9

Extensions to the Concentration Index: Inequality Aversion and the Health Achievement Index

The concentration index is a useful tool for measuring inequalities in the health sector. It does, however, have limitations.

First, like the Gini coefficient, it has implicit in it a particular set of value judgments about aversion to inequality. This chapter shows how to operationalize Wagstaff's (2002) "extended" concentration index, which allows attitudes to inequality to be made explicit, and to see how measured inequality changes as the attitude to inequality changes.

The second drawback of the concentration index—and the generalization of it—is that it is just a measure of inequality. Although equity is an important goal of health policy, it is not the only one. It is not just health inequality that matters; the average level of health also matters. Policy makers are likely to be willing to trade one off against the other—a little more inequality might be considered acceptable if the average increases substantially. This points to a second extension of the concentration index (Wagstaff 2002): a general measure of health "achievement" that captures inequality in the distribution of health (or some other health sector variable) as well as its mean.

The extended concentration index

The regular concentration index C is equal to (Kakwani, Wagstaff, and van Doorslaer 1997)

$$C = \frac{2}{n \cdot \mu} \sum_{i=1}^{n} h_i R_i - 1, \qquad (9.1)$$

where n is the sample size, h_i is the ill-health indicator for person i, μ is the mean level of ill health, and R_i is the fractional rank in the living-standards distribution of the ith person. The value judgments implicit in C are seen most easily when C is rewritten in an equivalent way as

$$C = 1 - \frac{2}{n \cdot \mu} \sum_{i=1}^{n} h_i \left(1 - R_i\right). \qquad (9.2)$$

The quantity $h_i/n\mu$ is the share of health (or ill health) enjoyed (or suffered) by person i. This is then weighted in the summation by twice the complement of the person's fractional rank, that is, $2(1 - R_i)$. So, the poorest person has his or her health share weighted by a number close to two. The weights decline in a stepwise fashion, reaching a number close to zero for the richest person. The concentration index is simply one minus the sum of these weighted health shares.

The *extended* concentration index can be written as follows:

(9.3) $$C(\nu)=1-\frac{\nu}{n\cdot\mu}\sum_{i=1}^{n}h_i\left(1-R_i\right)^{(\nu-1)} \qquad \nu>1.$$

In equation 9.3, ν is the inequality-aversion parameter, which will be explained below. The weight attached to the ith person's health share, $h_i/n\mu$, is now equal to $\nu(1-R_i)^{(\nu-1)}$, rather than by $2(1-R_i)$. When $\nu = 2$, the weight is the same as in the regular concentration index; so $C(2)$ is the standard concentration index. By contrast, when $\nu = 1$, everyone's health is weighted equally. This is the case in which the value judgment is that inequalities in health do not matter. So, $C(1) = 0$ however unequally health is distributed across the income distribution. As ν is raised above 1, the weight attached to the health of a very poor person rises, and the weight attached to the health of people who are above the 55th percentile decreases. For $\nu = 6$, the weight attached to the health of persons in the top two quintiles is virtually zero. When ν is raised to 8, the weight attached to the health of those in the top *half* of the income distribution is virtually zero (Figure 9.1).

Computing the extended concentration index on microdata

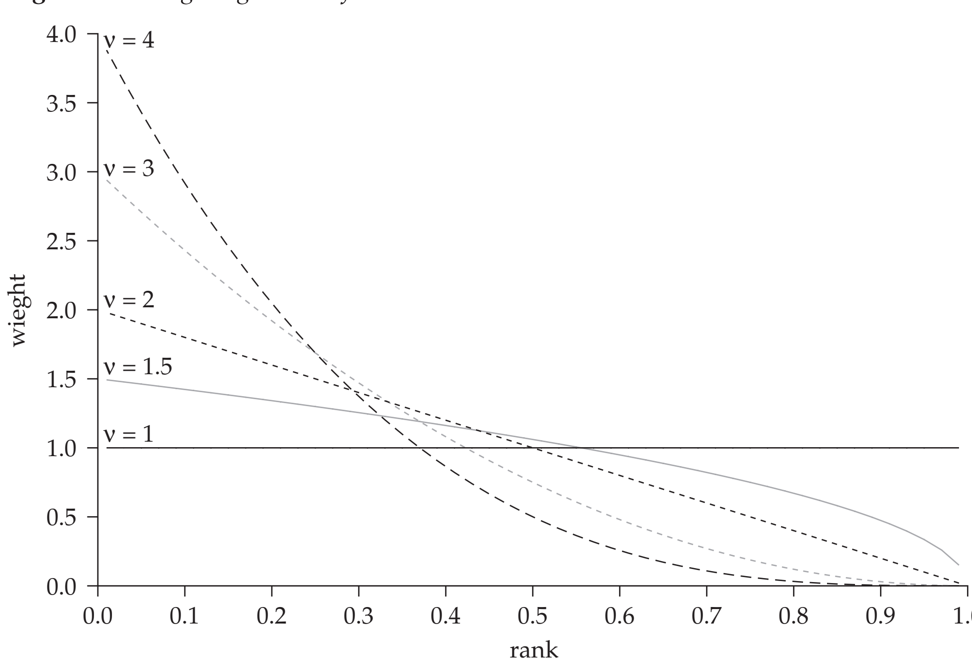
Like the regular concentration index, the extended concentration index can be written as a covariance (cf. equation 8.3). This is an easy way to compute the extended concentration index on microdata. The relevant covariance (Wagstaff 2002) is

(9.4) $$C(\nu)=-\frac{\nu}{\mu}\text{cov}\left(h_i,\left(1-R_i\right)^{\nu-1}\right).$$

This can be computed in a straightforward manner for different values of ν.

As an example, we compute the extended concentration index for values $\nu = 1$, $\nu = 2$, $\nu = 3$, $\nu = 4$, and $\nu = 5$ for (the negative of) height-for-age for children in the 1993 Vietnamese Living Standards Measurement Study. (The negative of the height-for-age variable captures malnutrition, the rate of which is higher among poorer children, so $C < 0$.) These are the same data used by Wagstaff, van Doorslaer, and

Figure 9.1 *Weighting Scheme for Extended Concentration Index*



Source: Authors.

Watanabe (2003). We know, of course, that $C(1) = 0$. We also know from Wagstaff, van Doorslaer, and Watanabe (2003) that for the year 1993 $C(2)$ is equal to –0.077 (cf. Wagstaff, van Doorslaer, and Watanabe 2003, p. 213). The Stata code to loop through $v = 1$, $v = 2$, $v = 3$, $v = 4$, and $v = 5$ is

```
sca drop _all
sum neghaz
scalar mean = r(mean)
forval i = 1/5 {
      ge adjrnk`i'=(1-rank)^(`i'-1)
      corr neghaz adjrnk`i' , covar
      sca ci`i' = -`i'*r(cov_12)/mean
}
sca li _all
```

Here `neghaz` is the negative of the height-for-age score (i.e., y_i), and `rank` is the fractional rank variable (i.e., R_i). The `sum` command stores the mean of `neghaz` in the scalar `mean`. For each of the values of v, the loop generates an adjusted rank variable, computes the required covariance, calculates the concentration index, and stores $C(v)$ in the scalar `ci`v. The `scalar list` command at the end produces the following:

```
sca   li _all
      ci5 = -.14068989
      ci4 = -.12858764
      ci3 = -.11006521
      ci2 = -.0771886
      ci1 = 0
      mean = 2.0298478
```

confirming that $C(1) = 0$ and $C(2) = -0.077$ for these data, and that as v is raised above 2, $C(v)$ becomes increasingly negative, reflecting the increasing weight that is being attached to the (ill-) health scores of poorer people.

Alternatively, the extended concentration index can be computed on microdata by means of a convenient regression (cf. equation 8.7). The appropriate convenient regression is

$$-v\,\mathrm{var}\left[\left(1-R_i\right)^{v-1}\right]\cdot\left[y_i/\mu\right]=\alpha_1+\beta_1\cdot\left(1-R_i\right)^{v-1}+u_i, \tag{9.5}$$

where β_1 is the extended concentration index. This is straightforward to set up and run once the desired values of v have been selected.

The Stata code to implement the convenient regression and loop through $v = 1$, $v = 2$, $v = 3$, $v = 4$, and $v = 5$ is

```
sum neghaz
scalar mean = r(mean)
forval i = 1/5 {
      ge adjrnk`i' = (1-rank)^(`i'-1)
      sum adjrnk`i'
      ge lhs`i' = -`i'*r(Var)*neghaz/mean
      reg lhs`i' adjrnk`i'
      sca ci`i' =_b[adjrnk`i']
}
sca li _all
```

The `scalar list` command at the end produces the following:

```
sca   li _all
      ci5 = -.14068989
      ci4 = -.12858764
      ci3 = -.11006521
      ci2 = -.0771886
      ci1 = 0
      mean = 2.0298478
```

which is identical to that produced by the covariance method.

Computing the extended concentration index on grouped data

The grouped-data analogue of equation 9.3 (Wagstaff 2002)[1] is as follows:

$$C(v) = v\sum_{t=1}^{T} f_t\left(1-R_t\right)^{(v-1)} - \frac{v}{\mu}\sum_{t=1}^{T} f_t h_t\left(1-R_t\right)^{(v-1)} \approx 1-\frac{v}{\mu}\sum_{t=1}^{T} f_t h_t\left(1-R_t\right)^{(v-1)} \tag{9.6}$$

where f_t is the sample proportion in the tth group, h_t is the average level of health or ill health of the tth group, and R_t is its fractional rank, defined as in chapter 8 as follows:

$$R_t = \sum_{\gamma=1}^{t-1} f_\gamma + \tfrac{1}{2} f_t, \tag{9.7}$$

indicating the cumulative proportion of the population up to the midpoint of each group interval.

This is easily implemented in a spreadsheet, as in table 9.1, taken from Wagstaff (2002). The example involves the distribution of under-five deaths in Bangladesh. The fractional rank variable, R, is derived using the formula above. In this case $v = 4$, and the column headed "`(1-R)^(v-1)`" gives the adjusted fractional rank for each asset group. The column headed "`h`" is the under-five mortality rate. The column headed "product" is the product of `f`, `h`, and `(1-R)^(v-1)`. The sum of these products (34.67) is then multiplied (in a cell not shown) by v, and divided by μ. The complement of this is the extended concentration index, in this case –0.0847, not dramatically different from $C(2)$, which is equal in this case to –0.0841.

Achievement—trading off inequality and the mean

The measure of "achievement" proposed in Wagstaff (2002) reflects the average level of health and the inequality in health between the poor and the better-off. It is defined as a weighted average of the health levels of the various people in the sample, in which higher weights are attached to poorer people than to better-off people. Thus achievement might be measured by the index:

$$I(v) = \frac{1}{n}\sum_{i=1}^{n} h_i v\left(1-R_i\right)^{(v-1)}, \tag{9.8}$$

[1]Note that equation A6 in Wagstaff (2002) contains a typo. Equation 9.6 above is the correct equation.

Table 9.1 *Inequality in Under-Five Deaths in Bangladesh*

Asset group	*No. births*	*f*	*R*	*(1-R)^(v-1)*	*h*	*Product*
1	2,950	0.22	0.11	0.71	141.1	21.85
2	3,191	0.24	0.34	0.29	146.9	10.11
3	2,695	0.20	0.56	0.09	135.2	2.36
4	2,581	0.19	0.75	0.02	122.3	0.35
5	2,029	0.15	0.92	0.00	76.0	0.00
	13,446				127.9	34.67

Source: Authors.

which is a weighted average of health levels, in which the weights are as graphed in Figure 9.1 and average to one. This index can be shown to be equal to (Wagstaff 2002) the following:

$$I(v) = \mu\big(1 - C(v)\big). \tag{9.9}$$

When h is a measure of ill health (so high values of $I(v)$ are considered bad) and $C(v) < 0$ (ill health is higher among the poor), inequality serves to raise the value of $I(v)$ above the mean, making achievement worse than it would appear if one were to look just at the mean. If ill-health declines monotonically with income, the greater the degree of inequality aversion, the greater the wedge between the mean, μ, and the value of the index $I(v)$.

Computing the achievement index

Given equation 9.9, there is nothing complicated about this. The Stata code below is the same as the code above for the extended concentration index, except that it adds a line to the loop that computes the achievement index for the current value of v and then adds a second loop that prints out a table, showing for each value of v, the values of $C(v)$ and $I(v)$. This can then be pasted into Excel or Word.

```
sum neghaz
scalar mean = r(mean)
forval i = 1/5 {
      ge adjrnk`i'=(1-rank)^(`i'-1)
      corr neghaz adjrnk`i' , covar
      sca ci`i' = -`i'*r(cov_12)/mean
      sca achiev`i' = mean*(1-ci`i')
}
forval i = 1/5 {
      di `i' _col(5) %5.3f ci`i' _col(20) %5.3f achiev`i'
_col(30)
}
```

For the example of child malnutrition, the last loop produces the following output:

```
1       0.000       2.030
2      -0.077       2.187
3      -0.110       2.253
4      -0.129       2.291
5      -0.141       2.315
```

The first column is the value of ν, the second is the value of $C(\nu)$, and the third is the value of $I(\nu)$. The latter is equal to μ when there is no aversion to inequality (i.e., $\nu = 1$). As ν increases above 1, measured inequality becomes more and more negative (in this example *h* is a "bad"), and $I(\nu)$ rises further and further above μ, meaning that the level of "disachievement" becomes larger and larger.

The spreadsheet computation of the achievement index is similarly straightforward, requiring just an extra cell in the spreadsheet above.

References

Kakwani, N. C., A. Wagstaff, and E. van Doorslaer. 1997. "Socioeconomic Inequalities in Health: Measurement, Computation and Statistical Inference." *Journal of Econometrics* 77(1): 87–104.

Wagstaff, A. 2002. "Inequality Aversion, Health Inequalities and Health Achievement." *Journal of Health Economics* 21(4): 627–41.

Wagstaff, A., E. van Doorslaer, and N. Watanabe. 2003. "On Decomposing the Causes of Health Sector Inequalities, with an Application to Malnutrition Inequalities in Vietnam." *Journal of Econometrics* 112(1): 207–23.