

13

Explaining Socioeconomic-Related Health Inequality: Decomposition of the Concentration Index

In the previous chapter we examined methods to explain the difference between two groups in the mean of some outcome variable of interest, which could be health or health care. By defining groups by socioeconomic status and using the method above, we can explain socioeconomic-related inequality in health or health care. But the degree of inequality captured is inevitably limited, given that group differences are examined. Measurement and explanation of inequality in health or health care across the entire distribution of some measure of socioeconomic status would be preferable. In chapter 8 we introduced the concentration index as a measure of socioeconomic-related inequality in health or health care. In this chapter we will explain how such inequality can be explained through decomposition of the concentration index.

Decomposition of the concentration index

For ease of exposition, we will refer to any health sector variable, such as health or health care use or payments, as “health” and to any (continuous) measure of socioeconomic status as “income.” Wagstaff, van Doorslaer, and Watanabe (2003) demonstrate that the health concentration index can be decomposed into the contributions of individual factors to income-related health inequality, in which each contribution is the product of the sensitivity of health with respect to that factor and the degree of income-related inequality in that factor. For any linear additive regression model of health (y), such as

$$(13.1) \quad y = \alpha + \sum_k \beta_k x_k + \varepsilon ,$$

the concentration index for y , C , can be written as follows:

$$(13.2) \quad C = \sum_k (\beta_k \bar{x}_k / \mu) C_k + GC_\varepsilon / \mu ,$$

where μ is the mean of y , \bar{x}_k is the mean of x_k , C_k is the concentration index for x_k (defined analogously to C), and GC_ε is the generalized concentration index for the error term (ε). Equation 13.2 shows that C is equal to a weighted sum of the concentration indices of the k regressors, where the weight for x_k is the elasticity of y

with respect to x_k $\left(\eta_k = \beta_k \frac{\bar{x}_k}{\mu} \right)$. The residual component—captured by the last term—

reflects the income-related inequality in health that is not explained by systematic variation in the regressors by income, which should approach zero for a well-specified model.

Wagstaff, van Doorslaer, and Watanabe (2003) use equation 13.2 to decompose income-related inequality in child malnutrition in Vietnam in 1993 and 1998. As in chapters 10 and 12, malnutrition is measured by the height-for-age z-scores (HAZ) of children younger than 10 years of age, and the measure of living standards is household consumption per capita. The z-scores are multiplied by -1 such that a greater value indicates more malnourishment. The specification of the regression model (equation 13.1) is very similar to that used in chapters 10 and 12. Here we include commune fixed effects to pick commune-level determinants of nutritional status. A summary of the results is presented in table 13.1. The (negative) concentration indices in the last row show that there was inequality in HAZ to the disadvantage of the poor in each year and that this inequality increased over time. The entries in each column are derived from equation 13.2 and give, for each year, the elasticity of HAZ with respect to each factor, the concentration index for each factor, and the total contribution of each factor to the HAZ concentration index. In each year, most of the consumption-related inequality in HAZ is explained by the direct effect of household consumption and by commune-level correlates of both malnutrition and consumption. The large elasticities of HAZ with respect to these factors are responsible for their large contribution to the HAZ concentration index. In contrast, there is a great deal of consumption-related inequality in access to both safe drinking water and satisfactory sanitation, but there is little sensitivity of HAZ to variation in these factors, and so they make little contribution to the HAZ concentration index.

Table 13.1 Decomposition of Concentration Index for Height-for-Age z-Scores of Children <10 Years, Vietnam, 1993 and 1998

	1993			1998		
	Elasticities	Concentration indices	Contributions	Elasticities	Concentration indices	Contributions
Child's age (in months)	1.137	0.020	0.023	1.630	0.018	0.030
Child's age squared	-0.634	0.030	-0.019	-0.880	0.028	-0.025
Child = male	0.022	0.003	0.000	0.045	0.014	0.001
(log)household consumption p.c.	-0.936	0.038	-0.035	-1.288	0.040	-0.052
Safe drinking water	-0.003	0.312	-0.001	-0.017	0.256	-0.004
Satisfactory sanitation	-0.009	0.468	-0.004	-0.006	0.508	-0.003
Years schooling household head	-0.017	0.065	-0.001	-0.015	0.094	-0.001
Years schooling mother	-0.037	0.075	-0.003	-0.003	0.108	-0.000
Fixed commune effects	1.477	-0.024	-0.035	1.534	-0.031	-0.047
"Residual"			-0.002			0.002
Total			-0.077			-0.099

Source: Wagstaff, van Doorslaer, Watanabe (2003, table 2).

Computation

The decomposition (equation 13.2) can be computed easily in Stata. First create the weighted fractional rank variable (`rank`) and estimate concentration index (`CI`) for the health variable (`y`) using the code provided in chapter 8. Generate a global `X` that refers to all the regressors in equation 13.1, estimate this regression, and create a scalar equal to the (weighted) mean of the health variable.

```
global X "varlist"
qui regr y $X [pw=weight]
sum y [aw=weight]
sca m_y=r(mean)
```

Then the factor specific elasticities, concentration indices, and contributions in equation 13.2 can be computed and displayed with the following loop,¹

```
foreach x of global X {
  qui {
    sca b_`x' = _b[`x']
    corr rank `x' [pw=weight], c
    sca cov_`x' = r(cov_12)
    sum `x' [pw=weight]
    sca elas_`x' = (b_`x'*r(mean))/m_y
    sca CI_`x' = 2*cov_`x'/r(mean)
    sca con_`x' = elas_`x'*CI_`x'
    sca prcnt_`x' = con_`x'/CI
  }
  di "`x' elasticity:", elas_`x'
  di "`x' concentration index:", CI_`x'
  di "`x' contribution:", con_`x'
  di "`x' percentage contribution:", prcnt_`x'
}
```

The final term in equation 13.2 can be obtained as a residual—the difference between the concentration index and the sum of the factor contributions.

Decomposition of change in the concentration index

Wagstaff, van Doorslaer, and Watanabe (2003) also proposed two approaches to explaining changes in income-related inequality over time. A first approach is to apply an Oaxaca-type decomposition (Oaxaca 1973) (see chapter 12). This can also be used to examine differences in inequality across cross-sectional units (van Doorslaer and Koolman 2004). Applying Oaxaca's method to equation 13.2 gives the following:

$$(13.3) \quad \Delta C = \sum_k \eta_{kt} (C_{kt} - C_{kt-1}) + \sum_k C_{kt-1} (\eta_{kt} - \eta_{kt-1}) + \Delta(GC_{et} / \mu_t),$$

where t indicates time period and Δ denotes first differences. As discussed in chapter 12, the Oaxaca decomposition is not unique and an alternative to equation 13.3 would be to weight the difference in concentration indices by the first period

¹We thank Xander Koolman, who originally wrote this code.

elasticity and weight the difference in elasticities by the second period concentration index.

This approach allows one to decompose change in income-related inequality in health into changes in inequality in the determinants of health, on the one hand, and changes in the elasticities of health with respect to these determinants, on the other. But it does not allow one to disentangle changes going on within the elasticities. To address this limitation, Wagstaff, van Doorslaer, and Watanabe (2003) consider the total differential of equation 13.2, allowing for changes in turn in the regression parameters, the means, and the concentration indices of the regressors. The change in the concentration index can be approximated (for small changes) by the following:

$$(13.4) \quad dC = -\frac{C}{\mu}d\alpha + \sum_k \frac{\bar{x}_k}{\mu}(C_k - C)d\beta_k + \sum_k \frac{\beta_k}{\mu}(C_k - C)d\bar{x}_k + \sum_k \frac{\beta_k \bar{x}_k}{\mu}dC_k + d\frac{GC_\varepsilon}{\mu}.$$

Note that the effect on C of a change in β_k , or in \bar{x}_k , depends on whether x_k is more unequally or less unequally distributed than y . This reflects two separate channels of influence—the direct effect of the change in β_k (or \bar{x}_k) on C and the indirect effect operating through μ . An increase in inequality in \bar{x}_k (i.e., C_k) will increase the degree of inequality in y . The impact is an increasing function of β_k and \bar{x}_k and a decreasing function of μ .

Wagstaff, van Doorslaer, and Watanabe (2003) use both equation 13.3 and equation 13.4 to decompose the change in income-related inequality in HAZ in Vietnam between 1993 and 1998. The results are summarized in table 13.2. Estimates of the percentage contribution of each determinant to the total change in C (third from last and last columns) are broadly similar across the two methods, with some important discrepancies. The Oaxaca-type method attributes more of the change to

Table 13.2 *Decomposition of Change in Concentration Index for Height-for-Age z-Scores of Children <10 Years, Vietnam, 1993–98*

	Decomposition of change in concentration index						
	Total differential approach (13.4)					Oaxaca-type approach (13.3)	
	β 's	Means of x 's	CI's	Total	Percent	Total	Percent
Child's age (in months)	0.003	0.011	-0.002	0.012	-57	0.007	-30
Child's age squared	0.003	-0.010	0.001	-0.006	29	-0.006	26
Child = male	0.001	0.000	0.000	0.001	-5	0.001	-3
Household consumption	-0.005	-0.005	-0.002	-0.011	52	-0.016	74
Safe drinking water	-0.002	0.000	0.000	-0.003	14	-0.003	16
Satisfactory sanitation	0.003	-0.002	0.000	0.001	-5	0.001	-5
Years schooling hhold. head	0.001	0.000	-0.001	0.000	0	0.000	1
Years schooling mother	0.005	0.000	-0.001	0.004	-19	0.003	-11
Fixed commune effects	0.000	-0.014	-0.010	-0.025	119	-0.012	55
"Residual"				0.005	-24	0.005	-24
Total	0.010	-0.021	-0.016	-0.021	100	-0.022	100

Source: Wagstaff, van Doorslaer, and Watanabe (2003, tables 3 and 4).

household consumption, whereas the differential approach gives more weight to changes occurring at the commune level. From the individual components of the total differential method (columns 2–4), we see that whereas changes in the means and concentration indices of the determinants of malnutrition have, on balance, tended to increase income-related inequality in HAZ, the opposite appears to be true of changes in the regression coefficients.

Computation

The components of equation 13.3 could be computed by running the regression and loop given in the previous section for each year of data, labeling the scalars to distinguish between their values in each year and taking the differences between them appropriately weighted, as in equation 13.3. The same general procedure could be used for equation 13.4, but differences between the year-specific regression coefficients and variables means would also have to be computed. That the total differential decomposition holds only for small changes must be kept in mind. Extrapolation to actual changes gives just an approximation to the change in the concentration index.

Extensions

As discussed in chapter 5, one is often interested in income-related inequality in a health sector variable after standardizing for correlates of income, such as age and gender. To assess equity in the distribution of health care (see chapter 15), it is also necessary to standardize for differences in “need.” The regression decomposition method is a convenient way of making such a standardization. One simply needs to deduct the contributions of the standardizing variables (included in the regression along with others) from the total concentration index. Van Doorslaer, Koolman, and Jones (2004) have demonstrated that this is equivalent to the two-step approach to indirect standardization discussed in chapter 5. Application of this approach to the measurement of inequity in health care use is discussed in chapter 15. This approach has been used to measure and decompose age-sex standardized income-related inequalities in self-reported health in Canada (van Doorslaer and Jones 2003) and in 13 European countries (van Doorslaer and Koolman 2004); to compare England, Wales, and Scotland during the 1979–1995 period (Gravelle and Sutton 2003); and to investigate the causes of changes in mental health in Great Britain (Wildman 2003).

Standard errors for the various components of the concentration index decomposition may be obtained by bootstrapping (van Doorslaer and Koolman 2004). Jones and Lopez-Nicolas (2004) extend this decomposition to a longitudinal setting, distinguishing between short-term inequality and the covariance between income and health through time.

The decomposition method relies on linearity of the underlying regression model. When the model is inherently nonlinear, it may be possible to base the decomposition on a linear approximation to the model. Van Doorslaer, Koolman, and Jones (2004) have used the “partial effects” representation of nonlinear count models to assess the degree of horizontal inequity in health care use in 12 European countries. This representation has the advantage of being a linear additive model of actual utilization, but it holds only by approximation, and the decomposition is

not unique but depends on the values at which partial effects are calculated. The approach is presented and discussed in chapter 15. Alternatively, Wan (2004) generalizes the regression-based decomposition method for application to any inequality measure with few restrictions on the underlying regression model.

References

- Gravelle, H., and M. Sutton. 2003. "Income Related Inequalities in Self-Assessed Health in Britain: 1979–1995." *Journal of Epidemiology and Community Health* 57(2): 125–29.
- Jones, A.M., and A. Lopez-Nicolas. 2004. "Measurement and Explanation of Socio-Economic Inequality in Health with Longitudinal Data." *Health Economics* 13: 1015–1030.
- Oaxaca, R. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14: 693–709.
- van Doorslaer, E., and A. M. Jones. 2003. "Inequalities in Self-Reported Health: Validation of a New Approach to Measurement." *Journal of Health Economics* 22.
- van Doorslaer, E., and X. Koolman. 2004. "Explaining the Differences in Income-Related Health Inequalities across European Countries." *Health Economics* 13(7): 609–28.
- van Doorslaer, E., X. Koolman, and A. M. Jones. 2004. Explaining Income-Related Inequalities in Doctor Utilization in Europe. *Health Economics* 13(7): 629–47.
- 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): 219–27.
- Wan, G. 2004. "Accounting for Income Inequality in Rural China: A Regression Based Approach." *Journal of Comparative Economics* 32: 348–63.
- Wildman, J. 2003. "Income Related Inequalities in Mental Health in Great Britain: Analysing the Causes of Health Inequality over Time." *Journal of Health Economics* 22(2): 295–312.