



Abul Naga, R., 2009. Statistical Inference for Multidimensional Inequality Indices. Working Paper. Bath, UK: Department of Economics, University of Bath, (2/09).

Link to official URL (if available):

<http://www.bath.ac.uk/economics/research/workingpapers.html>

Opus: University of Bath Online Publication Store

<http://opus.bath.ac.uk/>

This version is made available in accordance with publisher policies.
Please cite only the published version using the reference above.

See <http://opus.bath.ac.uk/> for usage policies.

Please scroll down to view the document.

Statistical Inference for Multidimensional Inequality Indices

Ramses H. Abul Naga

No. 2/09

BATH ECONOMICS RESEARCH PAPERS

Department of Economics and International Development



Statistical Inference for Multidimensional Inequality Indices

Ramses H. Abul Naga *

August 29, 2008

Abstract

We use the delta method to derive the large sample distribution of multidimensional inequality indices. We also present a simple method for computing standard errors and obtain explicit formulas in the context of two families of indices.

Keywords: multidimensional inequality indices, large sample distributions, standard errors.

JEL codes: D63, C43

*Department of Economics and International Development (University of Bath). I am grateful to the Swiss National Science Foundation for financial support under grant number 100012-108208. I also wish to thank Pierre-Yves Geoffard for discussions. I am solely responsible for any errors or omissions.

Ever since the work of Kolm (1977), there has been a growing theoretical literature on the measurement of multidimensional inequality (e.g. Tsui 1995, 1999; Abul Naga and Geoffard 2006, Gadjos and Weymark, 2006). Empirical work in the area (for instance Justino et al. 2004) is however considerably more scarce. One reason perhaps is that simple methods for undertaking basic inference in relation to multidimensional inequality indices are lacking¹. In this paper therefore we build on the results of Cowell (1999) obtained in the single attribute context, in order to derive the large sample distribution of multidimensional inequality indices. We also present a simple method for computing standard errors and obtain explicit formulas in the context of two families of indices.

1. Framework

Consider data on p attributes of well-being in relation to a sample of n individuals. Individual i has resources $x_i \doteq (x_{i1}, \dots, x_{ip})$, where $x_i \in \mathbb{R}_{++}^p$. The joint distribution

is a matrix $X \doteq \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in M^n$, the set of all $n \times p$ matrices with strictly positive elements. We also let $\bar{x} \doteq (\bar{x}_1, \dots, \bar{x}_p)$ denote the vector of sample means.

A p -dimensional inequality index is a real valued function $\hat{I}(X) : M^n \rightarrow \mathbb{R}_+$. A difficulty in developing inference for inequality indices noted previously in the unidimensional context (Cowell, 1999) is that the inequality index will rarely present itself in the form of a linear in X function.

For example, the multivariate generalization of the class of Atkinson-Kolm-Sen [AKS] indices proposed by Tsui (1995) takes the following form in the context of two attributes:

$$I^{AKS}(X) \doteq 1 - \left(\frac{1}{n\bar{x}_1^\alpha \bar{x}_2^\beta} \sum_{i=1}^n x_{i1}^\alpha x_{i2}^\beta \right)^{\frac{1}{\alpha+\beta}} \quad (1.1)$$

where α and β are parameter values chosen by the data analyst in order for the index to satisfy certain ethical properties². From the statistician's perspective

¹An important exception is Anderson (2004) which derives tests of inequality dominance in a multidimensional context.

²For instance, for the index $I(\cdot)$ to satisfy a property of aversion to correlation increasing transformations of X , α and β must be chosen to be strictly negative real numbers.

the inequality index is a transformation of three sample moments of X , namely, \bar{x}_1 , \bar{x}_2 and $s \doteq \frac{1}{n} \sum_{i=1}^n x_{i1}^\alpha x_{i2}^\beta$.

More generally, in the context of p attributes \bar{x} is a vector of p sample means, and we can let s be a $1 \times l$ vector of functions:

$$s \doteq \left[\frac{1}{n} \sum_{i=1}^n g_1(x_{i1}, \dots, x_{ip}) \quad \cdots \quad \frac{1}{n} \sum_{i=1}^n g_l(x_{i1}, \dots, x_{ip}) \right] \quad (1.2)$$

Let $k \doteq p + l$ denote the dimension of the vector of moments $w \doteq [\bar{x} \ s]$. We can now express a p -dimensional inequality index $\hat{I}(X)$ as a function $F(\bar{x}, s)$, where $F: \mathbb{R}_{++}^k \rightarrow \mathbb{R}_+$.

2. Large sample distribution

We know from the Analogy Principle that under random sampling the inequality index calculated in the data $\hat{I}(X)$ will result in a consistent estimator of the level of inequality underlying a given population.

Define σ as the population counterpart to s :

$$\sigma \doteq \left[E[g_1(x_i)] \quad \cdots \quad E[g_l(x_i)] \right] \quad (2.1)$$

Let $\mu \doteq (\mu_1, \dots, \mu_p)$ denote the vector of population means and let $\omega \doteq [\mu \ \sigma]$. We also let $cov(y)$ denote the covariance matrix of a random vector y . Finally, we define the $1 \times k$ Jacobian vector J as follows

$$J \doteq \frac{\partial F}{\partial \omega} = \left[\partial F / \partial \mu \quad \partial F / \partial \sigma \right]$$

In deriving our main result, we shall make the following assumptions:

[A1] The observations x_1, \dots, x_n are independently and identically distributed with mean vector μ_o and cross-moment vector σ_o where μ_o and σ_o are finite p -dimensional and l -dimensional vectors.

[A2] $cov(n^{1/2}w) = V_o$, where V_o is a finite $k \times k$ positive semi-definite matrix.

[A3] The function $F(\mu, \sigma)$ does not involve n and is continuously differentiable at the point $\omega_o \doteq [\mu_o \ \sigma_o]$.

Our purpose here is to obtain the large sample distribution of $\hat{I}(X) = F(\bar{x}, s)$ as a function of $I_o \doteq F(\mu_o, \sigma_o)$. For this purpose, we shall use the *delta* method (Goldberger, 1991, chapter 9).

Proposition 1 Under Assumptions [A1-A3], $\hat{I}(X) = F(\bar{x}, s)$ converges in distribution to a normal variate:

$$n^{1/2}[F(\bar{x}, s) - F(\mu_o, \sigma_o)] \longrightarrow N(0, J_o V_o J_o') \quad (2.2)$$

where J_o is the Jacobian vector given by

$$J_o \doteq \frac{\partial F}{\partial \omega} \Big|_{\omega=\omega_o}$$

Proof From [A1], the law of large numbers implies that $plim(\bar{x}) = \mu_o$ and $plim(s) = \sigma_o$. In turn, in relation to [A2] the central limit theorem entails that $n^{1/2}[\omega - \omega_o] \longrightarrow N(0, V_o)$. Hence, using [A3] the delta method entails that $n^{1/2}[\hat{I}(X) - I_o] = n^{1/2}[F(\bar{x}, s) - F(\mu_o, \sigma_o)]$ converges to a normal distribution with mean zero and variance $J_o V_o J_o'$ where $J_o = J(\omega_o)$ \square .

3. Standard errors

Let \hat{V} be any consistent estimator of V_o and let $\hat{J} \doteq J(\bar{x}, s)$. In the light of (2.2), we can then obtain an asymptotic standard error for $\hat{I}(X)$ as *s.e.* $[\hat{I}(X)] \doteq (\hat{J} \hat{V} \hat{J}')^{1/2}$. If we define \bar{g}_j as the j th component of s , i.e. $\bar{g}_j = \sum_{i=1}^n g_j(x_i)/n$ and we construct the $n \times k$ matrix

$$Z = \begin{bmatrix} x_{11} - \bar{x}_1 & \cdots & x_{1p} - \bar{x}_p & g_1(x_1) - \bar{g}_1 & \cdots & g_l(x_1) - \bar{g}_l \\ \vdots & & \vdots & \vdots & & \vdots \\ x_{n1} - \bar{x}_1 & \cdots & x_{np} - \bar{x}_p & g_1(x_n) - \bar{g}_1 & \cdots & g_l(x_n) - \bar{g}_l \end{bmatrix} \quad (3.1)$$

then this asymptotic standard error may be computed as follows:

$$s.e. [\hat{I}(X)] = \left(\frac{1}{n} \hat{J} Z' Z \hat{J}' \right)^{1/2} \quad (3.2)$$

where, in this context, $\hat{V} = Z' Z/n$. The formula for \hat{V} is fairly general. However, the form of the Jacobian vector \hat{J} will vary depending on the structure of the inequality index.

3.1. Indices based on the Cobb-Douglas form

We now return to the simple context of two attributes ³ and consider first inequality indices of the form

$$\hat{I}(X) \doteq F \left(\frac{1}{n\bar{x}_1^\alpha \bar{x}_2^\beta} \sum_{i=1}^n x_{i1}^\alpha x_{i2}^\beta \right) \quad (3.3)$$

The index (1.1) proposed by Tsui (1995) is one such function with $F(t) = 1 - t^{1/(\alpha+\beta)}$. The multidimensional generalized entropy index (Tsui, 1999) also has the form (3.3) with $F(t) = \eta t$, for some scalar η . As a corollary to Proposition 1, we derive the form of the Jacobian vector in the context of (3.3).

Corollary 2 *For inequality indices in the class (3.3), the Jacobian vector takes the form*

$$\hat{J} = F'(t) \cdot \left[-\alpha t / \bar{x}_1 \quad -\beta t / \bar{x}_2 \quad 1 / \bar{x}_1^\alpha \bar{x}_2^\beta \right] \quad (3.4)$$

where $t \doteq \frac{1}{n\bar{x}_1^\alpha \bar{x}_2^\beta} \sum_{i=1}^n x_{i1}^\alpha x_{i2}^\beta$.

In the context of (1.1) for instance, the standard error of the inequality index may readily be estimated using the formula (3.2) with:

$$F'(t) = -\frac{1}{\alpha + \beta} t^{\frac{1}{\alpha+\beta}-1} \quad (3.5)$$

$$Z = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & x_{11}^\alpha x_{12}^\beta - \frac{1}{n} \sum_{i=1}^n x_{i1}^\alpha x_{i2}^\beta \\ \vdots & \vdots & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & x_{n1}^\alpha x_{n2}^\beta - \frac{1}{n} \sum_{i=1}^n x_{i1}^\alpha x_{i2}^\beta \end{bmatrix} \quad (3.6)$$

In the context of the Tsui (1999) index, the standard error formula is as above with the only difference that $F' = \eta$ replaces (3.5) as the index is linear in t .

3.2. Indices which are correlation neutral

Indices which are entirely characterized by the marginal distributions of the two attributes are by definition neutral with respect to correlation increasing transformations of the joint distribution X . Such indices are of the form

$$\hat{I}(X) \doteq F \left(\frac{1}{n} \sum_{i=1}^n \phi_1(x_{i1}), \frac{1}{n} \sum_{i=1}^n \phi_2(x_{i2}), \phi_1(\bar{x}_1), \phi_2(\bar{x}_2) \right) \quad (3.7)$$

³The generalization to $p > 2$ attributes is straightforward given our earlier discussion.

Members of this important class include the bivariate generalization of the Gini index (Gadjos and Weymark, 2006) as well as the bivariate mean logarithmic deviation:

$$\hat{I}^{MLD}(X) \doteq 1 - \exp \left(\frac{\delta}{n} \sum_{i=1}^n \log(x_{i1}) + \frac{1-\delta}{n} \sum_{i=1}^n \log(x_{i2}) - \delta \log(\bar{x}_1) - (1-\delta) \log(\bar{x}_2) \right) \quad (3.8)$$

with $0 < \delta < 1$. There are two moments here other than the two sample means, so we can define $s_1 \doteq \sum_{i=1}^n \log(x_{i1})/n$ and $s_2 \doteq \sum_{i=1}^n \log(x_{i2})/n$.

We can readily obtain the asymptotic standard error of the inequality index as follows:

Corollary 3 *For the bivariate mean logarithmic deviation measure (3.8), the Jacobian vector takes the form*

$$\hat{J} = -\exp(t) \cdot \left[-\delta/\bar{x}_1 \quad -(1-\delta)/\bar{x}_2 \quad \delta \quad (1-\delta) \right] \quad (3.9)$$

where

$$t \doteq \frac{\delta}{n} \sum_{i=1}^n \log(x_{i1}) + \frac{1-\delta}{n} \sum_{i=1}^n \log(x_{i2}) - \delta \log(\bar{x}_1) - (1-\delta) \log(\bar{x}_2) \quad (3.10)$$

The standard error of the inequality index may readily be estimated using the formula (3.2) with:

$$Z = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \log(x_{11}) - \frac{1}{n} \sum_{i=1}^n \log(x_{i1}) & \log(x_{12}) - \frac{1}{n} \sum_{i=1}^n \log(x_{i2}) \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \log(x_{n1}) - \frac{1}{n} \sum_{i=1}^n \log(x_{i1}) & \log(x_{n2}) - \frac{1}{n} \sum_{i=1}^n \log(x_{i2}) \end{bmatrix} \quad (3.11)$$

4. References

Abul Naga R. and P.-Y. Geoffard (2006): "Decomposition of Bivariate Inequality Indices by Attributes", *Economics Letters* 90, 362-367.

Anderson G. (2004): "Indices and Tests for Multidimensional Inequality: Multivariate Generalizations of the Gini Coefficient and Kolmogorov-Smirnov Two Sample Test", Department of Economics, University of Toronto.

Cowell, F. (1999) "Estimation of inequality indices" in Silber, J. (ed.) *Handbook on Income Inequality Measurement*, Kluwer, Dordrecht.

Gajdos, T. and J. Weymark (2006): "Multidimensional generalized Gini indices", *Economic Theory* 26, 471-496.

Goldberger A. (1991): *A course in Econometrics*, Harvard University Press.

Justino P., J. Litchfield and Y. Niimi (2004): "Multidimensional Inequality: An empirical Application to Brazil", Working Paper 24, Poverty Research Unit at Sussex, University of Sussex.

Kolm S. (1977): "Multidimensional Egalitarianisms", *Quarterly Journal of Economics* 91, 1-13.

Tsui K-Y (1995): "Multidimensional Generalizations of the Relative and Absolute Inequality Indices: The Atkinson-Kolm-Sen Approach", *Journal of Economic Theory* 67, 251-265.

Tsui K-Y (1999): "Multidimensional Inequality and Multidimensional Entropy Measures: An Axiomatic Derivation", *Social Choice and Welfare* 16, 145-157.