

# 7

## *Concentration Curves*

In previous chapters, we assessed health inequality through variation in mean health across quintiles of some measure of living standards. Although convenient, such a grouped analysis provides only a partial picture of how health varies across the full distribution of living standards. A complete picture can be provided using a concentration curve, which displays the share of health accounted for by cumulative proportions of individuals in the population ranked from poorest to richest (Kakwani 1977; Kakwani et al. 1997; Wagstaff et al. 1991). The concentration curve can be used to examine inequality not just in health outcomes but in any health sector variable of interest. It can also be used to assess differences in health inequality across time and countries. For example, it has been used to assess whether subsidies to the health sector are targeted toward the poor and whether the targeting is better in some countries than in others (O'Donnell et al. 2007; Sahn and Younger 2000). It has also been used to assess whether child mortality is more unequally distributed to the disadvantage of poor children in one country than in another (Wagstaff 2000) and whether inequalities in adult health are more pronounced in some countries than in others (van Doorslaer et al. 1997). Many other applications are possible.

In this chapter we explain how to compute a concentration curve. We also explain how to test whether a concentration curve departs significantly from an equal distribution and whether there is a statistically significant difference between two concentration curves that may represent different health services, time periods, or countries. This requires computation of standard errors of the concentration curve ordinates.

### **The concentration curve defined**

The two key variables underlying the concentration curve are the health variable, the distribution of which is the subject of interest, and a variable capturing living standards against which the distribution is to be assessed. Measurement of key health sector variables and of household living standards has been considered in earlier chapters. The health variable must be measured in units that can be aggregated across individuals. This is not necessary for the living standards measure, which is used only to rank individuals from richest to poorest.

The data could be at the individual level (e.g., raw household survey data), in which case values of both the health variable and the living standards variable are available for each observation. Alternatively, the data could be grouped, in which case,

for each living-standard group (e.g., income quintile), the mean value of the health variable is observed. The ranking of the groups (which group is poorest, which group is second poorest, and so on) and the percentage of the sample falling into each group are known. In the case of grouped data, the only advantage of the concentration curve over a table of group means is that it gives a graphical representation of the data.

The concentration curve plots the cumulative percentage of the health variable ( $y$ -axis) against the cumulative percentage of the population, ranked by living standards, beginning with the poorest, and ending with the richest ( $x$ -axis). In other words, it plots shares of the health variable against quantiles of the living standards variable. Examples are given in the figure in box 7.1 and in figures 7.1 and 7.2. So, for example, the concentration curve might show the cumulative percentage of health subsidies accruing to the poorest  $p$  percent of the population. If everyone, irrespective of his or her living standards, has exactly the same value of the health variable, the concentration curve will be a 45-degree line, running from the bottom left-hand corner to the top right-hand corner. This is known as the line of equality. If, by contrast, the health sector variable takes higher (lower) values among poorer people, the concentration curve will lie above (below) the line of equality. The farther the curve is above the line of equality, the more concentrated the health variable is among the poor.

Concentration curves for the same variable in different countries or time periods can be plotted on the same graph. Similarly, curves for different health sector variables in the same country and time period can be plotted against each other. For example, the analyst may wish to assess whether inpatient care is more unequally distributed than primary care. If the concentration curve for one country (or time period or health service) lies everywhere above that for the other, the first curve is said to dominate the second, and the ranking by degree of inequality is unambiguous.<sup>1</sup> Alternatively, curves may cross, in which case neither distribution dominates the other. It is then still possible to make comparisons of degrees of inequality but only by resorting to a summary index of inequality, which inevitably involves the imposition of value judgments concerning the relative weight given to inequality arising at different points in the distribution (see chapter 8). Rankings by degree of inequality can then differ depending on the inequality index chosen.

### Graphing concentration curves—the grouped-data case

In the grouped-data case, the required data and the corresponding charts are easily produced in a spreadsheet program such as Microsoft Excel. The table in box 7.1 is pasted directly from Excel and contains all the data required to plot the concentration curve shown in the box. The curve is constructed in Excel using the XY (scatter) chart-type with the “scatter with data points connected by smoothed lines” option. The first series graphs the line of equality, the  $x$ -values and the  $y$ -values both being the cumulative percentage of the sample. The no-marker option is selected for the line of equality. The second series graphs the concentration curve, the  $x$ -values being the cumulative percentage of the sample, the  $y$ -values being the cumulative percentage of the health variable. It is important to include a 0 percent in both series. Both the  $x$ -axis and the  $y$ -axis need to be restricted to the range 0 to 100 percent.

---

<sup>1</sup>For an introduction to the concept of dominance, its relation to inequality measurement, and the related concept of stochastic dominance, see Deaton (1997).

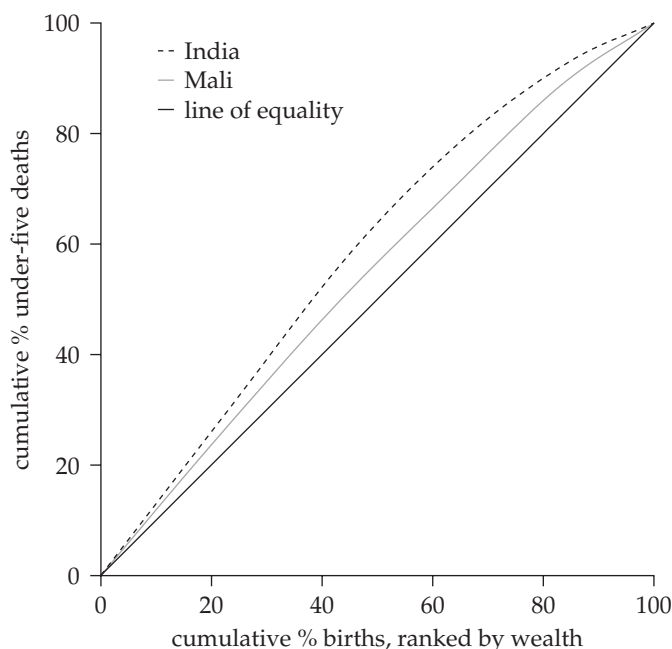
**Box 7.1** Example of a Concentration Curve Derived from Grouped Data

In this example, the sample comprises births, the living standards measure is the assets (wealth) index, and the health variable is deaths of children under five years of age. The data are from the demographic and health surveys of India and Mali. The table shows the number of births in each wealth index quintile during the period 1982–92 in India. Expressing these as percentages of the total number of births and cumulating them gives the cumulative percentage of births, ordered by wealth. That is what is plotted on the *x*-axis in the figure. Also shown are the under-five mortality rates (U5MR) for each of five wealth groups. Multiplying the U5MR by the number of births gives the number of deaths in each wealth group. Expressing these as a percentage of the total number of deaths and cumulating them gives the cumulative percentage of deaths. That is what is plotted on the *y*-axis in the figure. The concentration curve for India lies above (dominates) the line of equality, indicating that child deaths are concentrated among the poor. Also shown in the figure is the concentration curve for under-five deaths for Mali for the period 1985–95. The Mali curve lies everywhere below that of India (i.e., the India curve dominates the Mali curve), indicating there is less inequality in under-five mortality in Mali than in India.

*Under-Five Deaths in India, 1982–92*

| Wealth group  | No. of births | Rel % births | Cumul % births | U5MR per 1,000 | No. of deaths | Rel % deaths | Cumul % deaths |
|---------------|---------------|--------------|----------------|----------------|---------------|--------------|----------------|
|               |               |              | 0              |                |               |              | 0              |
| Poorest       | 29,939        | 23           | 23             | 154.7          | 4,632         | 30           | 30             |
| 2nd           | 28,776        | 22           | 45             | 152.9          | 4,400         | 29           | 59             |
| Middle        | 26,528        | 20           | 66             | 119.5          | 3,170         | 21           | 79             |
| 4th           | 24,689        | 19           | 85             | 86.9           | 2,145         | 14           | 93             |
| Richest       | 19,739        | 15           | 100            | 54.3           | 1,072         | 7            | 100            |
| Total/average | 129,671       |              |                | 118.8          | 15,419        |              |                |

*Concentration Curves for Under-Five Deaths in India and Mali*



Source: Authors.

## Graphing concentration curves—the microdata case

Graphing concentration curves in Stata can be done using the command `glcurve` (an ado file downloadable from the Stata Web site). However, better-looking charts can be obtained using the `twoway` command.

In the example that follows, we plot concentration curves for child malnutrition in Vietnam in 1992/93 and 1997/98, with malnutrition measured by the negative of height-for-age, as in Wagstaff et al. (2003). The dataset contains stacked data for the two years (`year` being 0 for 1992/93 and 1 for 1997/98), `neghaz` and `lnpcexp` being, respectively, the negative of height-for-age and the log of per capita expenditure for the year in question. We ignore below-sample weights, but they can be incorporated in both of the approaches below, as will be clear from Stata code elsewhere in the book.

The concentration curve can be produced directly using `glcurve` as follows:

```
glcurve neghaz, glvar(yord) pvar(rank) sortvar(lnpcexp)
replace by(year) split lorenz
```

The rank variable here is the `lnpcexp`. `glcurve` generates three new variables: `rank`, which is the child's rank in the expenditure distribution in each year, and `yord_0` and `yord_1`, which are respectively the  $y$ -ordinates for 1992/93 and 1997/98. Adding the `lorenz` option requests that  $y$ -ordinates be cumulative proportions of the health variable and not the cumulative means, which is the default.

Alternatively, these three variables could have been obtained through the following commands:

```
sort year lnpcexp
forval i = 0/1 {
    sum neghaz if year==`i'
    scalar nobs`i' = r(N)
}
ge rank=.
egen tmp = rank(lnpcexp) if year==0
replace rank=tmp/nobs0 if year==0
drop tmp
egen tmp = rank(lnpcexp) if year==1
replace rank=tmp/nobs1 if year==1

forval i = 0/1 {
    sum neghaz if year==`i'
    scalar s_malnut`i' = r(sum)
}
gen yord_0 = sum(neghaz)/s_malnut0 if year==0
gen yord_1 = sum(neghaz)/s_malnut1 if year==1
```

Whichever way the  $x$ -ordinates and  $y$ -ordinates are obtained, the concentration curve can be graphed using the full range of the options provided by `twoway` as follows:

```
ge rank2=rank
lab var yord_0 "1992/93"
lab var yord_1 "1997/98"
```

```

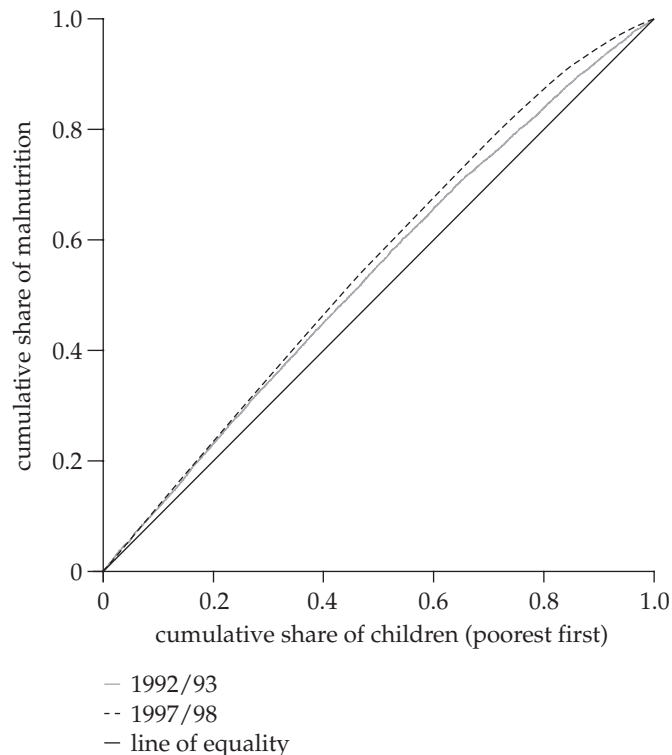
lab var rank "cumul share of children (poorest first)"
lab var rank2 "line of equality"

twoway (line yord_0 rank , sort clwidth(medthin) ///
clpat(solid))(line yord_1 rank, sort clwidth(medthin) ///
clpat(longdash) clcolor("153 204 0"))(line rank2 rank , ///
sort clwidth(medthin) clcolor(gray)), ///
ytitle(cumulative share of malnutrition, size(medsmall)) ///
yscale(titlegap(5)) xtitle(, size(medsmall)) ///
legend(rows(5)) xscale(titlegap(5)) ///
legend(region(lwidth(none))) plotregion(margin(zero)) ///
ysize(5.75) xsize(5) plotregion(lcolor(none))
graph export "$path0\cc curves 1992 and 1997.emf" , replace

```

The first line generates a duplicate rank variable that allows the line of equality to be plotted and labeled. The colors, pattern, and thickness of the concentration curves are controlled in the `twoway` command using the `clcolor`, `clwidth`, and `clpat` options. (The `///` in the code simply allows code to be continued over several lines.) The last line of code exports the graph in Windows Enhanced Meta Format (emf), which allows easy viewing from within Windows Explorer and easy insertion into a Word document or a PowerPoint presentation (using Insert, Picture, From File). Figure 7.1 shows the resultant concentration curve chart, which reveals,

**Figure 7.1** Concentration Curve for Child Malnutrition in Vietnam, 1992/93 and 1997/98



Source: Authors.

as indicated in Wagstaff et al. (2003), that inequality in child malnutrition increased somewhat in Vietnam between 1992/93 and 1997/98.

### Testing concentration curve dominance

Concentration curves are estimated from survey data and so are subject to sampling variability. Although visual inspection of a concentration curve in comparison with the 45-degree line or another concentration curve may give an impression of whether there is dominance, obviously this inspection is not sufficient to conclude whether or not dominance is statistically significant. To make inferences about dominance, the standard errors of the concentration curve ordinates must be computed in addition to their point estimates. If the analyst wishes to test dominance of a concentration curve against the Lorenz curve of income/consumption or against another concentration curve estimated from the same sample, then the standard errors for the differences between curve ordinates must be computed. This is complicated by the fact that, in such cases, the curves are dependent. The appropriate variance-covariance matrix, allowing for dependence between curves, has been derived by Bishop et al. (1994) and Davidson and Duclos (1997).

One decision rule that has been used in Lorenz (concentration) dominance tests has been to reject the null of nondominance in favor of dominance if there is at least one significant difference between curves (or a curve and the 45-degree line) in one direction and no significant difference in the other. For example, if there is at least one quantile point at which curve A lies significantly above curve B and there is no quantile point at which curve B lies above curve A, then it is concluded that A dominates B. If conventional critical values are used with such a decision rule, then there will be overrejection of the null because there is no correction for the fact that multiple comparisons are being made (Howes 1996). One solution is to use the same decision rule but to take multiple testing into account by using critical values from the studentized maximum modulus (SMM) distribution (Beach and Richmond 1985; Bishop et al. 1992; Stoline and Ury 1979). This is referred to as *the multiple comparison approach* (Dardanoni and Forcina 1999). An alternative criterion requires significant difference between ordinates at all quantile points to accept dominance (Howes 1996; Sahn and Younger 2000; Sahn et al. 2000). This is consistent with the intersection union principle (Dardanoni and Forcina 1999). Dardanoni and Forcina (1999) present Monte Carlo evidence showing that although this stricter rule reduces the probability of erroneously rejecting nondominance, it has greatly reduced the power of detecting dominance when true. If there is at least one significant difference between ordinates in each direction, then it is concluded that curves cross. If there are no significant differences in either direction, then, with the multiple comparison approach, null of nondominance is not rejected.

Besides the decision rule, the analyst must choose the number of quantile points at which ordinates are to be compared. If the number of comparison points is too restricted, then dominance across the full range of the distribution is not being tested. It is difficult, however, to find dominance at the extremes of distributions (Howes 1996). With reasonably large samples, a popular choice has been to test for differences at 19 evenly spaced quantiles from 0.05 to 0.95 (O'Donnell et al. forthcoming; Sahn and Younger 2000; Sahn et al. 2000).

The best statistical package for dominance testing is DAD, a specialist package for poverty and inequality analysis.<sup>2</sup> We have written our own Stata ado file for dominance testing.<sup>3</sup> The command follows the conventional Stata syntax,

```
dominance varlist [if] [in] [weight] [using filename],
  sortvar() [options]
```

If one variable is included in *varlist*, dominance of the concentration curve for this variable is tested against both the 45-degree line and the Lorenz curve of the living standards variable specified in *sortvar()*, which must be included. The default uses the multiple comparison approach decision rule, with comparisons at 19 equally spaced quantile points and a 5 percent significance level. The decision criterion can be changed to that of the intersection union principle with the option *rule(iup)*, or results using both decision rules can be requested with *rule(both)*. The number of comparison points can be changed with the option *level(#)*, with 20 being the maximum value. The significance level can be changed from 5 percent to 1 percent with *level(1)*.

Quintile (or decile) cumulative shares of the health variable and the living standards variable can be requested by the option *shares(quintiles)*. This will also report the *p*-value for a test of significant differences between the cumulative shares of the health and living standards variables at each quintile (decile). Differences between the shares of each variable and the population shares are also tested.

To illustrate, we use data from the 1995-6 National Sample Survey to test dominance of the concentration curve for the public health subsidy (*totsub*) in India against both the 45-degree and the Lorenz curve for equivalent household consumption (*hhexp\_eq*) as follows:<sup>4</sup>

```
dominance totsub [aw=weight], sortvar(hhexp_eq)
shares(quintiles)
```

Results confirm that the concentration curve is dominated by (lies below) the 45-degree line, but there is no dominance between the concentration curve of the subsidy and the Lorenz curve of household consumption. The cumulative quintile shares for the subsidy reported with the output are as follows:

Cumulative shares of *totsub*

```
Quantile Cum. share std. error Diff. from Diff. from
pop. share income share
```

```
p-value p-value
```

```
-----
q20 12.4911% 0.9554 0.0000 0.0366
q40 26.8893% 1.3009 0.0000 0.0979
q60 43.3710% 1.5738 0.0000 0.4922
q80 67.0052% 1.7219 0.0000 0.1175
-----
```

<sup>2</sup>The program can be downloaded free from <http://132.203.59.36/dad/>.

<sup>3</sup>The file can be downloaded from the Web site for this book. The program calls two other ados that need to be downloaded from the Stata Web site, *glcurve* and *locpoly*.

<sup>4</sup>For details of the application see O'Donnell et al (forthcoming).

The poorest 20 percent of individuals receive only 12.5 percent of the subsidy. The  $p$ -value in the fourth column confirms that this subsidy share is significantly less than the respective population share, and that is true at all quintiles. The  $p$ -values in the final column indicate that the subsidy shares are not significantly different from the consumption shares for all quintiles. Although there is a significant difference (at 5 percent) for the first quintile group, that is not inconsistent with the earlier finding of nondominance against the Lorenz curve, because in the dominance test critical values are adjusted for multiple comparisons.

If two variables are included in *varlist*, the program tests for dominance between the concentration curves of the two variables, allowing for dependence between the curves. In this case, the *shares()* option is not available.

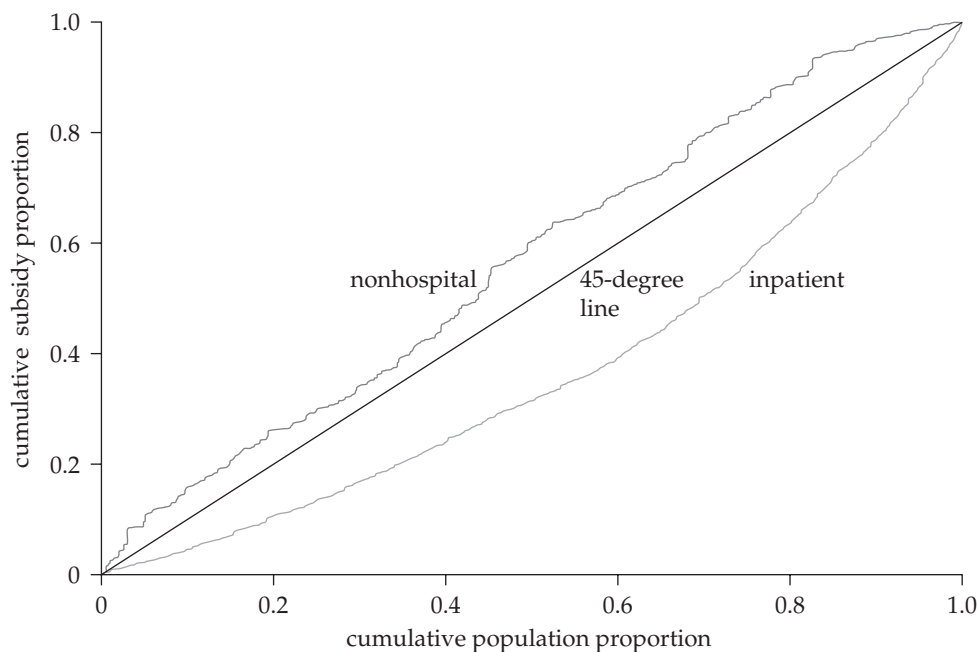
To illustrate, we examine dominance between the concentration curves for the subsidy to nonhospital care (*nonhsub*) and the subsidy to hospital inpatient care (*ipsub*) in India (see O'Donnell et al. [forthcoming]). From figure 7.2, it appears that the concentration curve for nonhospital care lies above that for inpatient care.

Figure 7.2 was produced with the following commands:

```
#delimit ;
glcurve nonhsub [aw=weight], sortvar(hhexp_eq) lorenz pvar(rank)
    glvar(nonh) nograph;

glcurve ipsub [aw=weight], sortvar(hhexp_eq) lorenz
    plot(line nonh rank, legend(label(2 "non-hospital")))
    line rank rank, legend(label(3 "45 degree"))
    legend(label(1 "inpatient")) l1(Cumulative subsidy proportion);
```

**Figure 7.2** Concentration Curves of Public Subsidy to Inpatient Care and Subsidy to Nonhospital Care, India, 1995–96



Source: Authors.



where the first `g1curve` command generates the  $x$  and  $y$  coordinates of the concentration curve for the nonhospital subsidy, and the second command plots this curve along with that for inpatient care and the 45-degree line.

We test for statistically significant dominance between the curves with

```
dominance nonhsub ipsub [aw=weight], sortvar(hhexp_eq)
```

The result confirms that the subsidy to nonhospital care dominates (is more pro-poor than) the inpatient subsidy.

The analyst may wish to test dominance between two concentration curves of the same variable estimated from independent samples. In this case, dominance testing is more straightforward because the ordinates of the respective concentration curves are independent. An important application is for cross-country comparisons of the distribution of a particular health or health care variable. For example, is the distribution of the public subsidy to health care less pro-poor in India than it is in Vietnam? To answer this question, we use data from the Indian 1995/96 NSS and the 1998 Vietnam Living Standards Survey.<sup>5</sup> The appropriate test can be carried out using the `dominance` command. In this case, the `using` option must be specified, and `varlist` must contain only one variable. Assume that we have the Indian data file loaded in Stata, that the Vietnam data file is named `Vietnam`, and that both files contain the variables `totsub` (total public health subsidy), `hhexp_eq` (equivalent household consumption), and `weight`. The syntax would then be as follows:

```
#delimit ;
dominance totsub [aw=weight] using Vietnam,
  sortvar(hhexp_eq) labels(India Vietnam);
```

where the `labels()` option gives labels to the two concentration curves that are used in the output. The result is reported as follows:

```
Test of dominance between concentration curves of totsub for
India and Vietnam
```

```
Data 1 Data 2 Sign. level # points Rule
```

```
-----
```

```
India Vietnam 5% 19 mca
```

```
Vietnam dominates India
```

This confirms that the distribution of the public health subsidy is more pro-poor in Vietnam than it is in India.

One other thing the analyst might want to do is to check dominance across years. For example, is it really the case that the 1997/98 concentration curve in Figure 7.2 lies everywhere above that of 1992/93? `dominance` requires the 1992/93 and 1997/98 data to be in two separate data files. The code below first saves the data set with the newly constructed rank variable, then saves the 1997/98 cases separately in the file `VN97.dta`. The code then loads the original stacked data set

---

<sup>5</sup>See O'Donnell et al. (2007) for details and for more cross-country comparisons of the distribution of public health subsidies in Asia.

containing the data for 1992/93 and 1997/98, drops the observations for 1997/98, and runs the dominance routine, requesting that results using both decision rules be reported.

```
save "$path0\ch 7.dta", replace
use "$path0\ch 7.dta", clear
keep if year==1
save "$path0\VN97.dta", replace
use "$path0\ch 7.dta", clear
keep if year==0
dominance neghaz using VN97, sortvar(rank) labels(1993 1997)
rule(both)
```

The output from Stata in this case is

Test of dominance between concentration curves of neghaz for 1993 and 1997

| <i>Data 1</i>       | <i>Data 2</i> | <i>Sign. level</i> | <i># points</i> | <i>Rule</i> |
|---------------------|---------------|--------------------|-----------------|-------------|
| 1993                | 1997          | 5%                 | 19              | mca         |
| 1997 dominates 1993 |               |                    |                 |             |
| 1993                | 1997          | 5%                 | 19              | iup         |
| nondominance        |               |                    |                 |             |

In this case, the 1997/98 concentration curve dominates (lies above) that of 1992/93 according to the less strict mca option, but not according to the stricter iup option. This reflects the fact that the two curves overlap toward the bottom of the income distribution.

## References

- Beach, C., M., and J. Richmond. 1985. "Joint Confidence Intervals for Income Shares and Lorenz Curves." *International Economic Review* 26: 439–50.
- Bishop, J. A., J. P. Formby, and P. D. Thistle. 1992. Convergence of the South and Non-South Income Distributions, 1969–1979. *American Economic Review* 82(1): 262–72.
- Bishop, J. A., K. V. Chow, and J. P. Formby. 1994. "Testing for Marginal Changes in Income Distributions with Lorenz and Concentration Curves." *International Economic Review* 35(2): 479–88.
- Dardanoni, V., and A. Forcina. 1999. "Inference for Lorenz Curve Orderings." *Econometrics Journal* 2: 49–75.
- Davidson, R., and J.-Y. Duclos. 1997. "Statistical Inference for the Measurement of the Incidence of Taxes and Transfers." *Econometrica* 65(6): 1453–65.
- Deaton, A. 1997. *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*. Baltimore, MD: Johns Hopkins University Press.
- Howes, S. 1996. "A New Test for Inferring Dominance from Sample Data." Mimeo. World Bank, Washington, DC.
- Kakwani, N. C. 1977. "Measurement of Tax Progressivity: An International Comparison." *Economic Journal* 87(345): 71–80.

- 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.
- O'Donnell, O., E. van Doorslaer, R. P. Rannan-Eliya, A. Somanathan, S. R. Adhikari, D. Harbianto, C. G. Garg, P. Hanvoravongchai, M. N. Huq, A. Karan, G. M. Leung, C.-W. Ng, B. R. Pande, K. Tin, L. Trisnantoro, C. Vasavid, Y. Zhang, and Y. Zhao. 2007. "The Incidence of Public Spending on Health Care: Comparative Evidence from Asia." *World Bank Economic Review* 21(1): 93–123.
- Sahn, D., and D. Younger. 2000. "Expenditure Incidence in Africa: Microeconomic Evidence." *Fiscal Studies* 21(3): 321–48.
- Sahn, D. E., S. D. Younger, and K. R. Simler. 2000. "Dominance Testing of Transfers in Romania." *Review of Income and Wealth* 46(3): 309–27.
- Stoline, M. R., and H. K. Ury. 1979. "Tables of the Studentized Maximum Modulus Distribution and an Application to Multiple Comparisons among Means." *Technometrics* 21(1): 87–93.
- van Doorslaer, E., A. Wagstaff, H. Bleichrodt, S. Calonge, U. G. Gerdtham, M. Gerfin, J. Geurts, L. Gross, U. Hakkinen, R. E. Leu, O. O'Donnell, C. Propper, F. Puffer, M. Rodriguez, G. Sundberg, and O. Winkelhake. 1997. "Income-Related Inequalities In Health: Some International Comparisons." *J Health Econ* 16(1): 93–112.
- Wagstaff, A. 2000. "Socioeconomic Inequalities in Child Mortality: Comparisons across Nine Developing Countries." *Bulletin of the World Health Organization* 78(1): 19–29.
- Wagstaff, A., P. Paci, and E. van Doorslaer. 1991. "On the Measurement of Inequalities in Health." *Soc Sci Med* 33(5): 545–57.
- 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.

