Concentration Curves

Introduction

The *concentration curve* [1, 2], and related *concentration index* (q.v.), provides a means of assessing the degree of income-related inequality in the distribution of a health variable. For example, it could be used to assess whether subsidies to the health sector are better targeted towards the poor in some countries than others [3], or whether child mortality is more unequally distributed to the disadvantage of poor children in one country than another [4], or whether inequalities in adult health are more pronounced in some countries than in others [5]. Many other applications are possible.

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 . Other Technical Notes discuss the measurement of key health sector variables and of household living standards. The data could be at 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 (e.g. 20% in each) is known.

The concentration curve plots the cumulative percentage of the health variable (y-axis) against the cumulative percentage of the sample, ranked by living standards, beginning with the poorest, and ending with the richest (x-axis). (See Figures 1-3 below for examples of a concentration curve.) So, for example, the concentration curve might show the cumulative percentage of health subsidies accruing to the poorest *p*% of the sample. If everyone, irrespective of his living standards, has exactly the same value of the health variable, the concentration curve will be a 45^0 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 amongst poorer people, the concentration curve will lie above (below) the line of equality. The further the curve is above the line of equality, the more concentrated the health variable is amongst the poor. Concentration curves for the same variable in different countries or time periods can be plotted on the same graph. If the concentration curve for one country (or time period) lies everywhere closer to the line of equality than another, the first curve is said to *dominate* the second and the ranking by degree of inequality is unambiguous. *Non-dominance* arises when concentration curves cross. In such cases, a summary measure of inequality is required in order to rank countries, or time periods, by degree of inequality (see Technical Note 7). If the variable takes on smaller values amongst the poor, the concentration curve will lie below the line of equality, and the further below the line of equality the concentration curve lies, the more concentrated amongst the better off the variable in question is.

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. Table 1 is pasted directly from Excel and contains all the data required to plot the concentration curve in Figure 1. 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% in both series. Both the x-axis and the y-axis need to be restricted to the range 0% to 100%.

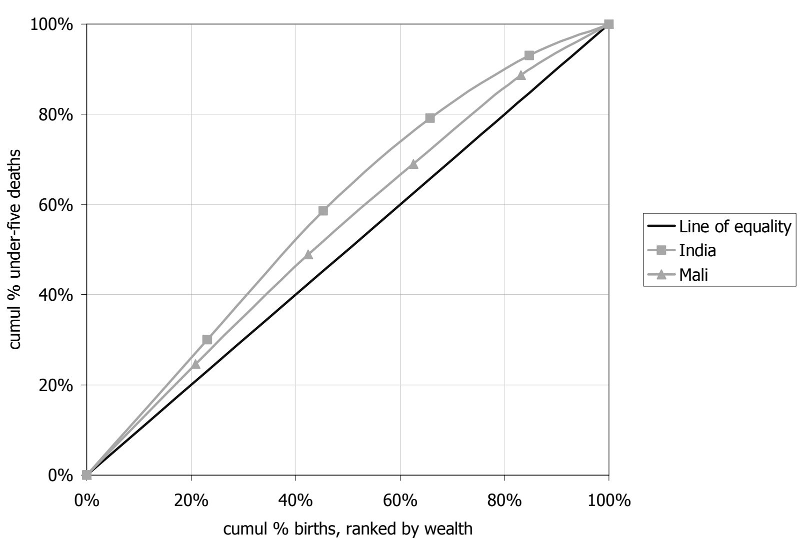
Box 1:
Example of a Concentration Curve Derived from Grouped Data

In this example, the sample comprises births, the living standards measure is wealth and the health variable is deaths of children under-five years. Table 1 shows the number of births in each wealth group over 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. This is what is plotted on the *x*-axis in Figure 1 below. 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 . This is what is plotted on the *y*-axis in Figure 1. The concentration curve for India lies above the *line of equality*, indicating that child deaths are concentrated amongst the poor. Also shown in Figure 1 is the concentration curve for under-five deaths for Mali for the period 1985-95. The Mali curve lies everywhere closer to the *line of equality* than that of India (i.e. the Mali curve *dominates* the India curve), indicating there is less inequality in under-five mortality in Mali than in India.

Table 1: Under-five deaths in India, 1982-92

Wealth group	No. of births	rel % births	cumul % births	U5MR per 1000	No. of deaths	rel % deaths	cumul % deaths
			0%				0%
Poorest	29939	23%	23%	154.7	4632	30%	30%
2nd	28776	22%	45%	152.9	4400	29%	59%
Middle	26528	20%	66%	119.5	3170	21%	79%
4th	24689	19%	85%	86.9	2145	14%	93%
Richest	19739	15%	100%	54.3	1072	7%	100%
Total/average	129671			118.8	15419		

Figure 1: Concentration curves for under-five deaths in India and Mali



Graphing concentration curves—the micro-data case

In the micro-data case, there are, in effect, two options for graphing concentration curves. One is to produce concentration curves on the micro-data using the graphing routines in packages such as Stata or

SPSS. The other is to use Stata or SPSS to manipulate the micro-data into groups and then construct the graphs in a spreadsheet program such as Excel, along the lines outlined above. This second option is useful when color is unavailable and more than one concentration curve is being graphed on one chart, since with several thousand data points, markers in an SPSS or Stata chart become indistinguishable. Excel also gives the user somewhat better control over the appearance of the graph than SPSS or Stata do. Against this must be offset the loss of precision which arises through aggregation.

SPSS commands for concentration curves

In this example, based on data from the 1998 Vietnam Living Standards Survey (VLSS), the concentration curve shows inequality in the distribution of medically attended deliveries in Vietnam. The sample consists births, the living standards variable is equivalent household consumption and the health variable is medically attended deliveries. The first step is to generate a variable indicating rank in the distribution of the living standards variable, `eqcons`. The `WEIGHT` command weights the data by the sample weight, `wt`. Failure to use the `WEIGHT` command will result in ranks which are not representative of the population distribution of consumption.. The `RANK` command creates a new variable `perrank`, which indicates the percentile location of each observation (child) in the distribution of equivalent consumption. This is the variable for the x-axis in the concentration curve chart.

```
WEIGHT BY wt .
RANK VARIABLES=eqcons (A) /PERCENT into peqcons /PRINT=YES
  /TIES=MEAN .
```

The next step is to generate the variable for the y-axis of the chart. In the example below the concentration curve shows inequality in medically attended deliveries in Vietnam. The variable `med_del` is a dummy variable taking a value of 1 if the child was delivered by a medically trained person. The `COMPUTE` statement generates a new variable, `mdelnum`, which indicates the number of weighted cases corresponding to each medically attended delivery. The cases are then ordered by the equivalent consumption variable using the `SORT CASES` command. The `CREATE` command then computes the cumulative sum of the weighted attended delivery variable `mdelnum`. The use and placement of the `SORT` command is crucial. The final `COMPUTE` statement simply expresses the cumulative sum as a percentage of the sum of weighted medically attended deliveries (8695264, obtained through the `DECSRIPTIVES` command), giving the variable `ccurve`. This is the variable for the y-axis.

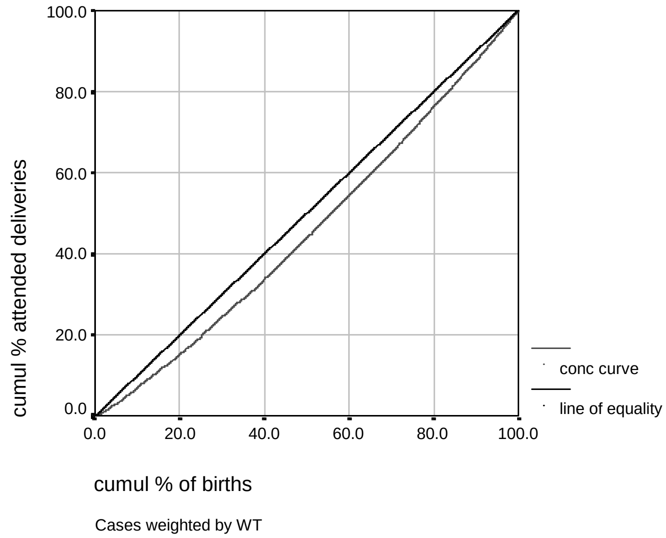
```
compute mdelnum = med_del * wt.
SORT CASES BY eqcons (A).
CREATE /mdelnu_1=CSUM(mdelnum).
DESCRIPTIVES  VARIABLES=mdelnum  /STATISTICS=MEAN SUM STDDEV MIN
MAX .
compute ccurve = (mdelnu_1 / 8695264) * 100.
```

Graphing the concentration curve can be done using the command:

```
GRAPH   /SCATTERPLOT(OVERLAY)=perrank perrank WITH perrank ccurve
(PAIR)  /MISSING=LISTWISE .
```

Double clicking on the resultant graph will allow it to be edited. On the axis options, title the axes, set the ranges to between 1 and 100, and check the grid box. Edit the legends to read ‘line of equality’ and ‘concentration curve’. The smallest possible marker needs to be selected. Some of the options can be saved in a chart template. Fig 2 shows the chart produced by SPSS for these data. The concentration curve reveals a slight tendency for better-off women in Vietnam to have their deliveries attended by medical staff.

Figure 2: Concentration curve for medically attended deliveries in Vietnam, SPSS



SPSS commands for production of grouped data

Another option is simply to write from SPSS an aggregated file containing the concentration curve ordinates for, say, 10 deciles or 25 equal-sized groups. The groups are formed by adding to the `RANK VARIABLES` command above the sub-command `/NTILES (25)`. Once the `perrank` and `ccurve` variables have been generated, the `AGGREGATE` command is used to write out a file containing the x-y ordinates for each group:

```
AGGREGATE  /OUTFILE='C:\fileneame.SAV'  /BREAK=ntiles25
  /peqcon_1 = MAX(perrank) /ccurve_1 = MAX(ccurve).
```

The resultant file can then be saved as an Excel file, and then loaded into Excel for graphing. Or the contents of the file can simply be pasted into Excel. The alternative is to produce a table giving the number of (weighted) cases and assisted deliveries in each of the 25 (or whatever) groups:

```
MEANS TABLES=mdelnum BY ntiles25  /CELLS COUNT  SUM.
```

The resultant output can then be pasted into Excel and the cumulative percentages and chart can be generated.

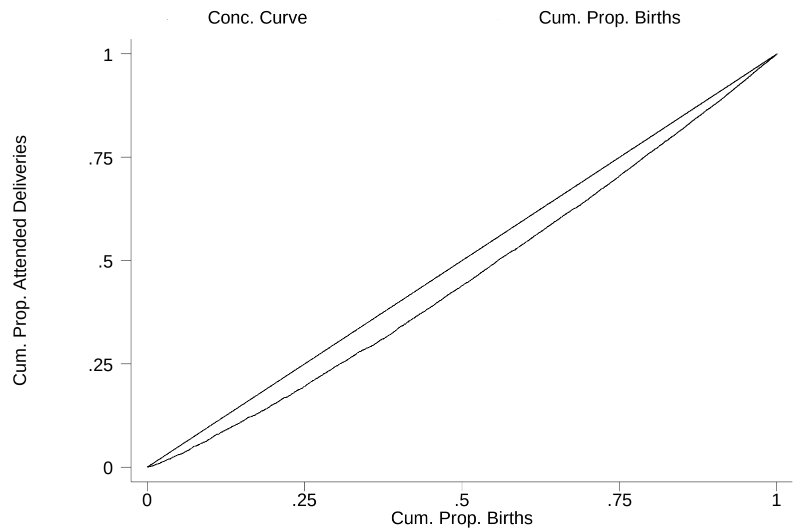
Stata commands

The command `GLCURVE` (a program downloadable from the Stata website) permits easy generation of the ordinates for the concentration curve chart. The commands below are for the medically attended deliveries example above. The `GLCURVE` command below generates a variable `rank`, which indicates the fractional rank in the distribution of births by equivalent consumption (`eqcons`) , as well as a generalized Lorenz ordinate `glmed, for medical attended deliveries`. The `generate` command divides this through by the weighted mean of `med_del` (.733401, produced by the chart generated by the `GLCURVE` command) to get the concentration curve ordinate. The `GRAPH` command then graphs the concentration curve diagram shown in Fig 3.

```
glcurve med_del [fw=wt] , glvar(glmed) pvar(rank) sortvar(eqcons)
      label variable rank "Cum. Prop. Births"
gen ccurve2 = glmed/.733401
label variable ccurve2 "Conc. Curve"
```

```
graph ccurve2 rank rank , sy(..) xscale(0,1) yscale(0,1) ylabel
(0 0.25 to 1) xlabel (0 0.25 to 1) l1(Cum. Prop. Attended
Deliveries)
```

Figure 3: Concentration curve for medically attended deliveries in Vietnam, Stata



Useful links

The concentration index, which is based on the concentration curve, enables health sector inequalities to be quantified— see Technical Note 7 in this series.

Bibliography

1. Wagstaff, A., P. Paci, and E. van Doorslaer, *On the measurement of inequalities in health.* Social Science and Medicine, 1991. **33**: p. 545-557.
2. Kakwani, N.C., A. Wagstaff, and E. Van Doorlsaer, *Socioeconomic inequalities in health: Measurement, computation and statistical inference.* Journal of Econometrics, 1997. **77**(1): p. 87-104.
3. Castro-Leal, F., J. Dayton, L. Demery, and K. Mehra, *Public spending on health care in Africa: do the poor benefit?* Bulletin of the World Health Organization, 2000. **78**(1): p. 66-74.
4. Wagstaff, A., *Socioeconomic inequalities in child mortality: comparisons across nine developing countries.* Bulletin of the World Health Organization, 2000. **78**(1): p. 19-29.
5. Van Doorslaer, E., A. Wagstaff, H. Bleichrodt, S. Caonge, U.-G. Gerdtham, M. Gerfin, J. Geurts, et al., *Income-related inequalities in health: Some international comparisons.* Journal of Health Economics, 1997. **16**: p. 93-112.