obs.	37226	36877	36815	36151	35469	34934	34752	34166
Year	2009	2010	2011	2012	2013	2014	2015	2016
Mean	10.2689	10.2044	10.1984	10.1922	10.1903	10.2203	10.2308	10.2946
S.D.	1.1287	1.2881	1.2468	1.2599	1.2715	1.2402	1.2013	1.1529
Maximum	13.1475	13.1226	13.6700	13.7084	14.1313	13.9219	13.5618	13.8585
Minimum	-3.3835	-2.8445	-2.2977	-2.5525	-3.2514	-3.2035	-2.7768	-2.0593
# of Obs.	34147	33872	32690	31827	31748	30785	30429	28137

To see the potential that the standard pairwise comparisons of the Lorenz curves across the years under consideration may not provide definitive rank ordering, observe the crossing and proximity between the Lorenz curves across the 16 years in figure 1. It was also observed that the Lorenz curves between 2001-2008 were in general lower than those between 2009-2016 which will be clear using the proposed inequality measures here, $\mathcal{I}_m$, below.

Figure 1: Lorenz Curves for U.S. between 2001-2016

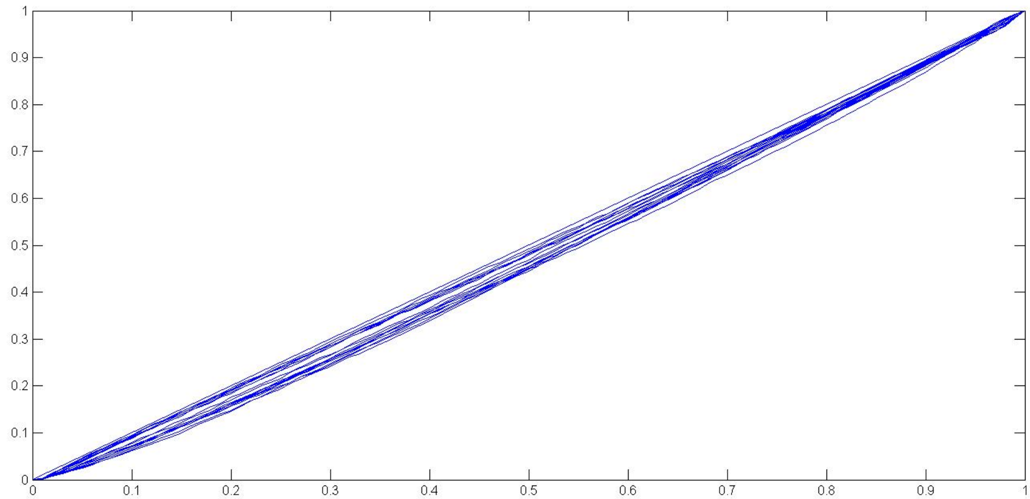

Table 2: log Income Lorenz Transvariation, 2001-2016

Year	2001	2002	2003	2004	2005	2006	2007	2008
$\mathcal{I}_m$	0.1702	0.2298	0.3274	0.3248	0.4474	0.4303	0.2821	0.2368
Rank	16	15	10	11	8	9	12	14
p-value, $\mathcal{I}_m = 1$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
p-value, $\mathcal{I}_m = 0$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Year	2009	2010	2011	2012	2013	2014	2015	2016
$ILorz_m$	0.5828	0.8139	0.8820	0.9738	0.9483	0.7764	0.7287	0.2559
Rank	7	4	3	1	2	5	6	13
p-value, $\mathcal{I}_m = 1$	0.0000	0.0000	0.1281	0.5488	0.0428	0.0000	0.0000	0.0000
p-value, $\mathcal{I}_m = 0$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.7304

Table 2 reports the results of the measure of inequality, $\mathcal{I}_m$ derived from the Relative Potential Lorenz Transvariation. The observed proximity of the measure to perfect equality in the latter half of the data examined reveals itself in the ordering, where the income distributions between 2009-2016 were closer to perfect equality than those between 2001-2008, with the exception of 2016. This could be interpreted as the result of the more egalitarian social policies of President Obama's administration, or the culmination of the educational policies of President Bush's administration.

The Lorenz curve however places weights on realizations at the tail, and it would hence be interesting to examine how the Generalized Lorenz curve, which magnifies the magnitude of incomes by the mean, might possibly offer a different set of rankings. This is reported in table 3, for the measure $\mathcal{GI}_m$. Interestingly, although the rankings did differ marginally, the overall observation that President Obama's presidency coincided with more equal income distributions remain, in the sense that the rankings for income distributions between 2009-2016 were in general higher than that between 2001-2008, with the exception of 2016. This highlights the stark difference in income distributions during the two presidencies, regardless of the use of Lorenz or Generalized Lorenz.

Table 3: log Income Generalized Lorenz Transvariation, 2001-2016

Year	2001	2002	2003	2004	2005	2006	2007	2008
$\mathcal{I}_m$	0.2398	0.2942	0.3739	0.3520	0.4808	0.4673	0.3374	0.2988
Rank	16	14	10	11	8	9	12	13
p-value, $\mathcal{I}_m = 1$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-value, $\mathcal{I}_m = 0$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Year	2009	2010	2011	2012	2013	2014	2015	2016
$\mathcal{I}_m$	0.6165	0.7874	0.8553	0.9495	0.9183	0.7674	0.7284	0.2817
Rank	7	4	3	1	2	5	6	15
p-value, $\mathcal{I}_m = 1$	0.0000	0.0000	0.0000	0.5834	0.1863	0.0000	0.0000	0.0000
p-value, $\mathcal{I}_m = 0$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1281

Finally, table 4 reports the p-values for cross year comparisons in the Lorenz Transvariation measure for $H_0 : \mathcal{I}_m = \mathcal{I}_{m'}$ for $m \neq m'$, that is it examines the statistical difference between the years under consideration.

Table 4: log Income Lorenz Transvariation, 2001-2016

Year		2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
	$\mathcal{I}_m$	0.1702	0.2298	0.3274	0.3248	0.4474	0.4303	0.2821	0.2368	0.5828	0.8139	0.8820	0.9738	0.9483	0.7764	0.7287	0.2559
2001	0.1702	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2002	0.2298	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2003	0.3274	0.0108	0.4126	0.2439	0.2448	0.0209	0.0377	0.3732	0.4584	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4713
2004	0.3248	0.0422	0.3129	0.0041	0.0077	0.0000	0.0000	0.2490	0.3671	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4804
2005	0.4474	0.0000	0.0000	0.3669	0.3997	0.0000	0.0000	0.1740	0.0071	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0363
2006	0.4303	0.0000	0.0708	0.0145	0.0240	0.0000	0.0000	0.3285	0.1415	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3445
2007	0.2821	0.2394	0.3714	0.0000	0.0000	0.0000	0.0000	0.0978	0.4123	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3777
2008	0.2368	0.0065	0.3133	0.0000	0.0000	0.0000	0.0000	0.0246	0.2574	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1657
2009	0.5828	0.0000	0.0000	0.0000	0.0000	0.3690	0.4372	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2010	0.8139	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3476	0.2153	0.0000	0.0000	0.1619	0.0075	0.0000
2011	0.8820	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3335	0.2971	0.0000	0.0048	0.1686	0.0293	0.0000
2012	0.9738	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2581	0.3830	0.4483	0.0000	0.0000	0.0000
2013	0.9483	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0256	0.3768	0.2650	0.0000	0.0000	0.0000
2014	0.7764	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2725	0.0156	0.0000	0.0000	0.4128	0.3919	0.0000
2015	0.7287	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3747	0.4746	0.1692	0.1915	0.2951	0.0930	0.0000
2016	0.2559	0.0000	0.0000	0.0837	0.0829	0.1325	0.1284	0.0000	0.0000	0.4631	0.0000	0.0000	0.0000	0.0000	0.0000	0.0096	0.0000

5 Conclusion

This paper provides the exact limit distributions of transvariation measures applied to the (Generalized) Lorenz curves. The idea behind transvariation is to ameliorate the common problem of a lack of resolution in the ranking of Lorenz curves, particularly when they cross which occurs with regular frequency in empirical work. These crossings in turn complicate comparisons of Lorenz curves since it would necessitate the estimation of their location. The convergence of the proposed transvariation measures in turn justifies the use of subsampling techniques in performing these inferences in lieu of the crossing, thus reducing the complexity, which was demonstrated in the example.

References

Anderson, G. and Leo, T. W. (2017a). On Providing a Complete Ordering of Non-Combinable Alternative Prospects. University of Toronto Discussion Paper.

Anderson, G. and Leo, T. W. (2017b). On Solving (Generalized) Gini and Lorenz Comparison Issues with Gini's (Adapted) Transvariation. University of Toronto Discussion Paper.

Anderson, G., Post, T., and Whang, Y.-J. (2017). Somewhere Between Utopia and Dystopia: Choosing From Multiple Incomparable Prospect. University of Toronto Discussion Paper.

Atkinson, A. B. (1970). On the Measurement of Inequality. *Journal of Economic Theory*, 2(3), 244–263.

Beach, C. M. and Davidson, R. (1983). Distribution-Free Statistical Inference with Lorenz Curves and Income Shares. *The Review of Economic Studies*, 50(4), 723–735.

Gini, C. (1916). Il Concetto di Transvariazione e le sue Prime Applicazioni. In C. Gini (Ed.), *Transvariazione*. Libreria Goliardica, (1–55).

Gini, C. (1959). *Transvariazione*. Libreria Goliardica, Rome.

Leshno, M. and Levy, H. (2002). Preferred by "All" and Preferred by "Most" Decision Makers: Almost Stochastic Dominance. *Management Science*, 48(8), 1074–1085.

Linton, O., Maasoumi, E., and Whang, Y.-J. (2005). Consistent Testing for Stochastic Dominance under General Sampling Schemes. *The Review of Economic Studies*, 72, 735–765.

Linton, O., Post, T., and Whang, Y.-J. (2014). Testing for the Stochastic Dominance Efficiency of a Given Portfolio. *The Econometrics Journal*, 17, S59–S74.

Politis, D. N. and Romano, J. P. (1994). Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions. *The Annals of Statistics*, 22(4), 2031–2050.

Politis, D. N., Romano, J. P., and Wolf, M. (1999). *Subsampling*. New York: Springer.

Seth, S. and Yalonetzky, G. (2016). Has the World Converged? A Robust Analysis of Non-Monetary Bounded Indictors. Leeds University Business School Discussion Paper.

Shorrocks, A. F. (1983). Ranking Income Distribution. *Economica*, 50(197), 3–17.

Zheng, B. (2016). Almost Lorenz Dominance. University of Colorado, Denver, Discussion Paper.