

THE STATA JOURNAL

Editors

H. JOSEPH NEWTON
Department of Statistics
Texas A&M University
College Station, Texas
editors@stata-journal.com

NICHOLAS J. COX
Department of Geography
Durham University
Durham, UK
editors@stata-journal.com

Associate Editors

CHRISTOPHER F. BAUM, Boston College
NATHANIEL BECK, New York University
RINO BELLOCCO, Karolinska Institutet, Sweden, and
University of Milano-Bicocca, Italy
MAARTEN L. BUIS, University of Konstanz, Germany
A. COLIN CAMERON, University of California–Davis
MARIO A. CLEVES, University of Arkansas for
Medical Sciences
WILLIAM D. DUPONT, Vanderbilt University
PHILIP ENDER, University of California–Los Angeles
DAVID EPSTEIN, Columbia University
ALLAN GREGORY, Queen's University
JAMES HARDIN, University of South Carolina
BEN JANN, University of Bern, Switzerland
STEPHEN JENKINS, London School of Economics and
Political Science
ULRICH KOHLER, University of Potsdam, Germany

FRAUKE KREUTER, Univ. of Maryland–College Park
PETER A. LACHENBRUCH, Oregon State University
JENS LAURITSEN, Odense University Hospital
STANLEY LEMESHOW, Ohio State University
J. SCOTT LONG, Indiana University
ROGER NEWSON, Imperial College, London
AUSTIN NICHOLS, Urban Institute, Washington DC
MARCELLO PAGANO, Harvard School of Public Health
SOPHIA RABE-HESKETH, Univ. of California–Berkeley
J. PATRICK ROYSTON, MRC Clinical Trials Unit,
London
PHILIP RYAN, University of Adelaide
MARK E. SCHAFFER, Heriot-Watt Univ., Edinburgh
JEROEN WEESIE, Utrecht University
IAN WHITE, MRC Biostatistics Unit, Cambridge
NICHOLAS J. G. WINTER, University of Virginia
JEFFREY WOOLDRIDGE, Michigan State University

Stata Press Editorial Manager

LISA GILMORE

Stata Press Copy Editors

DAVID CULWELL, SHELBI SEINER, and DEIRDRE SKAGGS

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go “beyond the Stata manual” in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*), *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

<http://www.stata-journal.com>

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

<http://www.stata.com/bookstore/sj.html>

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		Printed & electronic	
1-year subscription	\$115	1-year subscription	\$145
2-year subscription	\$210	2-year subscription	\$270
3-year subscription	\$285	3-year subscription	\$375
1-year student subscription	\$ 85	1-year student subscription	\$115
1-year institutional subscription	\$345	1-year institutional subscription	\$375
2-year institutional subscription	\$625	2-year institutional subscription	\$685
3-year institutional subscription	\$875	3-year institutional subscription	\$965
Electronic only		Electronic only	
1-year subscription	\$ 85	1-year subscription	\$ 85
2-year subscription	\$155	2-year subscription	\$155
3-year subscription	\$215	3-year subscription	\$215
1-year student subscription	\$ 55	1-year student subscription	\$ 55

Back issues of the *Stata Journal* may be ordered online at

<http://www.stata.com/bookstore/sjj.html>

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

<http://www.stata-journal.com/archives.html>

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.



Copyright © 2016 by StataCorp LP

Copyright Statement: The *Stata Journal* and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp LP. The contents of the supporting files (programs, datasets, and help files) may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

The articles appearing in the *Stata Journal* may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

Written permission must be obtained from StataCorp if you wish to make electronic copies of the insertions. This precludes placing electronic copies of the *Stata Journal*, in whole or in part, on publicly accessible websites, file servers, or other locations where the copy may be accessed by anyone other than the subscriber.

Users of any of the software, ideas, data, or other materials published in the *Stata Journal* or the supporting files understand that such use is made without warranty of any kind, by either the *Stata Journal*, the author, or StataCorp. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the *Stata Journal* is to promote free communication among Stata users.

The *Stata Journal* (ISSN 1536-867X) is a publication of Stata Press. Stata, **Stata**, Stata Press, Mata, **Mata**, and NetCourse are registered trademarks of StataCorp LP.

conindex: Estimation of concentration indices

Owen O'Donnell

Erasmus School of Economics
Erasmus University Rotterdam, the Netherlands
Tinbergen Institute, the Netherlands
and University of Macedonia, Greece

Stephen O'Neill

Department of Health Services Research and Policy
London School of Hygiene and Tropical Medicine, UK
stephen.oneill@lshtm.ac.uk

Tom Van Ourti

Erasmus School of Economics
Erasmus University Rotterdam, the Netherlands
and Tinbergen Institute, the Netherlands

Brendan Walsh

Division of Health Services Research and Management
School of Health Sciences
and City Health Economics Centre
City University London, UK

Abstract. Concentration indices are frequently used to measure inequality in one variable over the distribution of another. Most commonly, they are applied to the measurement of socioeconomic-related inequality in health. We introduce the user-written command `conindex`, which provides point estimates and standard errors of a range of concentration indices. The command also graphs concentration curves (and Lorenz curves) and performs statistical inference for the comparison of inequality between groups. We offer an accessible introduction to the various concentration indices that have been proposed to suit different measurement scales and ethical responses to inequality. We also demonstrate the command's capabilities and syntax by analyzing wealth-related inequality in health and health care in Cambodia.

Keywords: `st0427`, `conindex`, inequality, rank-dependent indices, concentration index, health, health care

1 Introduction

Concentration indices measure inequality in one variable over the distribution of another variable (Kakwani 1977). They are a particularly popular choice for the measurement of socioeconomic-related health inequality (Wagstaff, Paci, and van Doorslaer

1991; O'Donnell et al. 2008), as is evident from the 9,220 entries in Google Scholar with the keywords “concentration index” and “health”. In that case, the concentration index captures the extent to which health differs across individuals ranked by some indicator of socioeconomic status. A variety of concentration indices have been proposed to suit the measurement properties of the variable in which inequality is to be assessed and the assessor's ethical response to inequality (Wagstaff, Paci, and van Doorslaer 1991; Wagstaff 2002; Wagstaff 2005; Erreygers 2009b; Erreygers and Van Ourti 2011a,b; Erreygers, Clarke, and Van Ourti 2012).

We introduce the `conindex` command, which provides a simple unified means to estimate various concentration indices and their standard errors. It can graph the concentration curves that underlie some of the indices and test for differences in inequality across groups. It can also measure cross-sectional inequality in a cardinal variable over observations ranked by another variable that is at least ordinally measured. With repeated cross-section or panel data, one can use `conindex` to compare inequality across periods. One can also use it to estimate rank-dependent indices of univariate inequality, such as the Gini and generalized Gini.

Other user-written commands are available to calculate some rank-dependent inequality indices. `concindc` (Chen 2007) computes the most standard version of the concentration index for both individual and grouped data. The Lorenz curve and associated indices of univariate inequality can be computed and decomposed with a range of commands: `glcurve` (van Kerm and Jenkins 2007), `ineqerr` (Jolliffe and Krushelnytsky 1999), `ineqdeco` (Jenkins 1999), and `descogini` (Lopez-Feldman 2005). The comparative advantage of `conindex` is that it estimates a battery of concentration indices, which allows an analyst to select an index that fits the measurement properties of the variable of interest and is consistent with their normative principles concerning inequality. The indices are estimated by using the correspondence of each to a transformation of the covariance between the variable in which inequality is measured and the rank in the distribution over which inequality is assessed. This so-called convenient covariance approach (Kakwani 1980; Jenkins 1988; Kakwani, Wagstaff, and van Doorslaer 1997) can be implemented with both individual and grouped data while taking account of the sample design.

Before explaining the command, we define the various inequality indices it can compute and offer some guidance regarding the context in which each index is suitable. We illustrate the command's features by analyzing wealth-related inequality in health and health care in Cambodia.

2 Standard and generalized concentration indices

The concentration curve is the bivariate analogue of the Lorenz curve. It plots the cumulative proportion of one variable against the cumulative proportion of the population ranked by another variable. To facilitate more concise exposition, we will mostly refer to the variable of interest as `health` and the ranking variable as `income`. Income-related health inequality can be assessed by plotting the cumulative proportion of health across

individuals ranked from poorest to richest. Unlike the Lorenz curve, the concentration curve may lie above the 45° line if health, or more likely a measure of ill health, is more heavily concentrated among those with lower incomes (as in the hypothetical example in figure 1). The concentration index is twice the area between the concentration curve and the 45° line, indicating no relationship between the two variables.¹ It is defined as

$$C(h|y) = \frac{2\text{cov}(h_i, R_i)}{\bar{h}} = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{h_i}{\bar{h}} (2R_i - 1) \right\} \quad (1)$$

where h_i is the health variable in which inequality is measured, for example, health.²

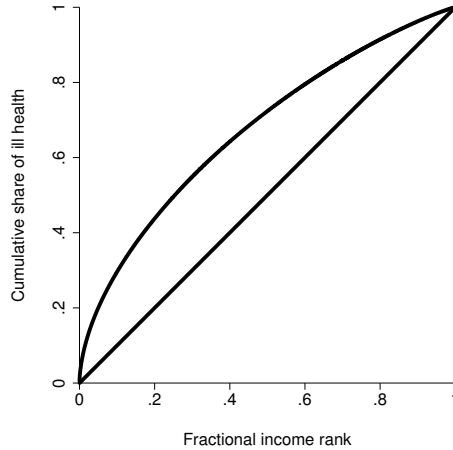


Figure 1. Hypothetical concentration curve

C ranges from $(1 - n)/n$, maximal pro-poor inequality (that is, all health is concentrated on the poorest individual), to $(n - 1)/n$, maximal pro-rich inequality.³

Equation (1) reveals that the concentration index can be interpreted as a weighted mean of (health) shares with the weights depending on the fractional (income) rank $(2R_i - 1)$.⁴ The Gini coefficient measure of univariate inequality arises as a special case of the concentration index when inequality is measured in the same variable that is used

1. A concentration index of 0 can arise either because health does not vary with income rank or because the concentration curve crosses the 45° line and pro-poor inequality in one part of the income distribution is exactly offset by pro-rich inequality in another part of the distribution.
2. The fractional rank varies between $1/2n$ and $1 - (1/2n)$ if there are no ties. In the case of ties, it equals the mean fractional rank of those individuals with the same value for y_i .
3. If given a welfare interpretation, then C respects the principle of income-related health transfers—social welfare falls when the health of a lower-income individual is reduced, and the health of a higher-income individual is raised by the same magnitude (Bleichrodt and van Doorslaer 2006).
4. C respects the principle of income-related health transfers. This is analogous to the principle of transfers for G and requires that a health transfer from a low- to a high-income individual will lower social welfare (Bleichrodt and van Doorslaer 2006).

for ranking. This is true for all indices we discuss in the remainder of this article and implies that **conindex** can be used to estimate univariate inequality.

The concentration index measures relative inequality and is invariant to equiproportionate changes in the variable of interest (health). This relative invariance is one extreme of the many normative positions one might take in measuring inequality (Kolm 1976). At the other extreme, absolute invariance corresponds to an inequality measure that is invariant to equal additions to health. Such a measure can be obtained through multiplication of the standard concentration index by the mean health leading to the generalized concentration index (Wagstaff, Paci, and van Doorslaer 1991).⁵ Multiplication by the mean gives this parameter an important role in the assessment of absolute inequality. When two distributions display the same level of relative inequality, the one with the higher mean will correspond to greater absolute inequality. The generalized concentration index GC can be expressed as

$$GC(h|y) = \frac{1}{n} \sum_{i=1}^n \{h_i(2R_i - 1)\}$$

and ranges between $\bar{h} \{(1 - n)/n\}$ (maximal pro-poor) and $\bar{h} \{(n - 1)/n\}$ (maximal pro-rich).

3 Taking account of measurement scale

The standard and generalized concentration indices are not necessarily invariant, or equivariant, under transformations of the variable of interest that are permissible for the level of measurement (that is, nominal, ordinal, cardinal, ratio, or fixed scale) (Erreygers and Van Ourti 2011a,b).⁶ Several variants of the standard and generalized concentration indices have been proposed for use with variables possessing different measurement properties. We differentiate between measurement levels at which permitted transformations affect the value of an index and levels at which transformations to different scales affect inequality orderings. Both have received attention (for example, Lambert and Zheng [2011]), but most applications focus on the former. We think the latter issue is more important because it deals with whether one bivariate distribution is evaluated to display greater inequality than another, irrespective of an arbitrary scaling.

5. The graphical representation of the generalized concentration index corresponding to figure 1 is the generalized concentration curve. According to the generalized concentration dominance criterion (Shorrocks 1983), a distribution with a higher mean cannot be dominated by one with a lower mean. But the ordering of two distributions by their generalized concentration indices does not necessarily correspond to their ordering by their means.

6. A function $f(\cdot)$ is invariant under transformation $g(\cdot)$ if $f\{g(x)\} = f(x)$. A function is equivariant if $f\{g(x)\} = |\partial g/\partial x| f(x)$.

3.1 Measurement level

In bivariate inequality measurement, an ordinal scale is sufficient for the variable that is used for the ranking of individuals. Rank-dependent indices can then be deployed to quantify inequality in variables measured at three levels:⁷

- Fixed: the measurement scale is unique (or fixed) with zero point corresponding to a situation of complete absence, for example, number of visits to a hospital within a given period.
- Ratio: the measurement scale is unique up to a proportional scaling factor with the zero point corresponding to a situation of complete absence, for example, life expectancy that could be measured in years, months, etc.
- Cardinal: the measurement scale is such that differences between values are meaningful but ratios are not, and the zero point is fixed arbitrarily, for example, temperature in Celsius or Fahrenheit or a health utility index.

For variables on a fixed scale, the standard and generalized concentration indices quantify inequality in the attribute of fundamental interest. Both are appropriate, with the choice between them depending on whether one is concerned about relative or absolute inequality. Changing the proportionality factor of a ratio-scaled variable will affect the value of the generalized concentration index but not that of the standard concentration index.⁸ The generalized concentration index should therefore be used with ratio-scaled data only when the variables compared in an inequality ordering are subject to the same scaling factor.⁹ Only in this case can one be sure that the inequality ordering given by the index applied to the variable is informative of the ranking of populations by inequality in the attribute of essential interest. Alternatively, because the generalized concentration index is equivariant under a proportional transformation of the variable, if one knows the differential scaling factors, then one can use them *ex post* to make the indices comparable across populations.

When the variable of interest is cardinal, the standard concentration index is not necessarily invariant to arbitrary retransformations of the variable.¹⁰ One can address this by using the modified concentration index MC (Erreygers and Van Ourti 2011a,b),

$$\text{MC}(h|y) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{h_i}{\bar{h} - h^{\min}} (2R_i - 1) \right\} \quad (2)$$

where h^{\min} is the lower limit of h_i and the index ranges between $(1 - n)/n$ and $(n - 1)/n$. Under ratio- or fixed-measurement scales ($h^{\min} = 0$), (2) simplifies to the standard concentration index in (1).

7. Measuring inequality in nominal and ordinal variables is not feasible using rank-dependent indices (Erreygers and Van Ourti 2011a,b).

8. Assuming that $h_i = \beta x_i$, one obtains $C(h|y) = 2\text{cov}(h_i/\bar{h}, R_i) = 2\text{cov}(\beta x_i/\beta \bar{x}, R_i) = C(x|y)$ and $\text{GC}(h|y) = 2\text{cov}(h_i, R_i) = 2\text{cov}(\beta x_i, R_i) = \beta \text{GC}(x|y)$.

9. We assume monotone transformations; hence, the proportional scaling factor is positive.

10. We restrict attention to positive linear transformations; that is, β must be positive in $h_i = \alpha + \beta x_i$.

One should use the modified concentration index when comparing inequality in an attribute using a variable that is inconsistently cardinalized for different populations or when comparing inequality in different cardinally scaled variables in the same population. If the cardinalization is constant, then inequality orderings made using the standard concentration index will be robust to the chosen cardinalization (although the index values will depend on the specific cardinalization chosen). Nevertheless, we advise that one also use the modified index in this case because it allows for an easier interpretation—the range is always $[(1 - n)/n, (n - 1)/n]$.

There is no easy modification to ensure the invariance of the generalized concentration index to retransformations of cardinal variables. However, provided the cardinalization adopted across populations or variables is the same, then the inequality ordering will be robust to the chosen cardinalization.

3.2 Bounded variables

Variables with a finite upper limit, such as years in school, a (health) utility index, or any binary indicator, complicate the measurement of inequality.¹¹ For instance, bounded variables can be represented either as attainments $a_i \in [a^{\min}, a^{\max}]$ or as shortfalls from the upper limit $s_i = a^{\max} - a_i$. Erreygers (2009b) introduced the “mirror” property, which requires that the magnitude of measured inequality represented by the absolute value of an index should not depend on whether the index is computed over attainments or shortfalls; that is, $I(a) = -I(s)$.¹²

The standard concentration index does not satisfy this condition, $C(s) = (-\bar{a}/\bar{s})C(a)$; hence, inequality in attainments do not mirror inequality in shortfalls except when $\bar{a} = \bar{s}$ (Erreygers 2009b).¹³ Moreover, inequality orderings based on the standard concentration index might depend on whether one uses shortfalls or attainments. More generally, the mirror condition is incompatible with the measurement of relative inequality (Erreygers and Van Ourti 2011a,b; Lambert and Zheng 2011). One must choose between satisfaction of the mirror condition and satisfaction of relative inequality invariance.¹⁴

The generalized concentration index satisfies the mirror condition, $GC(s) = -GC(a)$. However, as noted in section 2, the value of this index is not invariant to permissible transformations of ratio-scaled and cardinal variables. Erreygers (2009b) proposed a modification of the generalized concentration index that corrects this deficiency:

-
11. For discussion of the issues, particularly in relation to the measurement of health inequality, see Clarke et al. (2002); Wagstaff (2005); Erreygers (2009a,b,c); Wagstaff (2009); Erreygers and Van Ourti (2011a,b); Wagstaff (2011a,b); and Kjellsson and Gerdtham (2013a,b).
 12. Lambert and Zheng (2011) suggested a weaker condition requiring that the inequality ordering of populations by attainments is strictly the reverse of that by shortfalls.
 13. The same holds for the modified concentration index.
 14. Bosmans (Forthcoming) shows how one can overcome the impossibility of satisfying both conditions if one allows for different functional forms for the inequality index for attainments and the inequality index for shortfalls.

$$E(a|y) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{4a_i}{(a^{\max} - a^{\min})} (2R_i - 1) \right\} = -E(s|y) \quad (3)$$

This index ranges between -1 and $+1$.

Wagstaff (2005) noted that the range of the standard concentration index depends on the mean of the bounded variable and suggested rescaling the standard concentration index to ensure that it always lies in the range $[-1, 1]$:¹⁵

$$W(a|y) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{(a^{\max} - a^{\min})a_i}{(a^{\max} - \bar{a})(\bar{a} - a^{\min})} (2R_i - 1) \right\} = -W(s|y)$$

This index satisfies the mirror condition and so cannot be in line with the relative invariance criterion. Neither does it satisfy an absolute invariance criterion. In fact, the index is consistent with an inequality invariance condition consisting of a mixture of invariance with respect to proportionate changes in 1) attainments a_i and 2) shortfalls s_i . This may be considered to have paradoxical implications (Erreygers and Van Ourti 2011a,b), although Kjellsson and Gerdtham (2013b) argue that this invariance criterion is actually intuitive when one realizes that W can be written as the difference between the standard concentration indices for attainments and shortfalls.

Unlike for unbounded variables, the precise scaling of bounded variables does not affect the value of any rank-dependent inequality index provided that the bounding is considered. This is most easily understood from the realization that any bounded variable can be retransformed into an indicator of the proportional deviation from the minimum value, $b_i = (a_i - a^{\min})/(a^{\max} - a^{\min})$. This lies on the range $[0, 1]$ and records only “real” changes in the underlying attribute, not “nominal” ones due to the choice of measurement scale. Under this transformation, the Erreygers and Wagstaff indices simplify, respectively, to

$$E(b|y) = (1/n) \sum_{i=1}^n \{4b_i (2R_i - 1)\}$$

and

$$W(b|y) = (1/n) \sum_{i=1}^n [\{b_i/(1 - \bar{b})\} \bar{b}(2R_i - 1)]$$

3.3 Summing up

The main message of this section is not that the most appropriate inequality index depends on the measurement properties of the variable of interest but that those properties partly determine the ethical choices one faces when quantifying inequality, which

15. Wagstaff (2005) focused on the binary case.

is intrinsically a normative exercise. When the variable of interest has an infinite upper bound on a fixed scale, the main normative choice is between absolute and relative invariance. Matters are more complicated when the measurement scale is not unique. Applying the generalized concentration index to a ratio or cardinal variable requires one to accept that the inequality ordering may depend on the scaling adopted. This can be avoided for the relative inequality invariance criterion if one replaces the standard concentration index with the modified one. When the variable has a finite upper bound, one should first choose between relative inequality invariance and the mirror condition. If one prioritizes the relative invariance criterion (in attainments or shortfalls), then the standard concentration index or its modified version can be used. When priority is given to the mirror condition, one faces a choice between the Erreygers index, which focuses on absolute differences, and the Wagstaff index, which mixes concern for relative inequalities in attainments and relative inequalities in shortfalls.

If one considers no index to be normatively superior to all others, then one can check whether the inequality orderings are consistent across indices. If they are, all is well and good. However, such robustness does not hold in general.

4 Incorporating alternative attitudes to inequality

As noted in section 2, the standard concentration index can be interpreted as a weighted mean of the variable of interest with each individual's weight depending on its fractional rank; that is, $(2R_i - 1)$. This weight equals 0 for individuals with the median value of the ranking variable¹⁶ and is negative (positive) for individuals below (above) the median. Presuming the ranking variable is income, the weight increases linearly from $(1 - n)/n$ for the poorest individual to $(n - 1)/n$ for the richest. This linearity is consistent with a particular attitude toward inequality that need not command widespread support. Two extensions based on nonlinear weighting schemes can represent a variety of alternative ethical positions. The first approach makes it possible to vary the weight put on those at the top relative to those at the bottom of the distribution of the ranking variable. We refer to it as "sensitivity to poverty" because it allows more (or less) weight to be placed on the poorest individuals when income is used as the ranking variable. The second approach allows more (or less) weight to be placed on the extremes of the ranking distribution (for example, the very rich and very poor) vis-a-vis those in the middle. We term this approach "sensitivity to extremity".

4.1 Extended concentration index: sensitivity to poverty

Kakwani (1980) and Yitzhaki (1983) proposed a flexible extension of the univariate Gini index that incorporates a distributional sensitivity parameter v specifying the attitude toward inequality within the weight defined by $1 - v(1 - R_i)^{v-1}$. Pereira (1998) and Wagstaff (2002) suggested using the same weighting function in the context of the mea-

16. R_i for this individual is $(2i - 1)/2n = [2\{(n + 1)/2\} - 1]/2n = 1/2$; hence, the weight is $2(1/2) - 1 = 0$.

surement of income-related health inequality. This results in an extended concentration index identical to the standard concentration index except for the weighting function.¹⁷

$$EC(h|y; v) = \frac{1}{n} \sum_{i=1}^n \left[\frac{h_i}{\bar{h}} \left\{ 1 - v(1 - R_i)^{v-1} \right\} \right] \quad (4)$$

The distributional sensitivity parameter must take a value greater than or equal to 1. Larger values place more weight on the poorest individuals (when income is the ranking variable). The weighting function equals 0 for $v = 1$ and in that case gives an index of 0 regardless of the distribution of h , while $v = 2$ yields the standard concentration index. The extended concentration index ranges between $1 - v$ and 1,¹⁸ which suggests an intuitive interpretation of v as the distance between the weight given to the poorest and the richest individual. The weight given to the richest individual is always +1, while the weight given to the poorest individual becomes more negative for higher values of v . Hence, the weighting function in (4) is asymmetric around the individual with median income, unless $v = 2$. For $v \neq 2$, the individual with median income does not have a weight of 0. The individual given a weight of 0 will have a lower income than the one with the median income when $v > 2$.

4.2 Symmetric concentration index: Sensitivity to extremity

Erreygers, Clarke, and Van Ourti (2012) suggest extending the linear weighting function of the concentration index in such a way that two conditions are satisfied: 1) the individual with median income should play a pivotal role, obtaining a weight of zero, and 2) the weights for the other individuals should be inversely symmetric around median income. The poorest and richest individual should have the same weight but with an opposite sign. The second poorest and second richest should have the same weight with opposite signs, and so on. Under these conditions, varying attitudes toward inequality express one's sensitivity to extremity, that is, whether one is concerned merely with differences in the variable of interest at the middle of the income distribution or with differences between the extremes of the income distribution.

An index that satisfies both conditions and allows for varying degrees of sensitivity to extremity depending on the value of $\beta > 1$, which is analogous to the parameter v in the extended index, is¹⁹

-
17. When one uses a finite number of observations to calculate the extended concentration index and $v \neq 2$, a small-sample bias arises. Erreygers, Clarke, and Van Ourti (2012) develop an alternative way to calculate the extended concentration index (and its generalized version) to address this small-sample bias. Their approach is applied in the `conindex` command. We refer the reader to the appendix in Erreygers, Clarke, and Van Ourti (2012) for more details.
 18. EC ranges between $1 - v$ and 1 when $n \rightarrow +\infty$. For a finite value of n , the lower and upper limit of EC are $1 - v \{(2n - 1)/2n\}^{v-1}$ and $1 - v(1/2n)^{v-1}$.
 19. As with the extended concentration index, a small-sample bias arises when $\beta \neq 2$ (see footnote 17). We have implemented the approach explained in the appendix of Erreygers, Clarke, and Van Ourti (2012) in the `conindex` command. Note that the approach is also used for the generalized version.

$$S(h|y; \beta) = \frac{1}{n} \sum_{i=1}^n \left(\frac{h_i}{\bar{h}} \right) \left[\beta 2^{\beta-2} \left\{ \left(R_i - \frac{1}{2} \right)^2 \right\}^{\frac{\beta-2}{2}} \left(R_i - \frac{1}{2} \right) \right] \quad (5)$$

If $1 < \beta < 2$, more weight is placed on the middle of the income distribution, whereas for $\beta > 2$, the extremes are weighted more at the expense of the middle. If $\beta = 2$, the symmetric index equals the standard concentration index. When β becomes very large, the symmetric index will be very similar to the range index, which is sensitive only to the difference between the upper and lower end of the income distribution. This corresponds to one of the earliest measures of health inequality (for example, Townsend and Davidson [1982]). The range of the symmetric index is $[-(\beta/2), +(\beta/2)]$, which provides an intuitive interpretation of the β parameter as the absolute deviation between the weight given to the poorest and richest individual.²⁰

The choice between the symmetric and extended indices is normative. The symmetric index gives equal weight (but with an opposite sign) to individuals that are equally far apart from the pivotal individual with median rank, while the extended index prioritizes the lower regions of the ranking (income) distribution. Applied to income-related health inequality, the symmetric index is increasingly sensitive to a change that raises the health of a richer individual and reduces that of a poorer individual by an equal magnitude the further those individuals are from the pivotal individual. In contrast, the extended concentration index will be increasingly sensitive the closer the location of such a “health transfer” to the bottom of the income distribution. Erreygers, Clarke, and Van Ourti (2012) argue that the symmetric index is more concerned about the association between income and health, while the extended concentration index puts priority on the income distribution and only then analyzes health differences within the prioritized region of the income distribution.²¹

4.3 Generalizing the extended and symmetric indices

Erreygers, Clarke, and Van Ourti (2012) consider counterparts of the extended and symmetric indices that satisfy the mirror condition. They refer to the resulting measures as generalized indices because they satisfy an absolute inequality invariance criterion and define these on the transformed bounded variable $b_i = (a_i - a^{\min}) / (a^{\max} - a^{\min})$:

20. This range is entirely correct only when $n \rightarrow +\infty$. For a finite number of observations, the range is $\left[\beta 2^{\beta-2} \left[\{(1-n)/2n\}^2 \right]^{(\beta-2)/2} \{(1-n)/2n\}; \beta 2^{\beta-2} \left[\{(n-1)/2n\}^2 \right]^{(\beta-2)/2} \{(n-1)/2n\} \right]$.

21. Equations (4) and (5) reveal that the symmetric and extended concentration indices consider health shares and hence are sensitive to relative health differences. The “absolute” counterparts of these indices have not explicitly been introduced by Erreygers, Clarke, and Van Ourti (2012) (or Pereira [1998] and Wagstaff [2002] in the case of the extended concentration index) but are trivially derived by replacing the health shares with the health levels. Similarly, the measurement scale of unbounded variables is important for the extended and symmetric indices, but the discussion essentially mimics that in section 3.1. Modifications such as in sections 3.1 and 3.2 can be derived. Because these modifications are not integrated in the `conindex` command, we do not discuss these indices explicitly.

$$\text{GEC}(b|y; v) = \frac{1}{n} \sum_{i=1}^n \left(\frac{v^{\frac{v}{v-1}}}{v-1} b_i \right) \left\{ 1 - v(1 - R_i)^{v-1} \right\} \quad (6)$$

$$\text{GS}(b|y; \beta) = \frac{1}{n} \sum_{i=1}^n 4b_i \left[\beta 2^{\beta-2} \left\{ \left(R_i - \frac{1}{2} \right)^2 \right\}^{\frac{\beta-2}{2}} \left(R_i - \frac{1}{2} \right) \right] \quad (7)$$

with $v \geq 1$ and $\beta > 1$. For $v = \beta = 2$, both indices simplify to the Erreygers index (3). Erreygers, Clarke, and Van Ourti (2012) show that both indices always range between -1 and $+1$.²²

5 Estimation and inference

Each of the rank-dependent inequality indices discussed above can be expressed as a transformation of the covariance between the variable of interest (h_i) and the fractional rank (R_i) of the ordering variable. For example, the standard concentration index is twice the covariance divided by the mean of the variable of interest [see equation (1)]. Because the slope coefficient of a simple least-squares regression is the covariance divided by the variance of the regressor, each inequality index can be obtained from a regression of a transformation of the variable of interest on the rank. For example, the standard concentration index is the least-squares estimate of α_1 in the model

$$\frac{2\sigma_R^2}{\bar{h}} h_i = \alpha_0 + \alpha_1 R_i + \epsilon_i \quad (8)$$

where σ_R^2 is the variance of R and ϵ_i is an error term. The standard error of the least-squares estimate of α_1 serves as a standard error of the estimate of the concentration index.²³

An advantage of this approach is that Stata readily allows for sampling weights as well as robust and clustered standard errors. Appropriate rescalings of the dependent variable lead to the other indices considered in section 3.²⁴ Erreygers, Clarke, and Van

22. Extensions of these indices that simultaneously satisfy the mirror condition and the inequality invariance criterion underlying the Wagstaff index are not discussed, because these have not been introduced in the literature before. However, in principle, it would be feasible to derive such indices.

23. **conindex** does not consider the sampling variability of the estimate of the mean of the variable of interest used in constructing the dependent variable of the regression. Typically, this makes very little difference to the standard error of an estimated concentration index. The command calculates the population formula for σ_R^2 and not the sampling formula; that is, there is no degrees of freedom correction. The command does not implement the approach of Kakwani, Wagstaff, and van Doorslaer (1997) to account for serial correlation because this approach has not been extended to also allow for sample design issues such as clustering. For more details on these issues, see O'Donnell et al. (2008, chap. 8).

24. One should replace $(2\sigma_R^2)/\bar{h}$ in (8) by $2\sigma_R^2$ for GC, by $(2\sigma_R^2)/(\bar{h} - h^{\min})$ for MC, by $(8\sigma_R^2)/(\bar{a} - a^{\min})$ for E , and by $\{2\sigma_R^2(a^{\max} - a^{\min})\}/\{(a^{\max} - \bar{a})(\bar{a} - a^{\min})\}$ for W .

Ourti (2012) do not provide standard errors for the extended and symmetric indices; therefore, standard errors for these indices are not reported.

A final note concerns ties in the ranking variable that arise when different observations have the same value for the ranking variable. `conindex` accounts for this by calculating the fractional rank from the proportion of individuals with a given value of the ranking variable (y), such that $R_i = (\sum_{i=1}^n sw_i)^{-1} [q(y_i - 1) + 0.5 \{q(y_i) - q(y_i - 1)\}]$, where sw_i denotes the sampling weight of individual i and $q(y_i) = \sum_{k=1}^n 1(y_k \leq y_i) sw_k$ equals the proportion of individuals with at least the value y_i (Van Ourti 2004). While `conindex` automatically adjusts for ties in computing the point estimate, it purposefully does not do so in generating the standard error. This is because two individuals with the same value of the ranking variable may or may not be entirely independent observations. In the former case, one should not correct the standard errors, but if the observations are dependent (because they belong to the same household, for example), then the `cluster()` option should be used.²⁵ The occurrence of ties in the ranking variable is similar to the case of grouped data estimation of the standard concentration index. With grouped data, one row in the data matrix will include group mean values of the variable of interest and the ranking variable as well as the sample weight indicating the relative size of the group. One can apply `conindex` directly to such grouped data.

6 The conindex command

6.1 Syntax

The syntax for `conindex` is

```
conindex varname [if] [in] [weight] [, rankvar(varname) robust
      cluster(varname) truezero generalized bounded limits(#1 #2)
      wagstaff erreygers v(#) beta(#) graph loud compare(varname)
      keeprank(string) ytitle(string) xtitle(string)]
```

`fweights`, `aweight`s, and `pweight`s are allowed; see [U] 11.1.6 **weight**.

by *varlist*: is allowed and can be used to calculate indices for groups defined by multiple variables.

6.2 Description

`conindex` computes a range of rank-dependent inequality indices, including the Gini coefficient, the concentration index, the generalized (Gini) concentration index, the modified concentration index, the Wagstaff and Erreygers normalized concentration indices

25. Where ties occur between observations in different clusters, clustered standard errors may be unstable because they are obtained from a regression at the group level with groups defined by the unique values of the ranking variable.

for bounded variables, and the distributionally sensitive extended and symmetric concentration indices (and their generalized versions). There is no default index. Options define the index to be computed. One can use the **graph** option to obtain (generalized) Lorenz and (generalized) concentration curves. The default axis labels can be replaced with the **xtitle(string)** and **ytitle(string)** options.

For unbounded variables (that is, those with at least one infinite bound), the **truezero** option should be specified if the variable of interest is ratio scale (or fixed) and has a zero lower limit, in which case the standard concentration index is calculated. If instead the variable of interest is cardinal (with the zero point fixed arbitrarily), then the theoretical lower limit must be specified using the **limits(#)** option, where **#** is the minimum value. Note that one should not use the lowest value observed in the sample if this does not correspond to the theoretical lower bound. Specifying this option results in calculation of the modified concentration index.

The generalized concentration index derives from specifying the **generalized** option in conjunction with the **truezero** option.

For bounded variables (that is, those with both a finite lower and upper bound), the **bounded** option can be specified in conjunction with **limits(#1 #2)**, where **#1** and **#2** denote the theoretical minimum and maximum values of the variable of interest. The inequality indices are then calculated based on the standardized version of the variable of interest, $h^* = \{(h - \#_1) / (\#_2 - \#_1)\}$, and hence will be scale invariant.

The normalized concentration indices proposed by Wagstaff (2005) and Erreygers (2009b) may be obtained by specifying the **wagstaff** and **erreygers** options, respectively, in conjunction with the **bounded** and **limits(#1 #2)** options.

When a ranking variable is not provided using the **rankvar** option, **conindex** defaults to use *varname* to rank observations, leading to the calculation of unidimensional inequality indices (for example, the Gini coefficient).

The extended concentration index, which allows for alternative attitudes to inequality (Pereira 1998; Wagstaff 2002), is computed with the **truezero** and **v(#)** options, where **#** is the distributional sensitivity parameter. With **v(2)**, the extended concentration index is equivalent to the standard concentration index.

The symmetric concentration index is obtained with the **truezero** and **beta(#)** options. With **beta(2)**, the symmetric concentration index is equivalent to the standard concentration index.

The generalized version of the extended and symmetric concentration indices are obtained by combining the **v()** and **beta()** options with the **truezero** and **generalized** options.

All indices are calculated using the so-called convenient covariance approach (Kakwani 1980; Jenkins 1988; Kakwani, Wagstaff, and van Doorslaer 1997). Robust- and cluster-corrected standard errors can be obtained with the usual options. Standard errors for the extended and symmetric indices are not calculated by the current version of **conindex**.

The value of an index can be compared across groups defined by a single variable (for example, **urban**), and the null of homogeneity tested using the **compare()** option. The prefix **bysort varlist:** can be used to calculate the indices for groups defined by multiple variables (for example, **urban** and **hhsiz**).

The fractional rank may be preserved using the **keeprank(string)** option, where *string* is the name given to the rank variable created.

6.3 Options

rankvar(varname) specifies the variable by which individuals are ranked. *varname* must be at least an ordinal variable. When a ranking variable is not provided using the **rankvar()** option, **conindex** defaults to using *varname* to rank observations, leading to the calculation of unidimensional inequality indices (for example, the Gini coefficient).

robust requests Huber/White/sandwich standard errors.

cluster(varname) requests clustered standard errors that allow for intragroup correlation.

truezero declares that the variable of interest is ratio scaled (or fixed), leading to computation of the standard concentration index.

generalized requests the generalized concentration (Gini) index, measuring absolute inequality. This option can be used only in conjunction with **truezero**.

bounded specifies that the dependent variable is bounded. This option must be used in conjunction with the **limits()** option.

limits(#1 #2) must be used to specify the theoretical minimum (*#1*) and maximum (*#2*) for bounded variables. If the **bounded** and **truezero** options are not specified, then **limits(#1)** should be used to specify the minimum value to obtain the modified concentration index.

wagstaff in conjunction with **bounded** and **limits(#1 #2)** requests the Wagstaff index.

erreygers in conjunction with **bounded** and **limits(#1 #2)** requests the Erreygers index.

v(#) requests the extended concentration index be computed. This option can be used only in conjunction with **truezero**. With **v(2)**, the standard concentration index is computed. If the **v(#)**, **truezero**, and **generalized** options are specified, one obtains the generalized extended concentration index. In the latter case, with **v(2)**, the extended concentration index simplifies to the Erreygers index.

beta(#) requests the symmetric concentration index be computed. This option can be used only in conjunction with **truezero**. With **beta(2)**, the standard concentration index is computed. If the **beta(#)**, **truezero**, and **generalized** options are speci-

fied, one obtains the generalized symmetric concentration index. In the latter case, with `beta(2)`, the symmetric concentration index leads to the Erreygers index.

`graph` requests that a concentration curve be displayed. If no ranking variable is specified, a Lorenz curve is produced. In conjunction with `generalized`, one obtains the generalized Lorenz or concentration curve.

`loud` shows the output from the regression used to generate the inequality indices.

`compare(varname)` computes indices specific to groups specified by *varname*. Two tests of the null hypothesis of equality of the index values across groups are produced: an *F* test that is valid in small samples but requires an assumption of equal variances across groups (Chow 1960) and a *z* test that relaxes the assumption of equal variances but is valid only in large samples (Clogg, Petkova, and Haritou 1995). If *varname* is not binary, then only the *F* test is given.

`keeprank(string)` creates a new variable that contains the fractional ranks, where *string* is the name of the variable to be created. When used in conjunction with the `compare()` option, the variable *string* will contain the fractional rank for the full sample, and the suffix *k* is added to *string* to indicate the fractional rank for group *k*.

`ytitle(string)` and `xtitle(string)` specify the titles to appear on the *y* and *x* axes, respectively.

6.4 Stored results

`conindex` stores the following in `r()`:

Scalars

<code>r(N)</code>	number of observations
<code>r(Nunique)</code>	number of unique observations for <code>rankvar()</code>
<code>r(CI)</code>	concentration index
<code>r(CIse)</code>	standard error of concentration index
<code>r(SSE_unrestricted)</code>	unrestricted sum of squared errors (with <code>compare()</code> option)
<code>r(SSE_restricted)</code>	restricted sum of squared errors (with <code>compare()</code> option)
<code>r(F)</code>	<i>F</i> statistic for joint hypothesis that concentration index is same for all groups (with <code>compare()</code> option)
<code>r(CI0)</code>	concentration index for group 0 (with <code>compare()</code> option if only two groups)
<code>r(CI1)</code>	concentration index for group 1 (with <code>compare()</code> option if only two groups)
<code>r(CIse0)</code>	standard error of concentration index for group 0 (with <code>compare()</code> option if only two groups)
<code>r(CIse1)</code>	standard error of concentration index for group 1 (with <code>compare()</code> option if only two groups)
<code>r(Diff)</code>	difference in concentration index between groups (with <code>compare()</code> option if only two groups)
<code>r(Diffse)</code>	standard error of difference in concentration index between groups (with <code>compare()</code> option if only two groups)
<code>r(z)</code>	<i>Z</i> statistic for hypothesis that concentration index is same for both groups (with <code>compare()</code> option if only two groups)

7 conindex: Example applications

We illustrate the functionality of `conindex` through examples using data from the 2010 Demographic and Health Survey (DHS) of Cambodia, which can be obtained from <http://www.dhsprogram.com/>. The Cambodian DHS covers a representative sample of women aged between 15 and 49. It asks each participating woman about her pregnancies in the last 10 years and also collects information at the household level. We construct a dataset of households to estimate inequality in the distribution of health care expenditures and a dataset of births to estimate inequality in infant mortality. Inequality in each variable is examined in relation to a wealth index (`wealthindex`) that is obtained from a principal components analysis of the households' possession of a battery of assets and durables as well as housing materials (Filmer and Pritchett 2001). This index has an ordinal interpretation and is used as the ranking variable.

In the household dataset, we construct a measure of health care expenditure per capita (`healthexp`) by summing out-of-pocket medical spending in the last month across individuals within the household and dividing by household size.²⁶ This measure will serve as an example of an unbounded variable with a ratio scale. From the child dataset, we construct a binary indicator of infant mortality (`u1mr`) that indicates whether each child born during the last 10 years survived to its first birthday.²⁷ Results below indicate that around 6% of children die within a year of birth. Average per capita monthly health expenditure is about 12,000 riel (€2.40), but the median value of 0 riel and the high maximum of 14,500,000 riel show that the distribution of health expenditures is right skewed.

```
. summarize healthexp [aweight=sampweight_hh]
```

Variable	Obs	Weight	Mean	Std. Dev.	Min	Max
healthexp	15,667	75391.2524	12010.62	116693	0	1.45e+07

```
. summarize u1mr [aweight=sampweight]
```

Variable	Obs	Weight	Mean	Std. Dev.	Min	Max
u1mr	14,598	14588.3669	.0606938	.238776	0	1

Figure 2 shows that more than 8% of Cambodian children in the lowest wealth quintile group die before they reach their first birthday. This is more than three times greater than the rate of infant mortality in the richest wealth quintile group. The

26. For each ill or injured household member, the respondent was asked to state the costs expended for transportation and treatment for each visit to a health care provider (for up to three visits and without differentiating between outpatient and inpatient care). These costs were reported only for living people who had been ill or injured during the last month and did not include costs incurred for people who had died in the 30 days preceding the interview.

27. For the summary statistics of health care expenditures, this implies that we consider the individual as the ultimate unit of observation, even though expenditures are measured at the household level. For the concentration indices of health expenditures, it also implies that household size will influence the fractional ranks. As shown by Edbert (1997) and illustrated by Decoster and Ooghe (2003) on income data, this has important normative implications in terms of the axioms of anonymity and principle of transfer (both of which would be violated if each household were weighted equally independently of its size). In practice, this means we report estimates based on 15,667 observations, but application of `sampweight_hh` ensures these are representative of 75,391 individuals.

mortality rate declines, but not monotonically, in moving to higher wealth groups. Health expenditure rises from around 7,000 riel per capita in the lowest wealth quintile group to more than 22,500 riel in the top group.

```
. xtile wealthquint_hh = wealthindex [pweight=sampweight_hh], n(5)
. xtile wealthquint = wealthindex [pweight=sampweight], n(5)
. graph bar (mean) healthexp [pweight=sampweight_hh], over(wealthquint_hh)
. graph bar (mean) ulmr [pweight=sampweight], over(wealthquint)
```

Table 1 (see page 132) summarizes all the indices discussed below. The concentration index of 0.248 confirms that medical spending is heavily concentrated among better-off sample households identified by a higher position in the wealth index distribution. As well as the point estimate, `conindex` returns a cluster-adjusted standard error and a p -value for a test that the index equals 0. In this example, the null is strongly rejected ($p < 0.001$). `conindex` can be used to graph a concentration curve by adding the `graph` option.²⁸

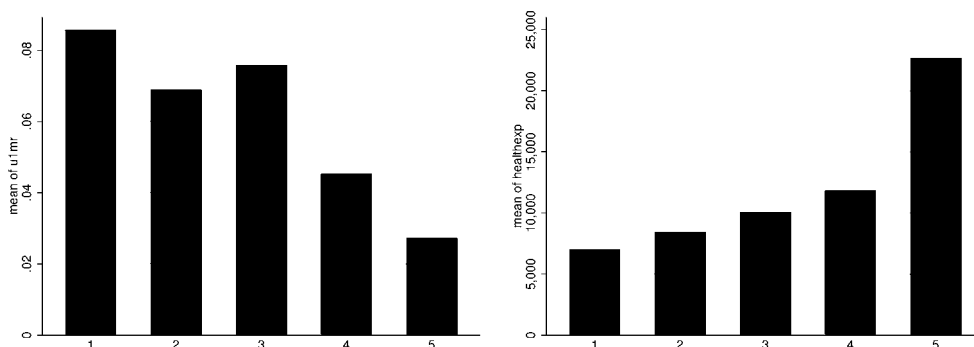


Figure 2. Infant mortality rate (left panel) and mean health care expenditure per capita (right panel) over wealth quintiles in Cambodia DHS, 2010

```
. conindex healthexp [aweight=sampweight_hh], rankvar(wealthindex) truezero
> cluster(PSU) graph ytitle(Cumulative share of healthexp)
> xtitle(Rank of wealthindex)
```

Index:	No. of obs.	Index value	Robust std. error	p-value
CI	15667	.24786719	.07246288	0.0007

(Note: Std. error adjusted for 611 clusters in PSU)

28. When graphing, `conindex` defaults to the variable label or, if it is unavailable, the variable name when labeling the axis. This can be overridden by specifying the `xtitle()` and `ytitle()` options and specifying the desired axis labels inside the parentheses. When `rankvar()` is not specified, `conindex` draws the Lorenz curve. Note also that the generalized concentration (Lorenz) curve will be drawn when the `generalized` option is also specified.

Figure 3 reveals that there is no ambiguity in the distribution of health expenditures. The concentration curve always lies below the diagonal, which indicates greater spending by those ranked higher according to the wealth index.

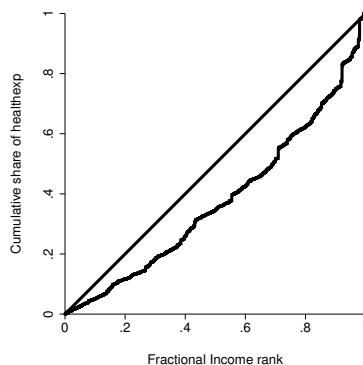


Figure 3. Concentration curve for out-of-pocket health care expenditure per capita against wealth index rank, Cambodia (DHS, 2010)

Because health care expenditures are unbounded and measured on a ratio scale, this estimate is robust to the proportionality factor arising from the choice of currency and can be used to rank inequality in medical spending in Cambodia against inequalities in other ratio-scale variables (for example, food expenditures) or health expenditure inequality in other countries. If one prefers that the measure of inequality in medical spending respect absolute invariance rather than relative invariance, then the generalized concentration index can be requested by typing

```
. conindex healthexp [aweight=sampweight_hh], rankvar(wealthindex) generalized
> truezero cluster(PSU)
```

Index:	No. of obs.	Index value	Robust std. error	p-value
Gen. CI	15667	2977.0381	870.32394	0.0007

(Note: Std. error adjusted for 611 clusters in PSU)

This gives an estimate of around 2,977 riel, which is obviously sensitive to the proportionality factor and cannot be used directly to compare inequality in medical spending across countries with different currencies.²⁹

The standard concentration index of infant mortality is negative, which indicates that infant deaths are concentrated among less wealthy households. The index for infant survival (`u1sr`) is correspondingly positive but differs greatly in absolute value from the index for mortality, which confirms that the mirror property does not hold and reflects imposition of relative invariance with respect to different variables (see also section 3.2).

29. The univariate Gini and generalized Gini index are obtained by omitting the `rankvar()` option in the `conindex` command.

Given the standard concentration index is insensitive to a proportional transformation of the variable of interest, the value used to indicate presence of a characteristic, for example, death = 1, is irrelevant provided the value used to indicate absence of that characteristic is fixed at 0.³⁰

```
. conindex u1mr [aweight=sampweight], rankvar(wealthindex) truezero cluster(PSU)
```

Index:	No. of obs.	Index value	Robust std. error	p-value
CI	14598	-.18890669	.02546028	0.0000

(Note: Std. error adjusted for 611 clusters in PSU)

```
. conindex u1sr [aweight=sampweight], rankvar(wealthindex) truezero cluster(PSU)
```

Index:	No. of obs.	Index value	Robust std. error	p-value
CI	14598	.01220632	.00164513	0.0000

(Note: Std. error adjusted for 611 clusters in PSU)

The generalized concentration indices of mortality and survival are equal in absolute value (see table 1), confirming that this index satisfies the mirror condition when it is applied to a binary variable. This is because the generalized concentration index for a binary variable equals one-fourth of the Erreygers index, which possesses the mirror property. But the generalized concentration index does not satisfy this condition in general. The Erreygers index is computed by specifying the **erreygers** option, along with two further options that indicate the variable is bounded and how it is coded.

```
. conindex u1mr [aweight=sampweight], rankvar(wealthindex) erreygers bounded  
> limits(0 1) cluster(PSU)
```

Index:	No. of obs.	Index value	Robust std. error	p-value
Erreygers norm. CI	14598	-.04586189	.00618113	0.0000

(Note: Std. error adjusted for 611 clusters in PSU)

```
. conindex u1sr [aweight=sampweight], rankvar(wealthindex) erreygers bounded  
> limits(0 1) cluster(PSU)
```

Index:	No. of obs.	Index value	Robust std. error	p-value
Erreygers norm. CI	14598	.04586189	.00618113	0.0000

(Note: Std. error adjusted for 611 clusters in PSU)

The Wagstaff index of infant mortality, which as explained above has different normative underpinnings, is computed by simply specifying **wagstaff** in place of **erreygers**.

30. Consult Erreygers and Van Ourti (2011a,b) and Wagstaff (2011a,b) for some discussion on this issue.

```
. conindex ulmr [aweight=sampweight], rankvar(wealthindex) wagstaff bounded
> limits(0 1) cluster(PSU)
```

Index:	No. of obs.	Index value	Robust std. error	p-value
Wagstaff norm. CI	14598	-.20111301	.02710541	0.0000

(Note: Std. error adjusted for 611 clusters in PSU)

```
. conindex ulsr [aweight=sampweight], rankvar(wealthindex) wagstaff bounded
> limits(0 1) cluster(PSU)
```

Index:	No. of obs.	Index value	Robust std. error	p-value
Wagstaff norm. CI	14598	.20111301	.02710541	0.0000

(Note: Std. error adjusted for 611 clusters in PSU)

The value of the Wagstaff index is close to that of the standard concentration index because the prevalence of infant deaths, at 6.1%, is close to 0; thus the index places greater weight on relative invariance with respect to presence of the characteristic (here, death) and so comes closer to the normative principle imposed by the standard concentration index. If the prevalence were 50%, then the Wagstaff index would give equal weight to relative invariance in attainments and shortfalls, which coincides with absolute invariance. In that case, its value would equal that of the Erreygers index (see Kjellsson and Gerdtham [2013b] for more discussion).

Table 1. Concentration indices estimated from the Cambodian DHS, 2010

	Health expenditure (healthexp)		Infant mortality (u1mr)		Infant survival (u1sr)	
Standard concentration index (C)	0.2479 (0.0725)		-0.1889 (0.0255)		0.0122 (0.0016)	
Generalized concentration index (GC)	2,977 (870)		-0.0115 (0.0015)		0.0115 (0.0015)	
Erreygers index (E)			-0.0459 (0.0062)		0.0459 (0.0062)	
Wagstaff index (W)			-0.2011 (0.0271)		0.2011 (0.0271)	
v, β	1.5	5	1.5	5	1.5	5
Extended concentration index (smrm EC(v))	0.1696	0.3818	-0.1201	-0.3187	0.0078	0.0206
Symmetric concentration index (SC(β))	0.2057	0.3943	-0.1683	-0.2492	0.0109	0.0161
Generalized extended concentration index (GEC(v))			-0.0492	-0.0362	0.0492	0.0362
Generalized symmetric concentration index (GSC(β))			-0.0409	-0.0605	0.0409	0.0605

Note: Robust standard errors that account for clustering at the level of the primary sampling unit (PSU) are in brackets.
The primary sampling units correspond to villages in the DHS.

The bottom panel of table 1 presents estimates of concentration indices with alternative attitudes to inequality to those underlying the standard concentration index. Setting the parameter v of the extended concentration index to 1.5 places relatively more weight on those residing in wealthier households, while setting the parameter to 5 gives more weight to the poorer observations. A value of 2 corresponds to the weighting implicit in the standard concentration index and so would result in an estimate equal to that of C . We use the same values for the β parameter of the symmetric index, where $\beta = 1.5$ corresponds to the case where more weight is placed on the middle of the wealth distribution, while $\beta = 5$ corresponds to a case where the extremes of the wealth distribution are more heavily weighted.³¹

The indices are computed as follows:

```
conindex varlist [aweight=sampweight], rankvar(wealthindex) truezero v(#) ///
      cluster(PSU)
conindex varlist [aweight=sampweight], rankvar(wealthindex) truezero      ///
      beta(#) cluster(PSU)
```

We emphasize that little can be learned from comparing extended indices computed for different values of v .³² Rather, one might check whether an inequality ordering across populations is robust to the choice of the value of v (β). If it is not, then a conclusion that a variable is more unequally distributed in one population than another needs to be made conditional on an explicit attitude toward inequality.

Generalized extended and symmetric indices are computed by simply adding the **generalized** option to the command lines immediately above. As is clear from the estimates in table 1, these indices satisfy the mirror condition.³³

```
conindex varlist [aweight= sampweight], rankvar(wealthindex) generalized ///
      truezero v(#) cluster(PSU)
conindex varlist [aweight= sampweight], rankvar(wealthindex) generalized ///
      truezero beta(#) cluster(PSU)
```

conindex allows estimates of all inequality indices to be compared across groups defined by a binary or categorical variable, and it tests the null of equality across groups. This is done by including the **compare()** option. For example, to compare wealth-related inequality in infant mortality across urban and rural locations, we can use

-
31. There is no particular reason to choose the same values for v and β . Our reason for doing so is that both v and β can be interpreted as the distance between the weights given to the least and most wealthy individual. See also sections 4.1 and 4.2.
 32. For instance, while their examples do not occur in the illustration in this article, Erreygers, Clarke, and Van Ourti (2012) report some empirical examples where initially pro-poor inequality reverses into pro-rich inequality when v is increased. The same reversal might also happen for the symmetric index.
 33. When $v = \beta = 2$, both indices are equal to the Erreygers index.


```
. conindex ulmr [aweight=sampweight], rankvar(wealthindex) erreygers bounded
> limits(0 1) cluster(PSU) compare(urban)
```

Index:	No. of obs.	Index value	Robust std. error	p-value
Erreygers norm. CI	14598	-.04586189	.00618113	0.0000

(Note: Std. error adjusted for 611 clusters in PSU)

For groups:

CI for group 1: urban = 0

Index:	No. of obs.	Index value	Robust std. error	p-value
Erreygers norm. CI	10969	-.02985274	.00724954	0.0000

(Note: Std. error adjusted for 420 clusters in PSU)

CI for group 2: urban = 1

Index:	No. of obs.	Index value	Robust std. error	p-value
Erreygers norm. CI	3629	-.02869979	.01006162	0.0048

(Note: Std. error adjusted for 191 clusters in PSU)

Test for stat. significant differences with Ho: diff=0 (assuming equal variances)

F-stat = .66985873	p-value= 0.4131
--------------------	-----------------

Test for stat. significant differences with Ho: diff=0 (large sample assumed)

Diff. = .00115295	Std. err. = .01240129	z-stat = 0.09	p-value = 0.9259
-------------------	-----------------------	---------------	------------------

The index estimated from the combined sample is displayed first. Then, the group-specific estimates are given. There is a significant concentration of infant mortality among the least wealthy in both rural and urban locations. The point estimates suggest that the degree of inequality is greatest in rural areas, but the difference with urban areas is small. Both tests fail to reject the null hypothesis that the index is the same in rural and urban locations.

8 Concluding remarks

This article introduced the user-written Stata command `conindex`, which calculates rank-dependent inequality indices while offering a great deal of flexibility in considering measurement scale and alternative attitudes to inequality. Estimation and inference is via a regression approach that can allow for sampling design, misspecification and grouped data and for testing for differences in inequality across populations.

Concentration indices are frequently used, particularly for the measurement of inequality in health by socioeconomic status. The indices estimated for different regions, periods, or groups could also be included in regression analyses as control variables. We hope that the greatly reduced computational cost offered by `conindex` will afford researchers the time to give greater consideration to their choice of index, ensuring that the one selected is appropriate for the scale of measurement and consistent with the normative position they are prepared to defend.

9 Acknowledgments

Owen O'Donnell and Tom Van Ourti acknowledge support from the National Institute on Aging under grant R01AG037398. We thank Ellen Van de Poel for assistance with the DHS data. The usual caveats apply, and all remaining errors are our responsibility.

10 References

- Bleichrodt, H., and E. van Doorslaer. 2006. A welfare economics foundation for health inequality measurement. *Journal of Health Economics* 25: 945–957.
- Bosmans, K. Forthcoming. Consistent comparisons of attainment and shortfall inequality: A critical examination. *Health Economics*.
- Chen, Z. A. 2007. `conindx`: Stata module to calculate concentration index with both individual and grouped data. Statistical Software Components S456802, Department of Economics, Boston College. <https://ideas.repec.org/c/boc/bocode/s456802.html>.
- Chow, G. C. 1960. Tests of equality between sets of coefficients in two linear regressions. *Econometrica* 28: 591–605.
- Clarke, P. M., U.-G. Gerdtham, M. Johannesson, K. Bingefors, and L. Smith. 2002. On the measurement of relative and absolute income-related health inequality. *Social Science and Medicine* 55: 1923–1928.
- Clogg, C. C., E. Petkova, and A. Haritou. 1995. Statistical methods for comparing regression coefficients between models. *American Journal of Sociology* 100: 1261–1293.
- Decoster, A., and E. Ooghe. 2003. Weighting with individuals, equivalent individuals or not weighting at all. Does it matter empirically? In *Inequality, Welfare and Poverty: Theory and Measurement*, ed. Y. Amiel and J. A. Bishop, vol. 9, 173–190. Amsterdam: JAI.
- Edbert, U. 1997. Social welfare when needs differ: An axiomatic approach. *Economica* 64: 233–244.
- Erreygers, G. 2009a. Can a single indicator measure both attainment and shortfall inequality? *Journal of Health Economics* 28: 885–893.

- . 2009b. Correcting the concentration index. *Journal of Health Economics* 28: 504–515.
- . 2009c. Correcting the concentration index: A reply to Wagstaff. *Journal of Health Economics* 28: 521–524.
- Erreygers, G., P. Clarke, and T. Van Ourti. 2012. “Mirror, mirror, on the wall, who in this land is fairest of all?”—Distributional sensitivity in the measurement of socioeconomic inequality of health. *Journal of Health Economics* 31: 257–270.
- Erreygers, G., and T. Van Ourti. 2011a. Putting the cart before the horse. A comment on Wagstaff on inequality measurement in the presence of binary variables. *Health Economics* 20: 1161–1165.
- . 2011b. Measuring socioeconomic inequality in health, health care and health financing by means of rank-dependent indices: A recipe for good practice. *Journal of Health Economics* 30: 685–694.
- Filmer, D., and L. H. Pritchett. 2001. Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of India. *Demography* 38: 115–132.
- Jenkins, S. P. 1988. Calculating income distribution indices from micro-data. *National Tax Journal* 41: 139–142.
- . 1999. ineqdeco: Stata module to calculate inequality indices with decomposition by subgroup. Statistical Software Components S366002, Department of Economics, Boston College. <https://ideas.repec.org/c/boc/bocode/s366002.html>.
- Jolliffe, D., and B. Krushelnysky. 1999. sg115: Bootstrap standard errors for indices of inequality. *Stata Technical Bulletin* 51: 28–32. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 191–196. College Station, TX: Stata Press.
- Kakwani, N., A. Wagstaff, and E. van Doorslaer. 1997. Socioeconomic inequalities in health: Measurement, computation, and statistical inference. *Journal of Econometrics* 77: 87–103.
- Kakwani, N. C. 1977. Measurement of tax progressivity: An international comparison. *Economic Journal* 87: 71–80.
- . 1980. *Income Inequality and Poverty: Methods of Estimation and Policy Applications*. New York: Oxford University Press.
- Kjellsson, G., and U.-G. Gerdtham. 2013a. Lost in translation: Rethinking the inequality equivalence criteria for bounded health variables. In *Research on Economic Inequality, Volume 21: Health and Inequality*, ed. P. R. Dias and O. O’Donnell, 3–32. Bingley, UK: Emerald.
- . 2013b. On correcting the concentration index for binary variables. *Journal of Health Economics* 32: 659–670.

- Kolm, S.-C. 1976. Unequal inequalities. I. *Journal of Economic Theory* 12: 416–442.
- Lambert, P., and B. Zheng. 2011. On the consistent measurement of attainment and shortfall inequality. *Journal of Health Economics* 30: 214–219.
- Lopez-Feldman, A. 2005. descogini: Stata module to perform Gini decomposition by income source. Statistical Software Components S456001, Department of Economics, Boston College. <https://ideas.repec.org/c/boc/bocode/s456001.html>.
- O'Donnell, O., E. van Doorslaer, A. Wagstaff, and M. Lindelow. 2008. *Analyzing Health Equity Using Household Survey Data: A Guide to Techniques and Their Implementation*. Washington, DC: The International Bank for Reconstruction and Development/The World Bank.
- Pereira, J. A. 1998. Inequality in infant mortality in Portugal, 1971–1991. In *Developments in Health Economics and Public Policy, Volume 6: Health, the Medical Profession, and Regulation*, ed. P. Zweifel, 75–93. Dordrecht: Kluwer Academic Publishers.
- Shorrocks, A. F. 1983. Ranking income distributions. *Econometrica* 50: 3–17.
- Townsend, P., and N. Davidson, eds. 1982. *Inequalities in Health: The Black Report*. Harmondsworth: Penguin.
- van Kerm, P., and S. P. Jenkins. 2007. Software Updates: Generalized Lorenz curves and related graphs: Update for Stata 7. *Stata Journal* 7: 280.
- Van Ourti, T. 2004. Measuring horizontal inequity in Belgian health care using a Gaussian random effects two part count data model. *Health Economics* 13: 705–724.
- Wagstaff, A. 2002. Inequality aversion, health inequalities and health achievement. *Journal of Health Economics* 21: 627–641.
- . 2005. The bounds of the concentration index when the variable of interest is binary, with an application to immunization inequality. *Health Economics* 14: 429–432.
- . 2009. Correcting the concentration index: A comment. *Journal of Health Economics* 28: 516–520.
- . 2011a. The concentration index of a binary outcome revisited. *Health Economics* 20: 1155–1160.
- . 2011b. Reply to Guido Erreygers and Tom Van Ourti's comment on 'The concentration index of a binary outcome revisited'. *Health Economics* 20: 1166–1168.
- Wagstaff, A., P. Paci, and E. van Doorslaer. 1991. On the measurement of inequalities in health. *Social Science and Medicine* 33: 545–557.
- Yitzhaki, S. 1983. On an extension of the Gini inequality index. *International Economic Review* 24: 617–628.

About the authors

Owen O'Donnell is a professor of applied economics in the Erasmus School of Economics at Erasmus University Rotterdam, a research fellow of the Tinbergen Institute, and an associate professor at the University of Macedonia (Greece).

Stephen O'Neill (corresponding author) is a research fellow in health economics in the Department of Health Services Research and Policy at the London School of Hygiene and Tropical Medicine.

Tom Van Ourti is a professor of applied health economics in the Erasmus School of Economics at Erasmus University Rotterdam and a research fellow of the Tinbergen Institute.

Brendan Walsh is a research fellow in health economics in the School of Health Sciences and the City Health Economics Centre at City University London.