University of Illinois — Department of Economics
Fall 2012 — Econ 508 — Roger Koenker

Lecture 19
"Inference about tail-behavior and measures of inequality"

An important topic, which is only rarely addressed in econometrics courses, is the measurement of inequality. This is a large topic which could easily occupy us for several weeks. I plan more of a surgical strike rather than an extended siege on the topic.

A standard model for size distributions in economics, and beyond, is the Pareto distribution

$$F(x) = 1 - (\xi/x)^{\alpha} \qquad x \geq \xi$$

which is also sometimes called Zipf's law. See Hill (1974) for an interesting discussion of how such a distribution might arise. There are many examples of applications in economics: distribution of incomes and size distributions of firms being only the most widely studied.[1]

There is a famous, or perhaps infamous would be more accurate, book by Zipf (1949) called *Human Behavior and the Principle of Least Effort* which offers a vast panoply of examples of the applicability of the Pareto law, including examples in linguistics, music and demography. Hill (1974) offers an interpretation in terms of the so-called Bose-Einstein model in which balls are allocated to cells in such a way that, given the current allocation, the probabilities of allocation to the various cells are proportional to the number of balls currently occupying each cell, i.e., growth proportional to current size. This is a model which has received considerable attention in the IO literature on models of firm growth. There is a recent review of the

[1] A wider class of models for the upper tail of size distributions consists of distributions with *regularly varying tails*, that is distributions whose tails behave like the Pareto distribution. For example, the Cauchy distribution has df, $1 - F(x) \sim (\pi x)^{-1}$ so its tail behaves very much like the Pareto with parameter $\alpha = 1$. More formally, a function F is regularly varying with index α if for $x > 0$,

$$\lim_{t \to \infty} \frac{F(tx)}{F(t)} = x^{\alpha},$$

if $\alpha = 0$ then we say that F is slowly varying. Resnick (2007) provides a thorough introduction to this line of inquiry, including extensive applications in finance.

firm growth literature where a variant of the Bose-Einstein model is called Gibrat's Law, by Sutton (1997). An interesting application which would be fun to explore as a thesis topic is the application of these methods to a comparison of the productivity of research in various academic fields over the last century. Parzen(1985) has suggested that "economics is becoming more scientific" on the basis that the tail exponent of its productivity distribution has decreased in recent years. Unfortunately, I've never been able to track done a reliable reference for this observation.

The Pareto model offers a simple means of measuring inequality by looking simply at the tail exponent α. The Pareto distribution is said to have algebraic tails, since the tails decline algebraically rather than exponentially as for the Gaussian, or exponential distributions.

MLE estimation of α.

$$
\begin{aligned}
f(x) &= \alpha(\xi/x)^{\alpha-1}\xi/x^2 \\
&= \alpha\xi^\alpha x^{-(\alpha+1)} \\
\log f(x) &= \log\alpha + \alpha\log\xi - (\alpha+1)\log x \\
\ell_n(\alpha) &= n\log\alpha + n\alpha\log\xi - (\alpha+1)\sum_{i=1}^n \log x_i \\
\nabla\ell_n(\alpha) &= \frac{n}{\alpha} + n\log\xi - \sum \log x_i \\
\nabla\ell_n(\alpha) &= 0 \Rightarrow \hat{\alpha} = (n^{-1}\sum_{i=1}^n \log(x_i/\xi))^{-1}
\end{aligned}
$$

QMLE estimation of α

Often we are unwilling to make a commitment to a global model of the size distribution, but might be willing to make inferences about only the upper tail of the distribution. Here, Hill (1975), comes to the rescue.

Suppose we think that the Pareto model is adequate for $x > \xi$, but don't necessarily believe it is appropriate below ξ. Alternatively, as is frequently the case, we may only have data for $x > \xi$ (the biggest firms, for example) and don't want to be bothered by the smaller ones. Hill proposes to construct random variables,

$$V_i = \log Y^{(i)} - \log Y^{(i+1)}$$

where $Y^{(i)}$ is the i^{th} reverse-order statistic, i.e., $Y^{(1)} = Y_{(n)}, Y^{(2)} = Y_{(n-1)}$, etc. Now, choose r such that $Y^{(r+1)} \geq \xi$ and compute

$$\hat{\alpha}_r = (r^{-1} \sum_{i=1}^{r} iV_i)^{-1}$$

Note that setting $y_i = \log Y^{(i)}$, we have[†]

$$\begin{aligned} \sum_{i=1}^{r} iV_i &= (y_1 - y_2) + 2(y_2 - y_3) + 3(y_3 - y_4) + \ldots + r(y_r - y_{r+1}) \\ &= y_1 + y_2 + y_3 + \ldots + y_r - ry_{r+1} \\ &= \sum_{i=1}^{r} \log(Y^{(i)}/Y^{(r+1)}) \end{aligned}$$

so $\hat{\alpha}_r$ is the MLE, conditional on only the first r (largest) order statistics. The theory of this is quite elegant and is based on a nice representation of the order statistics by Renyi. Choosing r is somewhat tricky, and is like choosing lag lengths or bandwidths for some other problems. One strategy is to compute $\hat{\alpha}_r$ for several r's and try to find a value which "stabilizes the estimate" – whatever that means.

[†]This is sometimes referred to as "summation-by-parts" for obvious reasons. Well, perhaps not *entirely* obvious, is maybe it is worth elaborating. Recall that the usual differentiation of products formula

$$Duv = uDv + vDu$$

yields the rule for integration by parts,

$$\int uDv = uv - \int vDu.$$

For summation we can do something similar, suppose u and v are defined on a discrete grid then,

$$\begin{aligned} \Delta(u(x)v(x)) &= u(x+1)v(x+1) - u(x)v(x) \\ &= u(x+1)v(x+1) - u(x)v(x+1) \\ &\quad +u(x)(v(x+1) - u(x)v(x) \\ &= u(x)\Delta v(x) + v(x+1)\Delta u(x). \end{aligned}$$

Note the mildly annoying shift of $v(x+1)$ needed to make this work out nicely. If we write $Sv(x) \equiv v(x+1), S$ for (forward) shift operator, then we can write the summation formula as

$$\sum u\Delta v = uv - \sum Sv\Delta u.$$

This is particularly simple when we have, as in the present case, since $u(x) = x$ so $\Delta u = 1$.

Now one might imagine having several samples at different time periods, for example, and one could compute estimates of $\hat{\alpha}$ for the various periods and compare, thus judging whether the distribution was becoming more or less concentrated. The tail behavior of asset returns has been a very controversial topic in finance since early work by Mandelbrot suggested algebraic tails might be an appropriate model. See, for example, the recent paper by McCulloch (1997) for an introduction to this literature.

On the Renyi Representation Result

Thm (Renyi) (1953) Let $\{Z_i\}$ be iid from F, f with $F(0) = 0$, and $Z^{(1)} \geq Z^{(2)} \geq \ldots \geq Z^{(n)}$ be the (reversed) order statistics, then

$$Z^{(i)} = F^{-1}\left(\exp\left(-\frac{e_1}{k} - \frac{e_2}{k-1} - \cdots - \frac{e_i}{k-i+1}\right)\right) \qquad \text{for } i = 1, \ldots, k$$

where e_i are iid exponential variates with mean 1.

Cor Inverting (solving for e_i) we have

$$e_j = (k - j + 1)[\log F(Z^{(j-1)}) - \log F(Z^{(j)})] \qquad j = 1, \ldots, k$$

where by definition $F(Z^{(0)}) = 1$.

Remark Since $F(Z) \sim U$ and $\log U \sim e$ all of this makes a certain amount of sense. It is also a fundamental result in the theory of rank statistics. It also has useful connections to auction theory.

Relationship to Gini coefficient and the Lorenz Curve

Another well known device for comparing measures of inequality is the Lorenz curve. It is usually described as a plot of the cumulative income earned by the poorest proportion, τ of the population. More formally, we may write,

$$\lambda(\tau) = \int_0^\tau F^{-1}(t)dt / \int_0^1 F^{-1}(t)dt$$

The function $\lambda(t)$ is clearly convex since it is the integral of a monotonic function. Several Lorenz curves for the Pareto distribution are illustrated below in Figure 1.

To the extent that the region between the curve and the 45 degree line is large the distribution F deviates from point mass one at some income.

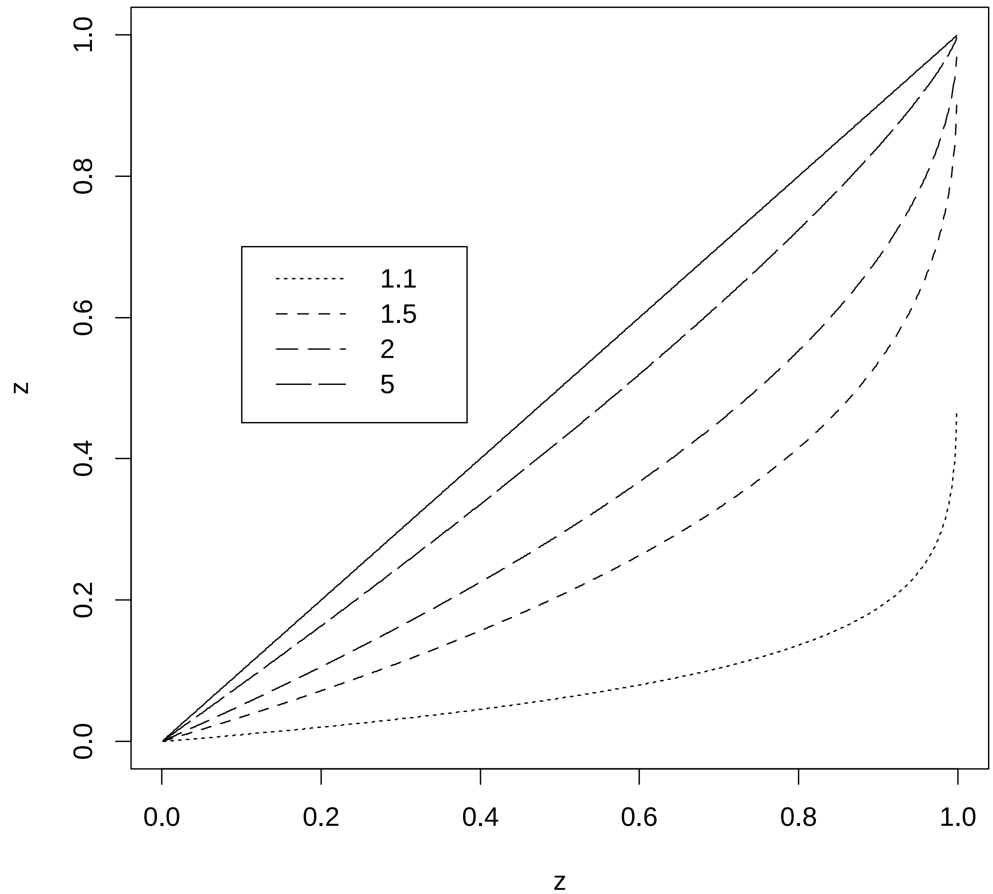

Figure 1: Several Lorenz Curves for the Pareto Distribution: The figure illustrates the Lorenz function for several different tail exponents of the Pareto distribution. The curves plots the cumulative proportion of income earned (y-axis) as a function of the cumulative proportion of the population ordered from poorest to richest.

A measure of departure from egalitarianism is therefore the twice the area between the curve and the diagonal. This is the Gini index (coefficient),

$$\gamma = 1 - 2\int_0^1 \lambda(t)dt.$$

In this form, the Gini coefficient is a measure of dispersion scaled to lie between zero and one. If this distribution, F, is degenerate at μ, then $\gamma = 0$. At the other extreme, if F puts mass $1/\mu$ at μ^2 and mass $(1 - 1/\mu)$ at 0, then as $\mu \to \infty, \gamma \to 1$.

There are a number of other interesting ways to express γ. Another way to express the geometric region represented by γ, i.e., double the region between the Lorenz curve and the 45 degree line in the figure, is to write

$$\gamma = \int_0^1 td\lambda(t) - \int_0^1 \lambda(t)dt.$$

This is simply the area of the region above the curve $\lambda(t)$ in the figure, minus the area of the region below. The area below the curve is clearly just $\int \lambda$, the area above the curve may be found by viewing the picture from the opposite side of the page and rotating it by 90 degrees. We are then integrating the function t with respect to the "density" $d\lambda(t)$ and we obtain the area above the curve in the original picture. This may give some geometric insight into integration by parts, since

$$\int_0^1 td\lambda + \int_0^1 \lambda dt = t\lambda(t)|_0^1 = 1,$$

may be seen as simply adding the areas above and below the curve $\lambda(t)$ in the unit square.

0.1 Total Time on Test (Read at your own Risk)

And this yields, substituting for $\int \lambda$,

$$\begin{aligned}
\gamma &= 2\int_0^1 td\lambda - 1. \\
&= 2\mu^{-1}\int_0^1 tF^{-1}(t)dt - 1 \\
&= 1 - 2\mu^{-1}\int_0^1 (1-t)F^{-1}(t)dt
\end{aligned}$$

(Note that $d\lambda/dt = \mu^{-1}F^{-1}(t)$ provided F is continuous.) The last expression is related to the reliability literature concept of cumulative rescaled total time on test. See Shorack and Wellner (1986, §23.5). Another intriguing expression arises from rewriting the intermediate step above as

$$\gamma = 2\mu^{-1}\int_0^1 tF^{-1}(t)dt - 1 = 2\mu^{-1}\int_{-\infty}^{\infty} x(F(x) - 1/2)dF(x)$$

which Olkin and Yitzhaki (1992) interpret as $\gamma = 2\text{Cov}\ (X, F_X(X))/\mu$ and relate to rank statistics.

0.2 Gini's Mean Difference

Finally, we should note that

$$\gamma = (2\mu)^{-1}\int_0^{\infty}\int_0^{\infty} |x - y|dF(x)dF(y)$$

so we can interpret γ as the ratio of the expected difference in two random draws from F, to the expected sum of the two draws, i.e.,

$$\gamma = \frac{E|X - Y|}{E(X + Y)}$$

This expression suggests that γ is a somewhat more robust alternative to the usual standard deviation as a measure of dispersion. To connect the two we note that,

$$\sigma = (\frac{1}{2}E(X - Y)^2)^{1/2}$$

where X, Y are independent with *df* F. Clearly σ places more weight on large discrepancies between X and Y than does γ. Neither quantity is formally robust in the sense of Hampel.

This version of the Gini coefficient yields a nice interpretation of γ in terms of social choice theory. We can think of γ as measuring the difference in income (wealth, etc.) of two randomly selected individuals selected from the population. This has some appeal on utilitarian grounds.

Example

The Pareto distribution provides a convenient example in which all the calculations are very simple. We have

$$\begin{aligned} F(X) &= 1 - (\xi/x)^{\alpha} \\ f(x) &= \alpha\xi^{\alpha}/x^{\alpha+1} \end{aligned}$$

so provided $\alpha = 1$,

$$\mu = \alpha\xi/(\alpha - 1).$$

The quantile function is

$$F^{-1}(t) = \xi(1-t)^{-1/\alpha}$$

and

$$\lambda(t) = 1 - (1-t)^{(\alpha-1)/\alpha}$$

so

$$\gamma = 1 - 2\int_0^1 \lambda(t)dt = 1/(2\alpha - 1).$$

We may also consider what happens when incomes are Pareto, but we choose to measure inequality by Gini on another scale. A simple, yet practically important, example involves the case in which, $Y = \log X$, that is, we measure inequality on the log income scale. If X has the Pareto distribution so,

$$F_X(x) = 1 - (\xi/x)^{\alpha} \qquad x \geq \xi$$

then

$$F_Y(y) = 1 - \xi^{\alpha}e^{-\alpha y} \qquad y \geq \log\xi$$

so Y has an exponential distribution, with mean

$$\mu = (1 + \alpha\log\xi)/\alpha$$

and the quantile function is

$$Q_Y(t) = \log\xi - \alpha^{-1}\log(1-t).$$

Thus the Lorenz curve for Y is

$$\lambda_Y(t) = \mu^{-1}\int_0^t (\log\xi - \alpha^{-1}\log(1-s))ds$$

and consequently the Gini for log income is,

$$\gamma_Y = 1 - 2\int_0^1 \lambda_Y(t)dt = \frac{1}{2(1+\alpha\log\xi)}$$

Note that if you consider the question, "how does γ change with α after the log transformation, the situation is actually quite similar to the levels version. In particular, the reader is invited to compare $d\log\gamma/d\alpha$ in the two cases.

Thus we get a nice "closed form" expression for γ and as expected as α *increases* giving a *thinner* tail we have a *smaller*, γ, indicating a more egalitarian distribution. Several examples of the Lorenz curve are illustrated in Figure 1 for different tail exponents of the Pareto distribution.

The approach of Hill can be adapted to the Lorenz-Gini approach to measuring equality. We can condition on only the upper tail of the distribution and reformulate the Lorenz curve and therefore the Gini based on this "censored" portion of the full distribution. This would be appropriate if we either (i) had data for only the upper tail, or (ii) we felt the functional form employed for the Lorenz curve, say the Pareto was only appropriate in this range. This is developed by Sen (1986).

There is a large literature on estimation of the Lorenz curve which essentially about suggesting convenient parametric functional forms for $\lambda(t)$. See Sen (1973) for an overview of the general issues surrounding inequality measurement. There is also a large related literature in IO having to do with measuring concentration of firms in particular markets.

References

Hill, B. (1974) The rank frequency form of Zipf's Law, *JASA*, 69, 1017-1026.

Hill, B. (1975) A simple general approach to inference about the tail of a distribution, *Annals of Stat*, 3, 1163-73.

Olkin, I, and S. Yitzhaki (1992) Gini regression analysis, *Int'l Stat Rev*, 60, 185-196.

McCulloch, J.H. (1997) Measuring tail tickness to estimate the stable index α: A critique, *JBES*, 15, 74-81.

Parzen, E. (1985) Personal Communication.

Resnick, S. (2007) *Heavy Tailed Phenomena*, Springer.

Shorack, G. and J. Wellner (1986) *Empirical Processes*, Wiley.

Sen, P.K. (1986) The Gini coefficient and poverty indexes: Some reconciliations, *JASA*. , 81, 1050-1057.

Sen, A. (1973) *On Economic Inequality*, Oxford U. Press.

Sutton, J. (1997) Gibrat's Legacy, *J. Econ. Lit.*, 35, 40-59.

Zipf, G.K. (1949) *Human Behavior and the Principle of Least Effort*, Addison-Wesley.