The Stata Journal (2016)
16, Number 4, pp. 837–866

Estimating Lorenz and concentration curves

Ben Jann
University of Bern
Bern, Switzerland
ben.jann@soz.unibe.ch

Abstract. Lorenz and concentration curves are widely used tools in inequality research. In this article, I present a new command, `lorenz`, that estimates Lorenz and concentration curves from individual-level data and, optionally, displays the results in a graph. The `lorenz` command supports relative, generalized, absolute, unnormalized, custom-normalized Lorenz, and concentration curves. It also provides tools for computing contrasts between different subpopulations or outcome variables. `lorenz` fully supports variance estimation for complex samples.

Keywords: st0457, lorenz, Lorenz curve, concentration curve, inequality, income distribution, wealth distribution, graphics, survey estimation

1 Introduction

Lorenz and concentration curves are widely used tools for analyzing economic inequality and redistribution (see, for example, Cowell [2011] and Lambert [2001]). Depending on the research question, different variants of the methodology are appropriate:

- The standard (relative) Lorenz curve illustrates the shape of a distribution and the degree of inequality from a relative, scale-free viewpoint. That is, a change in measurement units or proportional growth for all observations does not change the Lorenz curve. Because the Lorenz curve evaluates inequality independently of the overall outcome level, it can assess, for example, whether income inequality increased or decreased over time, how the structure of inequality changed, or how inequality differs across countries.
- A useful variant of the relative Lorenz curve is the equality gap (EG) curve, defined as the difference between the Lorenz curve and the line of perfect equality. It provides an immediate illustration of how a given distribution deviates from an equal distribution situation.
- The relative Lorenz curve depicts how a cake is distributed among a group, but it ignores the actual size of the cake. From the perspective of welfare economics, the size of the cake does matter. The generalized Lorenz (GL) curve is a Lorenz variant sensitive to the overall outcome level. It is designed so that gains in welfare lead to a vertical increase in the curve. Therefore, the GL curve is the best method for analyzing changes in inequality while accounting for the effects of overall growth.
- Just like the EG curve illustrates deviations from equality in terms of the relative Lorenz curve, the absolute Lorenz (AL) curve quantifies the EG curve in terms of

the GL curve. That is, the AL curve illustrates the loss in welfare compared with an equal distribution situation.

- Redistribution is a major concern in inequality research. For example, researchers are interested in how taxes change the shape of income distribution. The concentration curve, another variant of the Lorenz curve, is an important instrument in these analyses, because it identifies the winners and losers of redistribution. Concentration curves can also be used in many other contexts. Essentially, a concentration curve shows how one variable is distributed across groups defined in terms of another variable. For example, concentration curves can analyze how wealth is distributed across income groups or evaluate the degree to which bequests lead to a change in wealth inequality.

- Relative Lorenz and GL curves differ in how they are normalized. Relative Lorenz curves are normalized to 1 (100% of the cake); GL curves are normalized to the average of the outcome variable (the size of a piece of cake if the cake is distributed evenly). Depending on context, it can be useful to omit normalization all together [total Lorenz (TL) curve] or use custom normalization. Custom normalization is useful when comparing subpopulations or multiple outcome variables. For example, it allows expressing results for one group in relation to the outcome level in another group, or it allows expressing results for taxes in terms of percentages of total income.

Despite their frequent use in inequality research literature, Stata does not offer an official command for the estimation of Lorenz and concentration curves. However, several user-written commands do exist. For example, the `glcurve` command (Jenkins and Van Kerm 1999; Van Kerm and Jenkins 2001) can be used to plot generalized, relative Lorenz, or concentration curves, but the command does not provide information on sampling variances. The `svylorenz` command implements variance estimation (Jenkins 2006) for relative and GL curves, but the command does not support the estimation of concentration curves. Further available commands are `clorenz` (Abdelkrim 2005) and `alorenz` (Azevedo and Franco 2006), which both have their specific pros and cons. Yet, because none of these commands provide a comprehensive framework for estimating all variations of Lorenz curves described above, I implemented a new command, `lorenz`. The command computes and, optionally, graphs relative, total (unnormalized), generalized, AL, and concentration curves from individual-level data. `lorenz` has some unique features: for example, it provides standard errors and confidence intervals (CIs) for all variations and fully supports estimation from complex samples. Furthermore, `lorenz` is well suited for subpopulation analysis and offers options to compute contrasts between subpopulations or between outcome variables (including standard errors). In contrast to the other available commands, `lorenz` also offers custom normalization of results. Finally, `lorenz` integrates well with Stata and has been programmed so that it provides the typical features of an estimation command, including conventional `e()` returns for processing by postestimation commands.

In the remainder of the article, I will first discuss the relevant methods and formulas and then present the syntax and options of the `lorenz` command. I will then illustrate the `lorenz` command with examples.

2 Methods and formulas

2.1 Lorenz curve

Let X be the outcome variable of interest (for example, income). The cumulative distribution function of X is given as $F_X(x) = \Pr(X \leq x)$, and the quantile function (the inverse of the distribution function) is given as $Q_X(p) = F_X^{-1}(p) = \inf\{x|F_X(x) \geq p\}$ with $p \in [0, 1]$. For continuous X, the ordinates of the relative Lorenz curve are

$$L_X(p) = \frac{\int_{-\infty}^{Q_X^p} y\, dF_X(x)}{\int_{-\infty}^{\infty} x\, dF_X(x)}$$

(see, for example, Cowell [2000], Lambert [2001], and Hao and Naiman [2010]). Intuitively, a point on the Lorenz curve quantifies the proportion of total outcome of the poorest $p \times 100$ percent of the population. This can easily be seen in the finite population form of $L_X(p)$, which is

$$L_X(p) = \frac{\sum_{i=1}^{N} X_i I\left(X_i \leq Q_X^p\right)}{\sum_{i=1}^{N} X_i}$$

with $I(A)$ as an indicator function being equal to 1 if A is true and 0 otherwise.

Furthermore, let J_i be an indicator for whether observation i belongs to subpopulation j (that is, $J_i = 1$ if observation i belongs to subpopulation j and $J_i = 0$ otherwise). The finite population form of the Lorenz curve of X in subpopulation j in this case is

$$L_X^j(p) = \frac{\sum_{i=1}^{N} X_i I\left(X_i \leq Q_X^{p,j}\right) J_i}{\sum_{i=1}^{N} X_i J_i}$$

where $Q_X^{p,j}$ is the p quantile of X in subpopulation j. One obtains the population-wide Lorenz curve by setting $J_i = 1$ for all observations.

Lorenz curves are typically displayed graphically with p on the horizontal axis and $L_X(p)$ on the vertical axis, although Lorenz (1905) originally proposed an opposite layout.

2.2 Equality gap curve

The EG curve quantifies the degree to which the proportion of total outcome of the poorest $p \times 100$ percent of the population deviates from the proportion of total outcome these population members would get under an equal distribution. That is, the EG curve is equal to the difference between the equal distribution diagonal and the Lorenz curve. Formally, the (finite population form of the) EG curve of Y in subpopulation j is

$$\mathrm{EG}_X^j(p) = p - \frac{\sum_{i=1}^{N} X_i I\left(X_i \leq Q_X^{p,j}\right) J_i}{\sum_{i=1}^{N} X_i J_i} = p - L_X^j(p)$$

$\mathrm{EG}_X(p)$ is equal to the proportion of total outcome that would have to be relocated to the poorest $p \times 100$ percent to provide them an average outcome equal to the subpopulation average.

2.3 Total (unnormalized) Lorenz curve

In the finite population, one can define the (subpopulation specific) TL curve as

$$\mathrm{TL}_X^j(p) = \sum_{i=1}^{N} X_i I\left(X_i \leq Q_X^{p,j}\right) J_i$$

The TL curve quantifies the cumulative sum of outcomes among the poorest $p \times 100$ percent of the subpopulation.

2.4 Generalized Lorenz curve

The ordinates of the relative Lorenz curve refer to cumulative outcome proportions. Hence, $L_X(1) = 1$. In contrast, the ordinates of the GL curve, $\mathrm{GL}_X(p)$, refer to the cumulative outcome average. Hence, $\mathrm{GL}_X(1) = \overline{X}$, where $\overline{X}$ is the mean of X. Formally, one can define the GL curve as

$$\mathrm{GL}_X(p) = \int_{-\infty}^{Q_X^p} x \, dF_X(x)$$

the finite population form of which is

$$\mathrm{GL}_X(p) = \frac{1}{N} \sum_{i=1}^{N} X_i I\left(X_i \leq Q_X^p\right)$$

(see, for example, Shorrocks [1983], Cowell [2000], and Lambert [2001]). Furthermore, for subpopulation j, one can write the GL curve as

$$\text{GL}_X^j(p) = \frac{1}{\sum_{i=1}^{N} J_i} \sum_{i=1}^{N} X_i I\left(X_i \leq Q_X^{p,j}\right) J_i$$

where $\sum_{i=1}^{N} J_i$ is equal to the subpopulation size.

2.5 Absolute Lorenz curve

The AL curve quantifies the degree to which the GL curve deviates from the equal distribution line in terms of the cumulative outcome average (see, for example, Moyes [1987]). Formally, the (finite population form of the) AL curve of Y in subpopulation j is

$$\text{AL}_X^j(p) = \frac{1}{\sum_{i=1}^{N} J_i} \left\{ \sum_{i=1}^{N} X_i I\left(X_i \leq Q_X^{p,j}\right) J_i - p \sum_{i=1}^{N} X_i J_i \right\} = \text{GL}_X^j(p) - p \frac{\sum_{i=1}^{N} X_i J_i}{\sum_{i=1}^{N} J_i}$$

2.6 Concentration curve

The Lorenz curve of outcome variable X refers to cumulative outcome proportions of population members ranked by the values of X. Using an alternative ranking variable Y, while still measuring outcome in terms of X, leads to the so-called concentration curve. Formally, the (relative) concentration curve of X with respect to Y can be defined as

$$L_{XY}(p) = \frac{\int_{-\infty}^{Q_Y^p} \int_{-\infty}^{\infty} x f_{XY}(x,y)\, dx\, dy}{\int_{-\infty}^{\infty} x\, dF_X(x)}$$

where Q_Y^p is the p quantile of the distribution of Y and $f_{XY}(x,y)$ is the density of the joint distribution of X and Y (see, for example, Bishop, Chow, and Formby [1994]). In the finite population, the concentration curve simplifies to

$$L_{XY}(p) = \frac{\sum_{i=1}^{N} X_i I\left(Y_i \leq Q_Y^p\right)}{\sum_{i=1}^{N} X_i}$$

Furthermore, for subpopulation j, the concentration curve can be written as

$$L_{XY}^j(p) = \frac{\sum_{i=1}^{N} X_i I\left(Y_i \leq Q_Y^{p,j}\right) J_i}{\sum_{i=1}^{N} X_i J_i}$$

Total, generalized, or absolute concentration curves can be defined analogously.

2.7 Renormalization

Relative Lorenz curves are normalized with respect to the total of the analyzed outcome variable in the given population or subpopulation. It may be contextually useful to apply a different type of normalization. For example, when analyzing labor income, we may want to express results with respect to total income (labor income plus capital income). Likewise, when analyzing a subpopulation, we may want to express results relative to another subpopulation or relative to the overall population.

To normalize the Lorenz curve or the EG curve of X with respect to the total of Z (where Z may be the sum of several variables, possibly including X), let

$$L_X^{j,Z}(p) = \frac{\sum_{i=1}^{N} X_i I\left(X_i \leq Q_X^{p,j}\right) J_i}{\sum_{i=1}^{N} Z_i J_i} \qquad \text{and} \qquad \text{EG}_X^{j,Z}(p) = p - L_X^{j,Z}(p)$$

Likewise, for normalization with respect to a fixed (subpopulation) total τ, let

$$L_X^{j,\tau}(p) = \frac{\sum_{i=1}^{N} X_i I\left(X_i \leq Q_X^{p,j}\right) J_i}{\tau} \qquad \text{and} \qquad \text{EG}_X^{j,\tau}(p) = p - L_X^{j,\tau}(p)$$

To normalize the Lorenz curve of subpopulation j with respect to the total in subpopulation r (where subpopulation r may include subpopulation j), let

$$L_X^{jr}(p) = \frac{\sum_{i=1}^{N} X_i I\left(X_i \leq Q_X^{p,j}\right) J_i}{\sum_{i=1}^{N} X_i R_i}$$

where R_i is an indicator for whether observation i belongs to subpopulation r or not. For example, if r is the entire population (including subpopulation j), then $L_X^{j,r}(p)$ is the proportion of the population-wide outcome that goes to the poorest $p \times 100$ percent of subpopulation j. In contrast, because the EG curve is supposed to quantify the deviation from the equal distribution line, one should define the renormalized EG curve of subpopulation j with respect to the total in subpopulation r as

$$\text{EG}_X^{jr}(p) = p\frac{\sum_{i=1}^{N} J_i}{\sum_{i=1}^{N} R_i} - L_X^{jr}(p)$$

where $p\sum_{i=1}^{N} J_i / \sum_{i=1}^{N} R_i$ is the outcome share of the poorest $p \times 100$ percent of subpopulation j if all population members would receive the same outcome.

The normalized Lorenz curve ordinates $L_X^{jr}(p)$ express outcome shares in subpopulation j relative to the total outcome of subpopulation r. An alternative is to renormalize

Lorenz curve ordinates in a way such that they are relative to the total that would be observed in subpopulation j, if all members of subpopulation j would receive the average outcome of subpopulation r. This can be achieved by rescaling the total by relative group sizes; that is,

$$L_X^{j\overline{r}}(p) = \frac{\sum_{i=1}^{N} X_i I\left(X_i \le Q_X^{p,j}\right) J_i}{\frac{\sum_{i=1}^{N} J_i}{\sum_{i=1}^{N} R_i} \sum_{i=1}^{N} X_i R_i} \qquad \text{and} \qquad \text{EG}_X^{j\overline{r}}(p) = p - L_X^{j\overline{r}}(p)$$

There is a close relation between $L_X^{j\overline{r}}(p)$ and the GL curves; the ratio of $L_X^{j\overline{r}}(p)$ from two subpopulations is equal to the ratio of the GL curves for these subpopulations.

Combining normalization with respect to a different subpopulation and normalization with respect to the total of a different outcome variable or a fixed total leads to

$$L_X^{jr,Z}(p) = \frac{\sum_{i=1}^{N} X_i I\left(X_i \le Q_X^{p,j}\right) J_i}{\sum_{i=1}^{N} Z_i R_i} \qquad \text{EG}_X^{jr,Z}(p) = p\frac{\sum_{i=1}^{N} J_i}{\sum_{i=1}^{N} R_i} - L_X^{jr,Z}(p)$$

$$L_X^{j\overline{r},Z}(p) = \frac{\sum_{i=1}^{N} X_i I\left(X_i \le Q_X^{p,j}\right) J_i}{\frac{\sum_{i=1}^{N} J_i}{\sum_{i=1}^{N} R_i} \sum_{i=1}^{N} Z_i R_i} \qquad \text{EG}_X^{j\overline{r},Z}(p) = p - L_X^{j\overline{r},Z}(p)$$

and

$$L_X^{jr,\tau}(p) = \frac{\sum_{i=1}^{N} X_i I\left(X_i \le Q_X^{p,j}\right) J_i}{\tau} \qquad \text{EG}_X^{jr,\tau}(p) = p\frac{\sum_{i=1}^{N} J_i}{\sum_{i=1}^{N} R_i} - L_X^{jr,\tau}(p)$$

$$L_X^{j\overline{r},\tau}(p) = \frac{\sum_{i=1}^{N} X_i I\left(X_i \le Q_X^{p,j}\right) J_i}{\frac{\sum_{i=1}^{N} J_i}{\sum_{i=1}^{N} R_i}\tau} \qquad \text{EG}_X^{j\overline{r},\tau}(p) = p - L_X^{j\overline{r},\tau}(p)$$

You can apply analogous renormalizations to concentration curves. Simply replace $I(X_i \le Q_X^{p,j})$ in the above formulas with $I(Y_i \le Q_Y^{p,j})$.

2.8 Contrasts

Analyzing distributional differences is helpful for computing contrasts between Lorenz curves. For example, the difference

$$L_X(p) - L_Y(p)$$

may be used to evaluate whether distribution X Lorenz dominates distribution Y. Likewise, the difference

$$\text{GL}_X(p) - \text{GL}_Y(p)$$

may be used to evaluate whether distribution X GL dominates distribution Y. There is dominance if the difference is positive for all p. As shown by Atkinson (1970), if distribution X Lorenz dominates distribution Y, then distribution X can be seen as more equal than distribution Y under weak conditions. Likewise, if distribution X GL dominates distribution Y, then distribution X can be seen as preferable over distribution Y in terms of welfare under weak conditions (see, for example, Lambert [2001]).

It may be contextually practical to define contrasts as ratios, that is, $L_X(p)/L_Y(p)$, or as logarithms of ratios, that is, $\ln\{L_X(p)/L_Y(p)\}$.

2.9 Estimation procedure

In the preceding sections, I gave various formal definitions of quantities related to Lorenz curves. In this section, I provide a brief description of how we can estimate these quantities and their sampling variances from data.

Given is a sample X_i, $i = 1, \ldots, n$, with sampling weights w_i. Furthermore, let subscripts in parentheses refer to observations sorted in ascending order of X. We can then estimate $L_X(p)$ as

$$\widehat{L}_X(p) = (1-\gamma)\widetilde{X}_{i_p-1} + \gamma\widetilde{X}_{i_p}$$

where

$$\gamma = \frac{p - \widehat{p}_{i_p-1}}{\widehat{p}_{i_p} - \widehat{p}_{i_p-1}}, \quad \widetilde{X}_{i_p} = \frac{\sum_{i=1}^{i_p} w_{(i)}X_{(i)}}{\sum_{i=1}^{n} w_i X_i}, \quad \text{and} \quad \widehat{p}_{i_p} = \frac{\sum_{i=1}^{i_p} w_{(i)}}{\sum_{i=1}^{n} w_i}$$

and where i_p is defined such that $\widehat{p}_{i_p-1} < p \leq \widehat{p}_{i_p}$. Using this approach, we break ties in X proportionally and apply linear interpolation (corresponding to quantile definition 4 in Hyndman and Fan [1996]), where the distribution function of X is flat. Alternatively, we can avoid linear interpolation and estimate $L_X(p)$ as

$$\widehat{L}_X(p) = \widetilde{X}_{i_p} = \frac{\sum_{i=1}^{i_p} w_{(i)}X_{(i)}}{\sum_{i=1}^{n} w_i X_i}$$

(corresponding to quantile definition 1 in Hyndman and Fan [1996]). The first approach appears preferable over the second approach because the second approach requires arbitrary decisions on the sort order within ties of X to obtain stable results in the presence of sampling weights.

We can use analogous formulas to estimate total, generalized, absolute, or renormalized Lorenz curves. For concentration curves, the observations are sorted in order of Y instead of X; to enforce stable results, we can average X values within ties of Y.

Following Binder and Kovacevic (1995) and Kovačević and Binder (1997), we can obtain approximate variance estimates for Lorenz ordinates by estimating the total (see [R] **total**)—perhaps accounting for complex survey design (see [SVY] **svy estimation**)—of residual variables, defined as

$$u_i^j(p) = \frac{\left\{\left(X_i - \widehat{Q}_X^{p,j}\right) I\left(X_i \leq \widehat{Q}_X^{p,j}\right) + p\,\widehat{Q}_X^{p,j}\right\} J_i - a_i}{b}$$

where a_i and b are as described in table 1 (for details, see Jann [2016]).[1] For concentration curves, we can replace $I(X_i \leq \widehat{Q}_X^{p,j})$ with $I(Y_i \leq \widehat{Q}_Y^{p,j})$ and replace $\widehat{Q}_X^{p,j}$ with $\widehat{E}(X|Y = Q_Y^{p,j}, J = 1)$.[2] Furthermore, we can obtain variance estimates for contrasts by the delta method as outlined in Jann (2016).

1. When computing the u variables, the `lorenz` command presented in section 3 uses definition 4 in Hyndman and Fan (1996) to determine $\widehat{Q}_X^{p,j}$ (or definition 1, depending on the method used for estimating the Lorenz ordinates). Furthermore, analogously to the approach used for point estimation, ties in X are broken when determining $I(X_i \leq \widehat{Q}_X^{p,j})$ (based on observations sorted by w_i within ties, which is an arbitrary decision to enforce stable results). Depending on sample design, terms $1/(w_i n)$ and $\tau/(w_i n)$ in the formulas in table 1 require modification; an alternative is to set these terms to zero (see Jann [2016]). Finally, for EG curves, use $-u$ instead of u.
2. In the `lorenz` command presented in section 3, $E(X|Y = Q_Y^{p,j}, J = 1)$, the expected value of X at the p quantile of Y in subpopulation j, is estimated by local linear regression using the Epanechnikov kernel and the default rule-of-thumb bandwidth as described in [R] **lpoly**.

Table 1. Definitions of a_i and b

	a_i	b
$L_X^j(p), \mathrm{EG}_X^j(p)$	$X_i J_i \widehat{L}_X^j(p)$	$\sum_i w_i X_i J_i$
$\mathrm{TL}_X^j(p)$	$\frac{1}{w_i n}\widehat{\mathrm{TL}}_X^j(p)$	1
$\mathrm{GL}_X^j(p)$	$J_i \widehat{\mathrm{GL}}_X^j(p)$	$\sum_i w_i J_i$
$\mathrm{AL}_X^j(p)$	$\left\{\widehat{\mathrm{AL}}_X^j(p) + pX_i\right\} J_i$	$\sum_i w_i J_i$
$L_X^{j,Z}(p), \mathrm{EG}_X^{j,Z}(p)$	$Z_i J_i \widehat{L}_X^{j,Z}(p)$	$\sum_i w_i Z_i J_i$
$L_X^{j,\tau}(p), \mathrm{EG}_X^{j,\tau}(p)$	$\frac{\tau}{w_i n}\widehat{L}_X^{j,\tau}(p)$	τ
$L_X^{jr}(p)$	$X_i R_i \widehat{L}_X^{jr}(p)$	$\sum_i w_i X_i R_i$
$\mathrm{EG}_X^{jr}(p)$	$\left(\frac{\sum_k w_k X_k R_k}{\sum_k w_k J_k} J_i - \frac{\sum_k w_k X_k R_k}{\sum_k w_k R_k} R_i\right) p \frac{\sum_k w_k J_k}{\sum_k w_k R_k} + X_i R_i \widehat{L}_X^{jr}(p)$	$\sum_i w_i X_i R_i$
$L_X^{j\overline{r}}(p), \mathrm{EG}_X^{j\overline{r}}(p)$	$\left(X_i R_i - \frac{\sum_k w_k X_k R_k}{\sum_k w_k R_k} R_i + \frac{\sum_k w_k X_k R_k}{\sum_k w_k J_k} J_i\right) \times \frac{\sum_k w_k J_k}{\sum_k w_k R_k} \widehat{L}_X^{j\overline{r}}(p)$	$\frac{\sum_i w_i J_i}{\sum_i w_i R_i} \sum_i w_i X_i R_i$
$L_X^{jr,\tau}(p)$	$\frac{\tau}{w_i n}\widehat{L}_X^{jr,\tau}(p)$	τ
$\mathrm{EG}_X^{jr,\tau}(p)$	$\left(\frac{\tau J_i}{\sum_k w_k J_k} - \frac{\tau R_i}{\sum_k w_k R_k}\right) p \frac{\sum_k w_k J_k}{\sum_k w_k R_k} + \frac{\tau}{w_i n}\widehat{L}_X^{jr,\tau}(p)$	τ
$L_X^{j\overline{r},\tau}(p), \mathrm{EG}_X^{j\overline{r}}(p)$	$\left(\frac{\tau}{w_i n} - \frac{\tau R_i}{\sum_k w_k R_k} + \frac{\tau J_i}{\sum_k w_k J_k}\right) \frac{\sum_k w_k J_k}{\sum_k w_k R_k} \widehat{L}_X^{j\overline{r},\tau}(p)$	$\frac{\sum_i w_i J_i}{\sum_i w_i R_i}\tau$

(All sums are across the entire sample.)

3 The lorenz command

`lorenz` has three subcommands. `lorenz estimate` computes the Lorenz curve ordinates and their variance matrix; `lorenz contrast` computes differences in Lorenz curve ordinates between outcome variables or subpopulations based on the results by `lorenz estimate`; and `lorenz graph` draws a line graph from the results provided by `lorenz estimate` or `lorenz contrast`.

3.1 Syntax of lorenz estimate

The syntax of `lorenz estimate` is

`lorenz` [`estimate`] *varlist* [*if*] [*in*] [*weight*] [`,`
{`gap`|`sum`|`generalized`|`absolute`} `percent` `normalize(`*spec*`)` `gini`
`nquantiles(`#`)` `percentiles(`*numlist*`)` `pvar(`*pvar*`)` `step` `over(`*varname*`)`
`total` `contrast`[`(`*spec*`)`] `graph`[`(`*options*`)`] `vce(`*vcetype*`)` `cluster(`*clustvar*`)`
`svy`[`(`*subpop*`)`] `nose` `level(`#`)` `noheader` `notable` `nogtable` *display_options*]

`pweight`s, `fweight`s, and `iweight`s are allowed; see [U] **11.1.6 weight**. For each specified variable, `lorenz estimate` tabulates Lorenz curve ordinates along with their standard errors and CIs.[3] If one specifies the `over()` option (see below), only one variable is allowed in *varlist*. `lorenz` assumes subcommand `estimate` as the default; typing the word "`estimate`" is required only in the case of a name conflict between the first element of *varlist* and the other subcommands of `lorenz` (see below). Options are as follows.

Main

Only one instance of `gap`, `sum`, `generalized`, or `absolute` is allowed.

`gap` computes EG curves instead of relative Lorenz curves.

`sum` computes total (unnormalized) Lorenz curves instead of relative Lorenz curves.

`generalized` computes GL curves instead of relative Lorenz curves.

`absolute` computes AL curves instead of relative Lorenz curves.

`percent` expresses results as percentages instead of proportions. `percent` is not allowed in combination with `sum`, `generalized`, or `absolute`.

`normalize(`*spec*`)` normalizes Lorenz ordinates with respect to the specified total (not allowed in combination with `sum`, `generalized`, or `absolute`). *spec* is

[*over*:][*total*] [`, average`]

where *over* may be

`.`	the subpopulation at hand (the default)
#	the subpopulation identified by value #
`#`#	the #th subpopulation
`total`	the total across all subpopulations

3. Variance estimation is not supported for `iweight`s and `fweight`s. To compute standard errors and CIs in the case of `fweight`s, apply `lorenz` to the expanded data (see [D] **expand**).

and *total* may be

`.`	the total of the variable at hand (the default)
`*`	the total of the sum across all analyzed outcome variables
varlist	the total of the sum across the variables in *varlist*
#	a total equal to #

total specifies the variables to compute the total or sets the total to a fixed value. If multiple variables are specified, it uses the total across all specified variables (*varlist* may contain external variables that are not among the list of analyzed outcome variables). *over* selects the reference population to compute the total; *over* is allowed only if you specify the `over()` option (see below). Suboption `average` accounts for subpopulation sizes (sum of weights) so that the results are relative to the average outcome in the reference population; this is relevant only if *over* is present.

`gini` reports the Gini coefficients of the distributions (also known as concentration indices if you specify `pvar()`; see below) to be computed and reported in a separate table.[4]

Percentiles

`nquantiles(#)` specifies the number of (equally spaced) percentiles used to determine the Lorenz ordinates (plus an additional point at the origin). The default is `nquantiles(20)`. This is equivalent to typing `percentiles(0(5)100)`.

`percentiles(`*numlist*`)` specifies, as percentages, the percentiles to compute the Lorenz ordinates. The numbers in *numlist* must be within 0 and 100. You may apply shorthand conventions as described in [U] **11.1.8 numlist**. For example, to compute Lorenz ordinates from 0 to 100% in steps of 1 percentage point, type `percentiles(0(1)100)`. The numbers provided in `percentiles()` do not need to be equally spaced and do not need to cover the whole distribution. For example, to focus on the top 10% and use an increased resolution for the top 1%, type `percentiles(90(1)98 99(0.1)100)`.

`pvar(`*pvar*`)` computes concentration curves with respect to variable *pvar*. That is, it will determine the ordinates of the curves from observations sorted in ascending order of *pvar* instead of the outcome variable (and use average outcome values within ties of *pvar*).

`step` determines the Lorenz ordinates from the step function of the cumulative outcomes. The default is to use linear interpolation in regions where the step function is flat.

Over

`over(`*varname*`)` repeats results for each subpopulation defined by the values of *varname*. Only one outcome variable is allowed if you specify `over()`.

4. Variance estimation for Gini coefficients is not supported.

`total` reports additional overall results across all subpopulations. `total` is allowed only if you specify `over()`.

Contrast/graph

`contrast`[`(`*spec*`)`] computes differences in Lorenz ordinates between outcome variables or between subpopulations. *spec* is

[*base*] [`,` `ratio` `lnratio`]

where *base* is the name of the outcome variable or the value of the subpopulation used as the base for the contrasts. If *base* is omitted, `contrast()` computes adjacent contrasts across outcome variables or subpopulations (or contrasts with respect to the total if total results across subpopulations have been requested).

Use the suboption `ratio` to compute contrasts as ratios, or use the suboption `lnratio` to compute contrasts as logarithms of ratios. The default is to compute contrasts as differences.

`graph`[`(`*options*`)`] draws a line graph of the results. I describe *options* for `lorenz graph` below.

SE/SVY

`vce(`*vcetype*`)` determines how to compute standard errors and CIs. *vcetype* may be

`analytic`
`cluster` *clustvar*
`bootstrap` [`,` *bootstrap_options*]
`jackknife` [`,` *jackknife_options*]

The default is `vce(analytic)`. See [R] **bootstrap** and [R] **jackknife** for *bootstrap_options* and *jackknife_options*, respectively.

`cluster(`*clustvar*`)` is a synonym for `vce(cluster` *clustvar*`)`.

`svy`[`(`*subpop*`)`] accounts for the survey design for variance estimation; see [SVY] **svyset**. Specify *subpop* to restrict survey estimation to a subpopulation, where *subpop* is

[*varname*] [*if*]

The subpopulation is defined by observations for which *varname* $\neq 0$ and for which the `if` condition is met. See [SVY] **subpopulation estimation** for more information on subpopulation estimation.

The `svy` option is allowed only if Taylor linearization is the variance estimation method set by `svyset` (the default). For other variance estimation methods, use the usual `svy` prefix command; see [SVY] **svy**. For example, type `svy brr: lorenz` ... to use balanced repeated-replication variance estimation. The `svy` option is available because `lorenz` does not allow the `svy` prefix for Taylor linearization.

`nose` suppresses the computation of standard errors and CIs. Use the `nose` option to speed up computations, for example, when applying a prefix command that uses replication techniques for variance estimation, such as [SVY] **svy jackknife**. You cannot use the `nose` option together with `vce()`, `cluster()`, or `svy`.

Reporting

`level(#)` specifies the confidence level, as a percentage, for CIs. The default is `level(95)` or as set by `set level`.

`noheader` suppresses the output header; only the coefficient table is displayed.

`notable` suppresses the coefficient table.

`nogtable` suppresses the table containing the Gini coefficients.

display_options are standard reporting options such as `cformat()`, `pformat()`, `sformat()`, or `coeflegend`; see [R] **estimation options**.

3.2 Syntax of lorenz contrast

`lorenz contrast` computes differences in Lorenz ordinates between outcome variables or subpopulations. It requires results from `lorenz estimate` to be in memory, which will be replaced by the results from `lorenz contrast`.[5] The syntax is

`lorenz contrast` [*base*] [`, ratio lnratio graph` [`(`*options*`)`] *display_options*]

where *base* is the name of the outcome variable or the value of the subpopulation used as the base for the contrasts. If you omit *base*, `lorenz contrast` computes adjacent contrasts across outcome variables or subpopulations (or contrasts with respect to the total if total results across subpopulations have been requested). Options are as follows:

`ratio` causes contrasts to be reported as ratios. The default is to report contrasts as differences.

`lnratio` causes contrasts to be reported as logarithms of ratios. The default is to report contrasts as differences.

`graph` [`(`*options*`)`] draws a line graph of the results. I describe *options* for `lorenz graph` below.

display_options are standard reporting options such as `cformat()`, `pformat()`, `sformat()`, or `coeflegend`; see [R] **estimation options**.

5. Alternatively, to compute the contrasts directly, apply the `contrast()` option to `lorenz estimate` (see above).

3.3 Syntax of lorenz graph

`lorenz graph` draws a line diagram of Lorenz curves or Lorenz curve contrasts. It requires results from `lorenz estimate` or `lorenz contrast` to be in memory.[6] The syntax is

`lorenz graph` [`,` `proportion` `nodiagonal` `diagonal(`*line_options*`)` `keep(`*list*`)`
`prange(`*min max*`)` `gini(%`*fmt*`)` `nogini` *connect_options*
`labels("`*label1*`" "`*label2*`" ...)` `byopts(`*byopts*`)` `overlay` `o#(`*options*`)` `level(#)`
`ci(`*citype*`)` `ciopts(`*area_options*`)` `noci` `addplot(`*plot*`)` *twoway_options*]

Options are as follows.

Main

`proportion` scales the population axis as a proportion (0 to 1). The default is to scale the axis as a percentage (0 to 100).

`nodiagonal` omits the equal distribution diagonal included by default for graphing relative Lorenz or concentration curves. There is no equal distribution diagonal included for graphing EG, total, generalized, and AL curves. There is also no equal distribution diagonal for graphing contrasts.

`diagonal(`*line_options*`)` affects the rendition of the equal distribution diagonal, and *line_options* are as described in [G-3] ***line_options***.

`keep(`*list*`)` selects and orders the results to be included as separate subgraphs, where *list* is a list of the names of the outcome variables or the values of the subpopulations to be included. *list* may also contain `total` for the overall results if requested. Furthermore, you may use elements such as `#1`, `#2`, `#3`, etc., to refer to the 1st, 2nd, 3rd, etc., outcome variable or subpopulation.

`prange(`*min max*`)` restricts the range of the points to be included in the graph. It omits points whose abscissas lie outside *min* and *max*. *min* and *max* must be within [0,100]. For example, to include only the upper half of the distribution, type `prange(50 100)`.

`gini(%`*fmt*`)` sets the format for the Gini coefficients included in the subgraph or legend labels; see [D] **format**. The default is `gini(%9.3g)`. `gini()` includes Gini coefficients only if information on Gini coefficients is available in the provided results (that is, if you apply the `gini` option to `lorenz estimate`).

`nogini` suppresses the Gini coefficients. This is relevant only if you specify the `gini` option when calling `lorenz estimate`.

6. You may draw the graph directly using the `graph()` option on `lorenz estimate` or `lorenz contrast` (see above).

Labels/Rendering

connect_options affect the rendition of the plotted lines; see [G-3] ***connect_options***.

`labels("`*label1*`" "`*label2*`" ...)` specifies custom labels for the subgraphs of the outcome variables or subpopulations.

`byopts(`*byopts*`)` determines how subgraphs are combined; see [G-3] ***by_option***.

`overlay` includes results from multiple outcome variables or subpopulations in the same plot instead of creating subgraphs.

`o`#`(`*options*`)` affects the rendition of the line of the #th outcome variable or subpopulation if you specify `overlay`. For example, type `o2(lwidth(*2))` to increase the line width for the second outcome variable or subpopulation. *options* are the following:

connect_options	rendition of the plotted line (see [G-3] ***connect_options***)
[`no`]`ci`	whether to draw the CI
`ciopts(`*area_options*`)`	rendition of the CI (see below)

CIs

`level(`#`)` specifies the confidence level, as a percentage, for CIs. The default is the level used for computing the `lorenz estimate` results. `level()` cannot be used together with `ci(bc)`, `ci(bca)`, or `ci(percentile)`.

`ci(`*citype*`)` chooses the type of CIs to be plotted for results computed using the bootstrap technique. *citype* may be `normal` (normal-based CIs, the default), `bc` (bias-corrected [BC] CIs), `bca` (BC and accelerated CIs), or `percentile` (percentile CIs). `bca` is available only if you request BC_a CIs when running `lorenz estimate` (see [R] **bootstrap**).

`ciopts(`*area_options*`)` affects the rendition of the plotted confidence areas. *area_options* are as described in [G-3] ***area_options***.

`noci` omits CIs from the plot.

Add plots

`addplot(`*plot*`)` adds other plots to the generated graph; see [G-3] ***addplot_option***.

Y axis, X axis, Title, Caption, Legend, Overall

twoway_options are general twoway options, other than `by()`, as documented in [G-3] ***twoway_options***.

4 Examples

4.1 Basic application

By default, `lorenz` computes relative Lorenz curves on a regular-grid of 20 equally spaced points across the population (plus a point at the origin). The following example shows the results for wages in the 1988 extract of the U.S. National Longitudinal Study of Young Women data, shipped with Stata:

```
. sysuse nlsw88
(NLSW, 1988 extract)
. lorenz estimate wage
L(p)                              Number of obs   =      2,246
```

wage	Coef.	Std. Err.	[95% Conf.	Interval]
0	0	(omitted)		
5	.015106	.0004159	.0142904	.0159216
10	.0342651	.0007021	.0328882	.035642
15	.0558635	.0010096	.0538836	.0578434
20	.0801846	.0014032	.0774329	.0829363
25	.1067687	.0017315	.1033732	.1101642
30	.1356307	.0021301	.1314535	.1398078
35	.1670287	.0025182	.1620903	.171967
40	.2005501	.0029161	.1948315	.2062687
45	.2369209	.0033267	.2303971	.2434447
50	.2759734	.0037423	.2686347	.2833121
55	.3180215	.0041626	.3098585	.3261844
60	.3633071	.0045833	.3543191	.372295
65	.4125183	.0050056	.4027021	.4223345
70	.4657641	.0054137	.4551478	.4763804
75	.5241784	.0058003	.5128039	.5355529
80	.5880894	.0062464	.5758401	.6003388
85	.6577051	.0066148	.6447333	.6706769
90	.7346412	.0068289	.7212497	.7480328
95	.8265786	.0062687	.8142856	.8388716
100	1	.	.	.

The standard errors for the first point and the last point are 0 because these Lorenz ordinates are 0 and 1 by definition. This is why Stata flags the first point as "omitted" and prints "missing" for the standard error and CI of the last point.

To graph the estimated Lorenz curve, type (see figure 1):

```
. lorenz graph, aspectratio(1) xlabel(, grid)
```

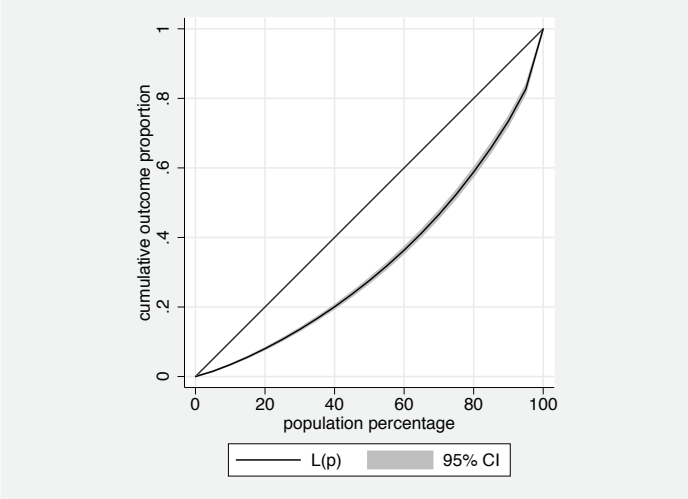

Figure 1. Lorenz curve of wages

The `aspectratio(1)` option enforces a square plot region, and `xlabel(, grid)` includes vertical grid lines.

The default for `lorenz` is a regular grid of evaluation points across the whole distribution. To use an irregular grid or to cover just a part of the distribution, use the `percentiles()` option. The following example focuses on the upper 20% and uses a different step size toward the tail of the distribution (figure 2).

```
. lorenz estimate wage, percentiles(80(2)94 95(1)100)
L(p)                                Number of obs   =      2,246
```

wage	Coef.	Std. Err.	[95% Conf.	Interval]
80	.5880894	.0062464	.5758401	.6003388
82	.6151027	.0063755	.6026003	.6276051
84	.6432933	.0065449	.6304586	.6561281
86	.672506	.006651	.6594633	.6855487
88	.70278	.0067917	.6894613	.7160987
90	.7346412	.0068289	.7212497	.7480328
92	.7684646	.0067952	.7551391	.7817901
94	.806114	.0064727	.7934209	.8188071
95	.8265786	.0062687	.8142856	.8388716
96	.8485922	.0060386	.8367504	.860434
97	.8730971	.0051329	.8630314	.8831629
98	.9046081	.0027287	.899257	.9099591
99	.9486493	.000697	.9472826	.9500161
100	1	.	.	.

```
. lorenz graph, recast(connect) msymbol(diamond)
```

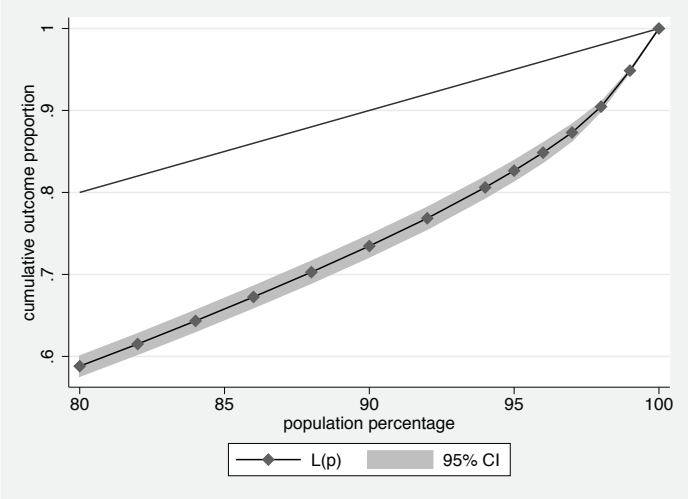

Figure 2. Upper part of the Lorenz curve of wages

The `recast(connect)` option made the evaluation points visible in the graph. The `msymbol(diamond)` option specified diamonds as marker symbols.

4.2 Subpopulation estimation

To compute results for multiple subpopulations, use the `over()` option. I analyzed wages by union status below (see figure 3):

```
. lorenz estimate wage, over(union)
  (output omitted)
. lorenz graph, aspectratio(1) xlabel(, grid)
```

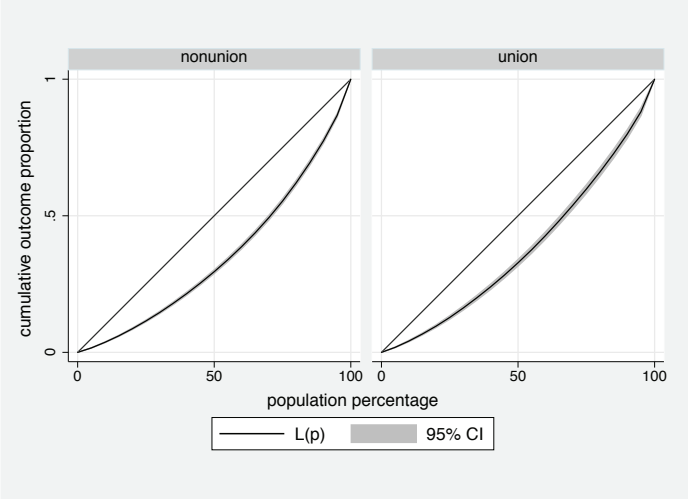

Figure 3. Lorenz curve of wages by union status

By default, `lorenz` places results for the subpopulations in separate subgraphs. To combine the results in a single plot, use the `overlay` option (figure 4):

```
. lorenz graph, aspectratio(1) xlabel(, grid) overlay
```

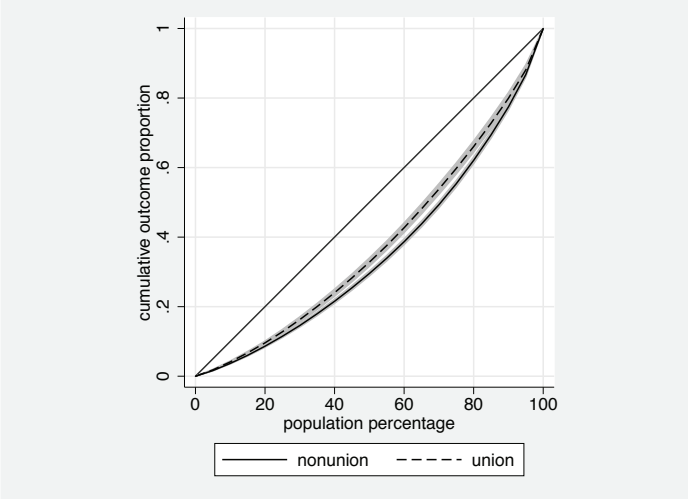

Figure 4. Overlayed Lorenz curves by union status

4.3 Contrasts and Lorenz dominance

A useful feature of `lorenz` is that it can compute contrasts between subpopulations or outcome variables. For example, to evaluate whether the wage distribution of unionized women Lorenz dominates the wage distribution of nonunionized women, type

```
. lorenz estimate wage, over(union)
  (output omitted)
. lorenz contrast 0
L(p)                                          Number of obs     =      1,878
            0: union = nonunion
            1: union = union
```

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
1						
0	0	(omitted)				
5	.0020273	.0009365	2.16	0.031	.0001905	.003864
10	.004292	.0016305	2.63	0.009	.0010942	.0074897
15	.0071636	.0023077	3.10	0.002	.0026376	.0116895
20	.0095728	.0030773	3.11	0.002	.0035375	.0156081
25	.0128985	.0038764	3.33	0.001	.0052959	.020501
30	.017007	.0046414	3.66	0.000	.0079042	.0261097
35	.0207488	.005331	3.89	0.000	.0102935	.031204
40	.024661	.0059814	4.12	0.000	.0129302	.0363918
45	.0284968	.0065591	4.34	0.000	.0156329	.0413607
50	.0326431	.0071647	4.56	0.000	.0185915	.0466948
55	.036453	.0077004	4.73	0.000	.0213506	.0515553
60	.0402741	.0082179	4.90	0.000	.0241569	.0563913
65	.0433946	.0086696	5.01	0.000	.0263914	.0603977
70	.0450269	.0090563	4.97	0.000	.0272654	.0627884
75	.043906	.0093882	4.68	0.000	.0254936	.0623184
80	.0397601	.009565	4.16	0.000	.021001	.0585193
85	.0334832	.0096968	3.45	0.001	.0144655	.0525008
90	.0248836	.0094742	2.63	0.009	.0063025	.0434646
95	.013423	.0083609	1.61	0.109	-.0029747	.0298208
100	0	(omitted)				

```
(difference to union = 0)
. lorenz graph
```

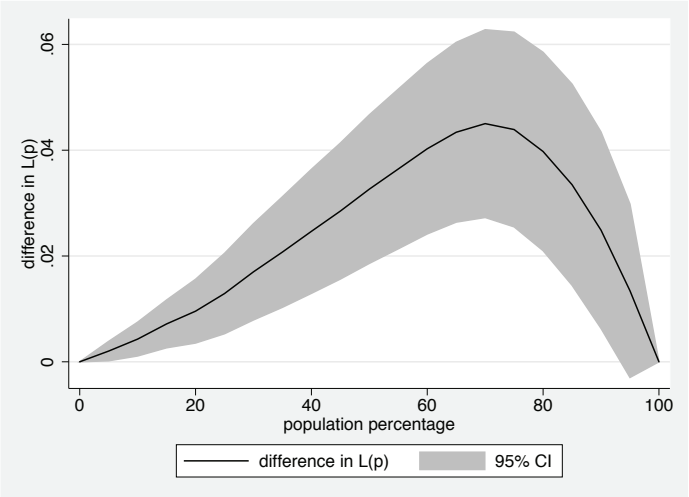

Figure 5. Difference in Lorenz curves by union status

The Lorenz curve for unionized women's wages lies above the Lorenz curve for nonunionized women's wages (see figure 5). One can conclude that the wage distribution for nonunionized women is less equal than the wage distribution for unionized women.

Lorenz dominance does not necessarily imply that one distribution is preferable over the other from a welfare perspective. To evaluate welfare ordering, one may find it useful to analyze GL dominance. The following example shows the GL curves of wages of unionized and nonunionized women (figure 6):

```
. lorenz estimate wage, over(union) generalized
  (output omitted)
. lorenz graph, overlay
```

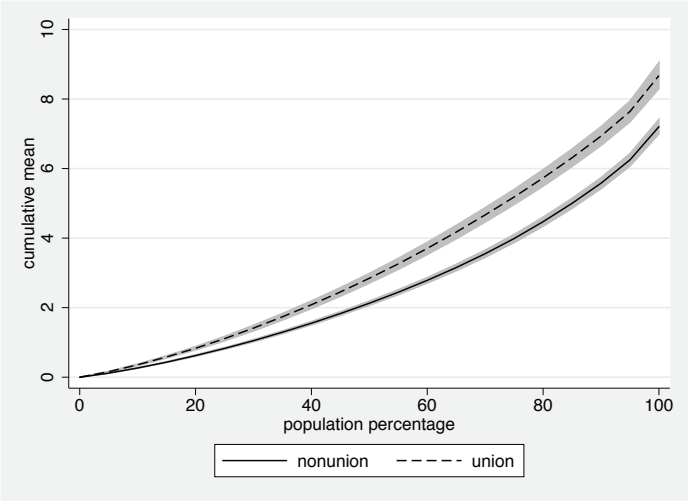

Figure 6. Generalized Lorenz curves by union status

To evaluate whether one distribution dominates the other, we can again take contrasts (figure 7):

```
. lorenz contrast 0
  (output omitted)
. lorenz graph
```

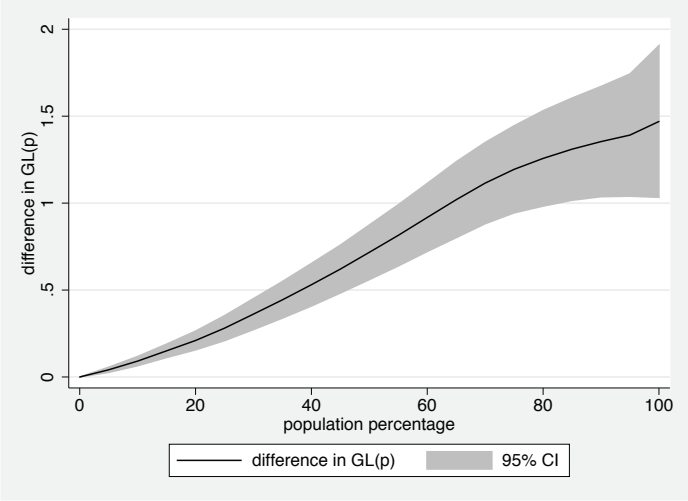

Figure 7. Difference in GL curves by union status

These results clearly show that the wage distribution for unionized women GL dominates the wage distribution for nonunionized women. Not only is the wage distribution for unionized women less unequal than the wage distribution for nonunionized women, it is also clearly preferable from a welfare perspective.

4.4 Concentration curves

Concentration curves illustrate how one variable is distributed across the population, ranked by another variable. As an example, consider the following household dataset with information on transfer income, capital income, and earnings. We may use the `pvar()` option to analyze how transfers and capital rents are distributed across households, while ranking households by earnings:

```
. use lorenz_exampledata
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
earnings	6,007	94673.66	72285.96	0	1965925
capincome	6,007	8346.58	50226.91	0	2374227
transfers	6,007	6375.284	15838.54	0	239408
couple	6,007	.4453138	.4970418	0	1

```
. lorenz estimate transfers capincome, pvar(earnings)
```

(*output omitted*)

```
. lorenz graph, aspectratio(1) xlabels(, grid) overlay legend(cols(1))
> ciopts(recast(rline) lpattern(dash))
```

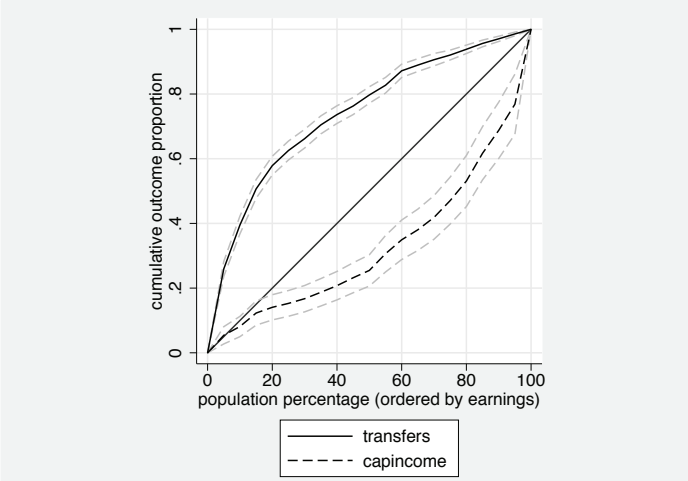

Figure 8. Concentration curves of transfers and capital income

We see, as expected, that the concentration curve for transfers lies above the equal distribution line (figure 8). That is, transfers benefit households with low earnings. For example, the bottom 50% of households in the earnings distribution receive about 80% of all transfers. Capital income is skewed toward high-earning households (the bottom 50% of households in the earnings distribution receive only about 25% of all capital income).

4.5 Renormalization

By default, Lorenz and concentration curves are normalized so that the last ordinate is equal to one, because 100% of the population possesses 100% of the outcome sum. When one analyzes subpopulations or multiple outcome variables, however, it may be useful to apply a different type of normalization with the `normalize()` option.

For example, the `normalize()` option can be used to subdivide the Lorenz curve of total income by income source as follows:

```
. generate totalinc = earnings + capincome + transfers
. generate earntrans = earnings + transfers
. lorenz estimate totalinc earntrans transfers, pvar(totalinc) normalize(totalinc)
  (output omitted)
. lorenz graph, aspectratio(1) xlabels(, grid) overlay
> labels("total income" "tansfers+earnings" "transfers")
> legend(position(3) stack cols(1))
```

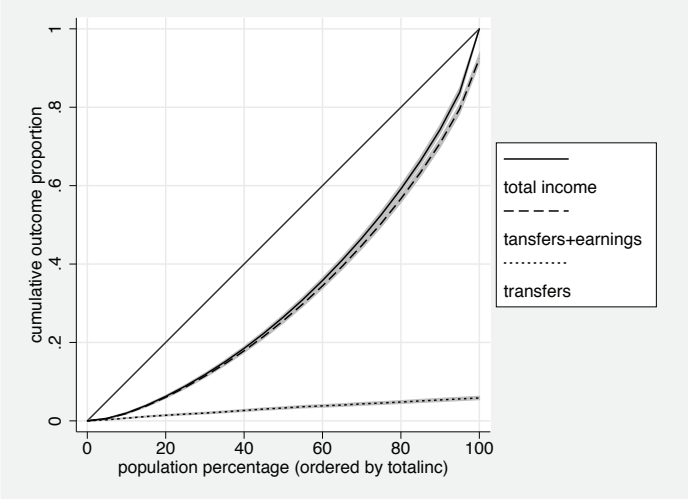

Figure 9. Lorenz curve subdivided by income source

The trick is to generate a series of variables accumulating the income sources step by step and then normalize results with respect to the variable containing the sum of all income sources. Furthermore, I specified the `pvar(totalinc)` option so that the same ordering of observations is used for all variables. The results are shown in figure 9. The bottom curve displays the part of cumulative total income due to transfers. The area between the bottom curve and the middle curve depicts the contribution of earnings. The area between the middle curve and the upper curve captures the contribution of capital income.

The data used in the last example come from two different types of tax subjects: single-person tax subjects and married couples. When analyzing income inequality among singles and couples, we might want to account for the different income levels of the two groups. The following example shows how to compute Lorenz curves for the two groups that are both normalized with respect to the same reference group.

```
. lorenz estimate totalinc, over(couple) normalize(1:, average)
  (output omitted)
. lorenz graph, aspectratio(1) xlabels(, grid) overlay
```

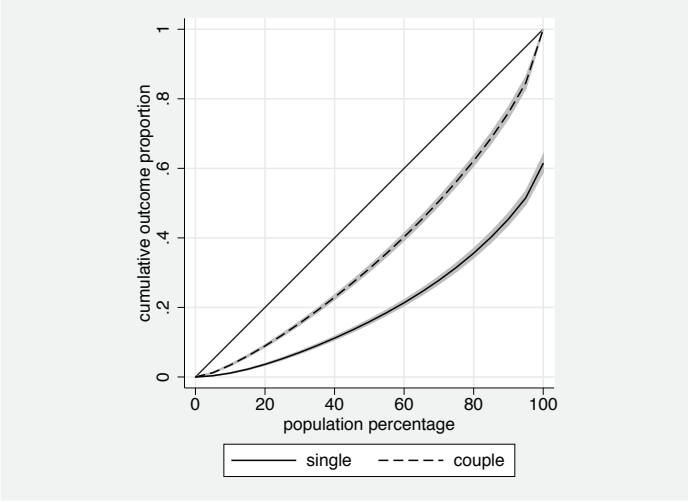

Figure 10. Renormalized Lorenz curves by type of tax subject

Expression "`1:`" in `normalize()` specifies that the subpopulation identified by value 1 (couples) be used as the reference group; suboption `average` implies normalization in terms of group averages instead of group totals (without the `average` suboption, the results would be affected by the group sizes). In figure 10, we see that on average, income declared by singles amounts to about 60% of income declared by couples (the rightmost ordinate of the renormalized Lorenz curve of singles is equal to about 0.6). We also see that the curve for singles lies everywhere below the Lorenz curve of couples; that is, the poorest x% of singles are always poorer than the poorest x% of couples.

Finally, the following example illustrates how to apply renormalization to both variables and subpopulations.

```
. lorenz estimate transfers, over(couple) pvar(earnings)
> normalize(total:earnings, average)
  (output omitted)
. lorenz graph, nodiagonal overlay
```

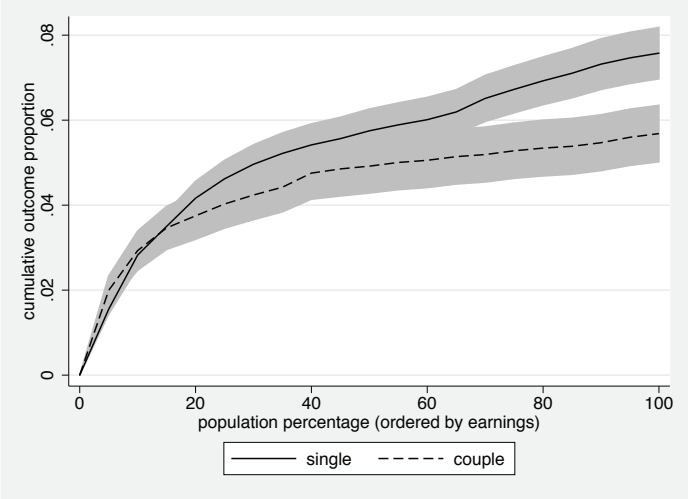

Figure 11. Renormalized concentration curves of transfers

In the above example, I analyze the relationship between transfers and labor income (earnings). The `normalize()` option normalizes all results with respect to the average of variable `earnings` in the overall population (expression "`total:`" selects the total population as the reference group). The results, displayed in figure 11, indicate that couples, on average, receive transfers in the range of a bit more than 5.5% of the population average of earnings; singles receive somewhat higher transfers (7.5% of the population average of earnings). For both groups, transfers are more concentrated among the poor; that is, the size of transfers decreases with earnings (both curves have a decreasing slope). The decrease in transfers, however, is more rapid for couples than for singles.

5 Acknowledgment

This research has been supported by the Swiss National Science Foundation (Grant No. 143399).

6 References

Abdelkrim, A. 2005. clorenz: Stata module to estimate Lorenz and concentration curves. Statistical Software Components S456515, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s456515.html.

Atkinson, A. B. 1970. On the measurement of inequality. *Journal of Economic Theory* 2: 244–263.

Azevedo, J. P., and S. Franco. 2006. alorenz: Stata module to produce Pen's Parade, Lorenz and Generalised Lorenz curve. Statistical Software Components S456749, Department of Economics, Boston College.
http://ideas.repec.org/c/boc/bocode/s456749.html.

Binder, D. A., and M. S. Kovacevic. 1995. Estimating some measures of income inequality from survey data: An application of the estimating equations approach. *Survey Methodology* 21: 137–145.

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: 479–488.

Cowell, F. A. 2000. Measurement of inequality. In *Handbook of Income Distribution, vol. 1*, ed. A. B. Atkinson and F. Bourguignon, 87–166. The Netherlands: Elsevier.

———. 2011. *Measuring Inequality*. 3rd ed. Oxford: Oxford University Press.

Hao, L., and D. Q. Naiman. 2010. *Assessing Inequality*. Thousand Oaks, CA: Sage.

Hyndman, R. J., and Y. Fan. 1996. Sample quantiles in statistical packages. *American Statistician* 50: 361–365.

Jann, B. 2016. Assessing inequality using percentile shares. *Stata Journal* 16: 264–300.

Jenkins, S. P. 2006. svylorenz: Stata module to derive distribution-free variance estimates from complex survey data, of quantile group shares of a total, cumulative quantile group shares. Statistical Software Components S456602, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s456602.html.

Jenkins, S. P., and P. Van Kerm. 1999. sg107: Generalized Lorenz curves and related graphs. *Stata Technical Bulletin* 48: 25–29. Reprinted in *Stata Technical Bulletin Reprints*, vol. 8, pp. 274–278. College Station, TX: Stata Press.

Kovačević, M. S., and D. A. Binder. 1997. Variance estimation for measures of income inequality and polarization—The estimating equations approach. *Journal of Official Statistics* 13: 41–58.

Lambert, P. J. 2001. *The Distribution and Redistribution of Income*. 3rd ed. Manchester: Manchester University Press.

Lorenz, M. O. 1905. Methods of measuring the concentration of wealth. *Journal of the American Statistical Association* 9: 209–219.

Moyes, P. 1987. A new concept of Lorenz domination. *Economics Letters* 23: 203–207.

Shorrocks, A. F. 1983. Ranking income distributions. *Economica* 50: 3–17.

Van Kerm, P., and S. P. Jenkins. 2001. Generalized Lorenz curves and related graphs: An update for Stata 7. *Stata Journal* 1: 107–112.

About the author

Ben Jann is a professor of sociology at the University of Bern, Switzerland. His research interests include social science methodology, statistics, social stratification, and labor market sociology. Recent publications include articles in *Sociological Methodology*, *Sociological Methods and Research*, the *Stata Journal*, *Public Opinion Quarterly*, and the *American Sociological Review*.