STATISTICAL INFERENCE FOR THE SEN-SHORROCKS-THON INDEX OF POVERTY INTENSITY

Kuan Xu Dalhousie University

The proposition of the modified Sen index (the SST index) for measuring poverty intensity represents an important advance in the literature of poverty measures. This index is useful in empirical research on income distribution and poverty because it satisfies a set of desirable properties while the Sen index and some other indices do not. This index is symmetric, monotonic, continuous, homogeneous of degree zero in incomes and the poverty line, and consistent with the transfer axiom. It also admits a useful geometric interpretation. Hence, there is a considerable interest on the part of economists to apply this measure to sample income data in order to draw valid statistical inference about poverty intensity. This paper examines the asymptotic distribution of the SST index estimator, and proposes a useful test for comparing two population SST indices over time or across states. This test can be used for evaluating the deprivation dominance relationship among income distributions.

1. INTRODUCTION

Research on poverty indices has received considerable attention since Sen (1976) proposed a poverty index and a set of desirable criteria for evaluating a poverty index. Poverty indices differ from the headcount ratio or income gap ratio in that they are designed to combine these two measures with a third dimension: income inequality amongst the poor. The measures of poverty intensity are useful for exploring the extent of poverty amongst the deprived population. That is one of the most important considerations in designing related social policies.

To measure poverty intensity, some economists attempt to refine existing poverty indices. The Sen index is not replication invariant, not continuous in individ-

Direct all correspondence to: Professor Kuan Xu, Department of Economics, Dalhousie University, Nova Scotia, Canada B3H 3J5.

Journal of Income Distribution, 8(1), 143-152

Copyright © 1998 JAI Press Ltd.

ISSN: 0926-6437

All rights of reproduction in any form reserved.

ual incomes, and fails to satisfy the transfer axiom. However, a desirable poverty index should be replication invariant. In other words, the value of the poverty index for the combination of two identical populations is the same as the value of the poverty index for any one of two identical populations. A desirable poverty index should be a continuous function of individual incomes. A desirable poverty index should also take a lower value when a transfer occurs from a poor individual to someone whose income is higher. Having noted these, Shorrocks (1995) proposed a modified Sen index for measuring the intensity of poverty. Zheng (1997) notes that this modification is identical to the limit of Thon's modified Sen index (1979). Hence, the index is referred to as the Sen-Shorrocks-Thon (or SST) index. This new index is replication invariant, continuous in individual incomes, homogeneous of degree zero in individual incomes and the corresponding poverty line. It is consistent with the transfer axiom. It is normalized to take values in the range from zero to one. It also admits a geometric interpretation.⁴

There is a considerable interest on the part of economists to apply this index to sample survey data in order to draw valid statistical inferences about poverty intensity. Economists rarely have the luxury of obtaining income data from the whole population. They have no choice but use sample survey data to get useful information about the underlying population. While other authors studied the statistical inferences of different aspects of poverty measures, this paper focuses on the method of statistical inference for the population SST index and the bootstrap method (see, for example, Bishop, Chow, & Zheng, 1995; Davidson & Duclos, 1998; Osberg & Xu, 1997, 1998; Rongve, 1997; Preston, 1995; Zheng, Cushing, & Chow, 1995).

This paper tries to answer the following questions: (1) Can we use the sample data to estimate the SST index? (2) If the answer to question (1) is a yes, how reliable is the estimate? (3) If the reliability of the SST index estimates can be characterized by a variance (standard deviation), is there any simpler way to estimate the variance (standard deviation)? (4) If there exists a simpler way, can one establish a test statistic so that two index estimates can be compared based on a sound statistic foundation? The answers to these questions must be provided based on the analysis of the SST index and some of the existing statistics literature. This paper shows that the SST index can be estimated based on sample data using equation (2) in section 2. Since the estimate may be affected by sampling variation, it is useful to understand the variance (standard deviation) of the SST index estimator. Lemmas 1 and 2 in section 3 provide the premises that are used to show that the SST index estimator is normally distributed (Proposition 1), and that there exists an analytical expression for computing the variance (standard deviation) of the SST index estimator (Proposition 2). The normality of the SST index estimator is an important condition for establishing a test for comparing two SST index estimates (Proposition 3). This paper also introduces the bootstrap method for estimating consistently the variance (standard deviation) of the SST index estimate. The bootstrap method is implemented in the following steps: (1) randomly draw a new sample from the original sample with replacement and compute the SST index based on the new sample; (2) repeat step (1) many (say 100) times; and (3) compute the variance (standard deviation) of the SST index estimate based on the collection of index estimates from the new samples. This paper uses the proposed procedure to evaluate changes in poverty intensity in the United States in 1969, 1979, and 1988. Although the individual mean income increased from the late 1960s to the late 1980s, poverty intensity increased significantly during the examined period, as shown by the SST index estimates and related test results.

The remainder of the paper is organized as follows. The SST index and some notation are introduced in section 2. In section 3, the asymptotic distribution of the SST index estimator is derived, and a statistical test for comparing two population SST indices is proposed. Some illustrative empirical examples are given in section 4. Finally, concluding remarks are offered in section 5.

2. THE SST INDEX AND SOME NOTATION

The population SST index is proposed assuming that all the income data of a population are known and nonstochastic. Let such an N-person population income vector be $Y = [Y_1, Y_2, ..., Y_N]'$, and the ordered population income vector $Y_0 = [Y_{(1)}, Y_{(2)}, ..., Y_{(N)}]'$ such that $Y_{(1)} \le Y_{(2)} \le ... \le Y_{(N)}$. Let Z > 0 be the population poverty line and Q be the number of people in the population whose income is less than the poverty line Z. For the i-th poor person, the poverty gap is given by $Z - Y_{(i)}$. The population SST index is defined as (see Shorrocks, 1995):

$$P(Y;Z) = \frac{1}{N^2} \sum_{i=1}^{Q} (2N - 2i + 1) \frac{Z - Y_{(i)}}{Z}.$$
 (1)

P(Y, Z) can be computed based on equation (1) if the population income vector is known.

However, the population incomes are usually unknown and inferences must be drawn from sample information. To consider the issue of statistical inference regarding the SST index, let $y = [y_1, y_2,..., y_n]'$ be a random sample of incomes from a random variable Y with a cumulative distribution function $F(u) = P(Y \le u)$ and a probability density function f. The ordered observations are given by $y_0 = [y_{(1)}, y_{(2)},..., y_{(n)}]'$. Let f be the number of poor people in the sample whose income is below the sample poverty line. There are two general approaches for treating the poverty line in the statistical inference. One treats the poverty line as given so that f in f in

$$P(y;z) = \frac{1}{n^2} \sum_{i=1}^{q} (2n - 2i + 1) \frac{z - y_{(i)}}{z} .$$
 (2)

Alternatively, the above index can be written as

$$P(y;z) = \frac{1}{n}W\left(1 - \left(\frac{1}{z}\right)\hat{y}_o\right),\tag{3}$$

where $W = \left[2 - \frac{1}{n}, 2 - \frac{3}{n}, \dots, 2 - \frac{2q-1}{n}\right]'$, **1** is a $q \times 1$ vector of ones, and $\bar{y}_0 = [y_{(1)}, y_{(2)}, \dots, y_{(q)}]'$ is a $q \times 1$ vector of incomes of the poor truncated from $y_0 = [y_{(1)}, y_{(2)}, \dots, y_{(n)}]$. The elements in $\left(1 - \left(\frac{1}{z}\right)\bar{y}_0\right)$ and W are arranged in descending order. This alternative expression of P(y; z) is used to derive the asymptotic variance of the SST index estimator.

3. ASYMPTOTIC DISTRIBUTION OF THE SST INDEX AND RELATED $Z_{P(Y;Z)}$ -TEST

Since the SST index estimator is a linear function of order statistics with a suitably chosen weight function, the result due to Stigler (1969, 1974) directly applies to P(y; z). Let J be such a weight function, and $y_{(i)}$ the i-th sample order statistic. Let $\sigma^2(J, F)$ be

$$\sigma^{2}(J,F) = S_{\infty}^{\infty} S_{\infty}^{\infty} J(F(u))J(F(v))[F(\min(u,v)) - F(u)F(v)]dudv, \tag{4}$$

where u and v are in the support of F. As shown in Lemma 1, Stigler (1969, 1974) demonstrates that a linear function of order statistics of the form

$$S_n = \frac{1}{n} \sum_{i=1}^{n} J(i/(n+1)) y_{(i)}$$
 (5)

is asymptotically normal under a set of conditions.⁵

Lemma 1 If Ey_i^2 is finite, and J is bounded and continuous almost everywhere for the inverse function of the cumulative distribution function F, then

$$\lim_{n \to \infty} n\sigma^2(S_n) = \sigma^2(J, F).$$
 (6)

Also, if $\sigma^2(J,F) > 0$, then

$$\frac{(S_n - ES_n)}{\sigma(S_n)} \xrightarrow{d} N(0, 1) \tag{7}$$

A result of Ghosh (1971), which makes minimal assumptions on the population, is useful for deriving the analytical expression for the asymptotic variance of the SST index estimator. Denote the truncated vector of order statistics $\bar{y}_0 = [y_{(1)}, y_{(2)}, \dots, y_{(q)}]'$ as the vector of sample quantiles $\bar{y}_0 = [y_{(1)}, y_{(2)}, \dots, y_{(q)}]'$, corresponding to a set of suitably chosen p_j 's such that $0 < p_1 < p_2 < \dots < p_q = 1$, where $[\cdot]$ is an integer operator. The population quantiles corresponding to the sample quantiles is denoted by $\xi = [\xi_{p_1}, \xi_{p_2}, \dots, \xi_{p_q}]'$. The following lemma shows that under minimal assumptions, the vector of sample quantiles has the asymptotic normal distribution with mean ξ and asymptotic variance-covariance matrix Ω .

Lemma 2 If $0 < f((\xi_{p_j}) \text{ for } j = 1, 2, ..., q, \text{ then}$

$$[\sqrt{n}(y_{([np_1])} - \xi_{p_1}), ..., \sqrt{n}(y_{([np_q])} - \xi_{p_q})]' = \sqrt{n}[y_0 - \xi] \xrightarrow{d} N(0, \Omega) , (8)$$

where the elements ω_{jj}' of Ω are given by

$$\omega_{jj}' = \left(\frac{p_j(1 - p_{j'})}{f(\xi_{p_j})f(\xi_{p_{j'}})}\right) \quad \text{for all } j \text{ and } j', \text{ and } j \le j'.$$
(9)

Given Lemma 1, Proposition 1 shows that the SST index estimator also has an asymptotic normal distribution. This result is useful for our discussion of the $Z_{P(y;z)}$ -test discussed later in the paper.

Proposition 1 If P(y;z) satisfies the conditions in Lemma 1, then

$$\frac{(P(y;z)-EP(y;z))}{\sigma(P(y;z))} \stackrel{d}{\longrightarrow} N(0,1).$$

Proof: P(y;z) is a special case of S_n and satisfies the conditions of Lemma 1. Hence, the above result follows.

The following proposition demonstrates the analytical expression of the asymptotic variance of the SST index estimator.

Proposition 2 The asymptotic variance of the SST index estimator is given by

$$\sigma^2(P(y;z)) = \frac{1}{n^3 z^2} W' \Omega W,$$

where Ω is defined in Lemma 2.

Proof: According to equation (3) and Lemma 2, P(y;z) is a linear function of y_0 . The asymptotic variance of P(y;z) is obtained by the delta method.

To estimate Ω , the probability density function evaluated at ξ_{p_j} , $f(\xi_{p_j})$, in equation (9) needs to be replaced with the empirical probability density function evaluated at $y_{(p_j)}$, $\hat{f}(y_{(p_j)})$ which is cumbersome to compute. A simple but computing-intensive alternative, the bootstrap method (see Efron, 1979, 1982; Efron & Tibshirani, 1986), is used to estimate consistently the variance (standard deviation) for the SST index estimator $[\hat{\sigma}^2(P(y;z))]$.

When two sample SST indices [say $P(y_a; z_a)$ and $P(y_b; z_b)$] are estimated for two sample income data sets [say $\{y_a, z_a\}$ and $\{y_b, z_b\}$] that are from two data generating processes [say Y_a and Y_b], the comparison between the two population SST indices [say $P(Y_a; Z_a)$ and $P(Y_b; Z_b)$] can be made directly, and a standard asymptotic $Z_{P(y;z)}$ -test can be implemented for

$$H_0: P(Y_a; Z_a) - P(Y_b; Z_b) = 0 \text{ against } H_A: P(Y_a; Z_a) - P(Y_b; Z_b) > 0.$$

The null hypothesis states that the population SST indices for Y_a and Y_b are identical, while the alternative hypothesis states that the population SST index for Y_a is greater than that of Y_b , which implies that poverty intensity for Y_b is lower than that of Y_a .

Proposition 3 If Y_a and Y_b are independent, the asymptotic $Z_{P(y;z)}$ test statistic has the following asymptotic distribution:

$$Z_{P(y;z)} = \frac{P(y_a;z_a) - P(y_b;z_b)}{\sqrt{\hat{\sigma}^2(P(y_a;z_a)) + \hat{\sigma}^2(P(y_b;z_b))}} \stackrel{d}{\to} N(0,1)$$

Proof: Since each SST index estimator has an asymptotic normal distribution, the $Z_{P(y;z)}$ -test statistic is a standardized difference of the two estimators, its asymptotic distribution must be normal.

Due to the limited space, the satisfactory simulation results of the finite sample properties and the power of the test are not reported here, but they are available upon request.

Table 1
The SST Index for the US PSID Data: 1969, 1979 and 1988

Year	69	79	88
Sample size	2,206	4,173	6,244
Maximum income (\$):	249,755	736,474	1,193,745
Minimum income (\$):	114	138	1
Mean income (\$):	26,596	28,932	30,817
Std. dev. of income (\$)	19,474	24,098	32,225
Poverty line (\$):	11,396	12,382	12,341.5
Nob. of the poor:	404	846	1,354
The SST index:	0.1201	0.1335	0.1557
Std. dev. of the SST index:	0.0038	0.0030	0.0026

4. ILLUSTRATIVE EXAMPLES

The method of statistical inference for the SST index is applied to the pre-tax individual income data for 1969, 1979, and 1988 generated from the Panel Study of Income Dynamics of the United States.⁹

Although the selection of the poverty line is an active area of research itself, ¹⁰ the poverty line used here is based on Smeeding (1991) who suggests that it be computed as half the median of an income distribution. ¹¹ Table 1 shows the maximum, minimum, mean, and standard deviation of the pre-tax income data, the poverty line, and the number of the poor for 1969, 1979, and 1988, respectively. These numbers do not provide sufficient information for evaluating changes in the poverty intensity. Sometimes they could be even misleading. To understand the state and changes of poverty intensity, the SST index and its statistical inference procedure could be used. As shown in Table 1, the sample SST index is 0.1201 for 1969, 0.1335 for 1979, and 0.1557 for 1988. Based on these sample estimates, it appears that poverty intensity in the United States increased over the examined years.

Without a measure of sampling variation of the sample estimator of the population SST index, a researcher cannot conclude whether a difference between two sample SST indices reflects the true difference between the two population SST indices or merely the discrepancy due to sampling variation. Hence, it is necessary to compute the variance (standard deviation) and perform the $Z_{P(y;z)}$ -test. In Table 1, the standard deviations are provided based on the bootstrap method. To test the null hypothesis that the SST index for 1988 is equal to that for 1979 against the alternative hypothesis that the two are not equal, the $Z_{P(y;z)}$ -test statistic turns out to be 5.5919 and its p-value is 1.1227 x 10⁻⁸. For the null hypothesis that the SST index for 1979 is equal to that for 1969 against the alternative hypothesis that the two are not equal, the $Z_{P(y;z)}$ -test statistic is valued at 2.7698 and its p-value is 0.0028. The results indicate that the changes (or increases) in poverty intensity over the years could not be viewed as statistically insignificant. The evidence sug-

gests that poverty intensity in the United States had indeed increased over the last three decades although the pre-tax income level had also increased from \$26,596 in 1969 to \$30,817 in 1988. Indeed, poverty was significantly more intense in 1988.

5. CONCLUDING REMARKS

The proposition of the SST index represents an important advance in the literature of poverty measures. This paper has proposed a method of statistical inference for the population SST index, which is necessary for applying the theoretical measure to sample data in empirical work. The statistical issues discussed in the paper may help to foster the application of the SST index in social and economic research. ¹²

In response to the need for applying the SST index of poverty intensity to sample data, this paper provides the statistical justification for its asymptotic properties. It shows that the asymptotic distribution of the SST index estimator is normal, and that a $\mathbb{Z}_{P(y;z)}$ -test can be used to evaluate changes in poverty intensity of one population over time or examine the dominance relationship between poverty profiles of two populations at one point in time.

The method of statistical inference for the population SST index is relatively easy to apply. Since the standard deviation of the SST index estimator is somewhat complex to compute, this paper suggests that it be computed by the bootstrap method. As the computer becomes more powerful and computing more economic, the computing-intensive method is relatively simply to implement. As illustrative examples, the method of statistical inference is applied to the United States sample income data for 1969, 1979, and 1988. The results show that the differences among the poverty intensity measures in 1969, 1979, and 1988 are statistically significant and that poverty intensity was significantly more intense in 1988.

ACKNOWLEDGMENTS: The research support from Social Sciences and Humanities Research Council of Canada (SSHRC) and Dalhousie University Research Fund are gratefully acknowledged. I would like to thank M. Cross, T. Iscan, L. Osberg, U.L.G. Rao, Ian Rongve, and Buhong Zheng for their helpful comments.

NOTES

- 1. See, among others, Atkinson (1987), Besley (1990), Blackorby and Donaldson (1980), Donaldson and Weymark (1986), Foster, Greer and Thorbecke (1984), Foster and Shorrocks (1988, and 1991), Takayama (1979) and Thon (1979, 1983). In addition, Kakwani (1980), Foster (1984), Hagenaars (1986), and Seidl (1988) have provided useful surveys of this literature.
 - 2. That is defined as the proportion of the poor in the total population.
 - 3. That is defined as the ratio of average income gap of the poor to the poverty line.

- 4. See Shorrocks (1995) for details regarding the properties of the index.
- 5. One underlying assumption is that the extremal order statistics do not contribute too much to S_n [see Stigler (1969, 1974)]. This reflects the basic characteristics of the poverty gap data because they are bounded and unlikely to have extreme values.
- 6. The p_j 's can be selected so that represent all sample observations of the income levels that are less than the poverty line. In actual implementation of this test, there is no need to specify the percentiles if the bootstrap method is adopted. The theoretical argument presented here is to show that the SST index estimator has an asymptotic normal distribution.
 - 7. See Silverman (1986) for a variety of the density estimation methods.
- 8. A researcher may change the greater-than sign for the alternative hypothesis to either less-than or not-equal-to sign depending on the purpose of the test.
- 9. The data are kindly made available by E. Maasoumi of the Department of Economics at the Southern Methodist University.
- 10. Fisher (1992) provides a survey on the development and history of the poverty thresholds for the US economy, in which the limitations of the poverty thresholds are evaluated. Vaughan (1993) explores the use of the public's views to set income poverty thresholds, and finds that, since the late 1960's, the income level of the official measure had fallen below the Gallup-based threshold until 1989.
- 11. The fixed poverty line may also be used, but this will not affect the purpose of this demonstration.
- 12. Osberg and Xu (1997, 1998) extend the work presented in this paper and show that the SST index can be decomposed into the headcount ratio, average poverty gap ratio, and overall Gini index of poverty gap ratios. Some economists at Statistics Canada also work on poverty issues using the SST index.

REFERENCES

- Atkinson, A.B. 1987. "On the Measurement of Poverty," Econometrica, 55: 749-764.
- Besley, T. (1990). "Means Testing versus Universal Provision in Poverty Alleviation Programmes," Economica, 57: 119-129.
- Bishop, J.A., Chow, K.V., and Zheng, B. 1995. "Statistical Inference and Decomposable Poverty Measures," *Bulletin of Economic Research*, 47: 329-340.
- Blackorby, C., and Donaldson, D. 1980. "Ethical Indices for the Measurement of Poverty," Econometrica, 48: 1053-1060.
- David, H.A. 1981. Order Statistics, Second Edition, New York: John Wiley & Sons, Inc.: 273-276.
- Davidson, R., and Duclos, J.Y. 1998. "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality," Mimeo, Queen's University and Université Laval, Canada.
- Donaldson, D., and Weymark, J.A. 1986. "Properties of Fixed Population Poverty Indices," International Economic Review, 27: 667-688.
- Efron, B. 1979. "Bootstrap Methods: Another Look at the Jackknife," Annals of Statistics, 7: 1-26.
- Efron, B. 1982. The Jackknife, the Bootstrap and Other Resampling Plans, Society for Industrial and Applied Mathematics.
- Efron, B. and Tibshirani, R. 1986. "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy," Statistical Science, 1: 54-77.
- Fisher, G.M. 1992. "The Development and History of the Poverty Thresholds," Social Security Bulletin, 55(4), Winter 1992: 3-14.

- Foster, J.E. 1984. "On Economic Poverty: A Survey of Aggregate Measures," Advances in Econometrics, 3: 215-251.
- Foster, J.E., Greer, J., and Thorbecke E. 1984. "A Class of Decomposable Poverty Indices," *Econometrica*, 52: 761-766.
- Foster, J.E., and Shorrocks, A.F. 1988. "Poverty Orderings," Econometrica, 56: 173-178.
- Foster, J.E., and Shorrocks, A.F. 1991. "Subgroup Consistent Poverty Indices," *Econometrica*, 59: 687-709.
- Ghosh, J.K. 1971. "A New Proof of the Bahadur Representation of Quantiles and an Application," Annals of Mathematical Statistics, 42: 1957-1961.
- Hagenaars, A.J.M. 1986. The Perception of Poverty. Amsterdam: North Holland.
- Kakwani, N. 1980. Income Inequality and Poverty: Methods of Estimation and Policy Applications. New York: Oxford University Press.
- Osberg, L., and Xu, K. 1997. "International Comparisons of Poverty Intensity: Index Decomposition and Bootstrap Inference," Department of Economics, Working Paper 97-03, Dalhousie University, Canada.
- Osberg, L., and Xu, K. 1998. "Poverty Intensity How Well Do Canadian Provinces Compare?," forthcoming in Canadian Public Policy.
- Preston, I. 1995. "Sampling Distribution of Relative Poverty Statistics," Applied Statistics, 44: 91-99.
- Rongve, I. 1997. "Statistical Inference for Poverty Indices with Fixed Poverty Lines," Applied Economics, 29: 387-392.
- Seidl, C. 1988. "Poverty Measurement: A Survey" in D. Bos, M. Rose, and C. Seidl (eds.) Welfare and Efficiency in Public Economics. Berlin: Springer-Verlag.
- Sen, A.K. 1976. "Poverty: An Ordinal Approach to Measurement," Econometrica, 44: 219-231.
- Shorrocks, A.F. 1995. "Revisiting the Sen Poverty Index," Econometrica, 63: 1225-1230.
- Silverman, B.W. 1986. Density Estimation for Statistics and Data Analysis. New York: Chapman and Hall.
- Smeeding, T.M. 1991. "Cross-National Comparisons of Inequality and Poverty Position," in L. Osberg (ed.) Economic Inequality and Poverty: International Perspective. New York: M. E. Sharpe, Inc.
- Stigler, S.M. 1969. "Linear Functions of Order Statistics," Annals of Mathematical Statistics, 40: 770-788.
- Stigler, S.M. 1974. "Linear Functions of Order Statistics with Smooth Weight Functions," *Annals of Statistics*, 2: 676-693. "Correction", 7: 466.
- Takayama, N. 1979. "Poverty, Income Inequality, and Their Measures: Professor Sen's Axiomatic Approach Reconsidered," *Econometrica*, 47: 747-759.
- Thon, D. 1979. "On Measuring Poverty," Review of Income and Wealth, 25: 429-439.
- Thon, D. 1983. "A Poverty Measure," The Indian Economic Journal, 30: 55-70.
- Vaughan, D.R. 1993. "Exploring the Use of the Public's Views to Set Income Poverty Thresholds and Adjust Them over Time," Social Security Bulletin, 56(2), Summer 1993: 22-46.
- Zheng, B. 1997. "Aggregate Poverty Measures," Journal of Economic Surveys, 11(2): 123-161.
- Zheng, B., Cushing, B.J., and Chow, K.V. 1995. "Statistical Tests of Changes in U.S. Poverty, 1975 to 1990," Southern Economic Journal, 62: 334-347.