DISCUSSION PAPER SERIES

IZA

IZA DP No. 2749

# Two Types of Inequality: Inequality Between Persons and Inequality Between Subgroups

Guillermina Jasso
Samuel Kotz

April 2007

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

# Two Types of Inequality: Inequality Between Persons and Inequality Between Subgroups

**Guillermina Jasso**
*New York University and IZA*

**Samuel Kotz**
*George Washington University*

Discussion Paper No. 2749
April 2007

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 2749
April 2007

# ABSTRACT

## Two Types of Inequality: Inequality Between Persons and Inequality Between Subgroups*

Social scientists study two kinds of inequality: inequality between persons (as in income inequality) and inequality between subgroups (as in racial inequality). This paper analyzes the mathematical connections between the two kinds of inequality. The paper proceeds by exploring a set of two-parameter continuous probability distributions widely used in economic and sociological applications. We define a general inequality parameter, which governs all measures of personal inequality (such as the Gini coefficient), and we link this parameter to the gap (difference or ratio) between the means of subdistributions. In this way we establish that, at least in the two-parameter distributions analyzed here, and for the case of two nonoverlapping subgroups, as personal inequality increases, so does inequality between subgroups. This general inequality parameter also governs Lorenz dominance. Further, we explore the connection between subgroup inequality (in particular, the ratio of the bottom subgroup mean to the top subgroup mean) and decomposition of personal inequality into between-subgroup and within-subgroup components, focusing on an important decomposable measure, Theil's MLD, and its operation in the Pareto case. This allows us to establish that all the quantities in the decomposition are monotonic functions of the general inequality parameter. Thus, the general inequality parameter captures the "deep structure" of inequality. We also introduce a whole-distribution graphical tool for assessing personal and subgroup inequality. Substantively, this work suggests that in at least some societies, characterized by special income distributions, whenever inequality disrupts social harmony and social cohesion, it attacks on two fronts, via subgroup inequality as well as personal inequality.

JEL Classification: C02, C16, D31, D6, I3

Keywords: continuous univariate distributions, two-parameter distributions, lognormal distribution, Pareto distribution, power-function distribution, Gini coefficient, Atkinson measure, Theil's MLD, coefficient of variation, Lorenz curve, decomposition of inequality measures, between component, within component

Corresponding author:

Guillermina Jasso
Department of Sociology
New York University
295 Lafayette Street, 4th Floor
New York, NY 10012-9605
USA
E-mail: gj1@nyu.edu

* We thank Yaw Nyarko and participants at the biennial meeting of the International Society for Justice Research, Berlin, Germany, August 2006, and at the annual Winter Conference of the Methodology Section of the American Sociological Association, Yale University, March 2007, especially Scott Boorman, Thomas DiPrete, and Tim F. Liao, for their valuable comments; and we thank the editor and three anonymous referees for their close reading and valuable comments. We also gratefully acknowledge New York University and George Washington University for their intellectual and financial support. This paper is forthcoming in *Sociological Methods and Research*.

## 1. INTRODUCTION

Discussions of social inequality address two distinct types of inequality, inequality between persons and inequality between subgroups (see, inter alia, the landmark work by Jencks et al. 1972). The quintessential example of the first type is income inequality, and its study has produced an extensive literature focused on both substantive and methodological aspects (e.g., Champernowne and Cowell 1998; Karoly and Burtless 1995; Kleiber and Kotz 2003). The quintessential examples of the second type are the race earnings gap and the gender earnings gap, and these are examined in a growing substantive literature, predominantly in the United States and the United Kingdom (e.g., Blau and Kahn 2000; Darity and Mason 1998; Goldin 1990, 2006; Harkness 1996; O'Neill 2003; Reskin and Bielby 2005). In the United States, the burgeoning interest in race and gender gaps has been stimulated in part by concern that -- notwithstanding the almost half century since President John F. Kennedy issued the groundbreaking Executive Order 10925 prohibiting discrimination on the basis of "race, creed, color, or national origin" (6 March 1961) and signed the Equal Pay Act (10 June 1963) extending to gender the protection against discrimination -- race and gender may still affect economic chances (U.S. Council of Economic Advisers 1998ab). These governmental actions, although limited to the United States, have had a wider international significance and impact.[1]

In both types of inequality, interest centers on inequality with respect to a positive quantitative variable $X$, such as wages, income, or wealth. The two types differ, however, in the entities under assessment. While persons are the focal units in the first type, the focal units in the second type are subgroups based on a particular qualitative variable $S$, such as gender, race, or nativity. Thus, inequality between persons refers to inequality in the distribution of $X$ in a set of persons – called personal inequality. Inequality between subgroups, on the other hand, refers to the discrepancy between the mean (or other measure of central tendency) of $X$ between two

[1] Jencks et al. (1972) discuss both types of inequality in the introductory chapter and while emphasizing "inequality between individuals, not inequality between groups" (Jencks et al. 1972:14), nonetheless introduce pertinent material on race gaps (e.g, in pp. 216-219).

subgroups of a group (or population) – called subgroup inequality. Tailored to a specific set of quantitative and qualitative characteristics, the two types of inequality are often called "*X* inequality" – e.g., earnings inequality or wealth inequality – and "*S X* gap" – e.g., gender earnings gap or nativity wealth gap or race wage gap, respectively.[2]

Personal inequality can be defined for a group or population as well as for the subgroups within the group. We refer to personal inequality in a group as plain unmodified "personal inequality" or, synonymously, as overall inequality. We refer to personal inequality in a subgroup as within-subgroup personal inequality, with more particular terms as appropriate, for example, bottom-subgroup personal inequality and top-subgroup personal inequality, or male personal inequality and female personal inequality, or terms based on particular measures of inequality, such as bottom-subgroup Gini coefficient or top-subgroup Theil's MLD. The term "subgroup inequality" is always reserved for measures of inequality between subgroups, such as the ratio of the bottom-subgroup mean to the top-subgroup mean.

A basic natural question is: What is the connection between overall inequality and subgroup inequality? Is one a function (hopefully monotone) of the other? Or are both generated by some other deeper feature of the distribution or population? There are several approaches one can take, including investigation of the mechanisms producing the quantitative variable *X* and its interplay with the *S* characteristic. Such an approach might explore whether persons who differ in the *S* characteristic (say, men and women, or immigrants and natives) have differential access to *X* itself (as when laws prescribe wages by sex or constrain asset accumulation by religion – as in the Middle Ages in Europe, say, and certain Muslim countries even today) or to the sources of *X* (as when education is differentially available on the basis of gender, nativity, race, or religion) or differ in some *X*-relevant characteristic (such as language or attachment to the labor force).

---

[2] The distinction between qualitative and quantitative characteristics, long appreciated in mathematics (e.g., Allen 1938:10-11) and in statistics and econometrics, has since the pioneering work of Blau (1974) come to be seen as structuring behavioral and social phenomena in a fundamental way. The combination of a qualitative and a quantitative characteristic in subgroup inequality provides a further instance of the scope and usefulness of the distinction.

Our approach is more modest, yet possibly of larger significance, involving greater generality. Because the distribution of any $X$ is subject to the same basic mathematical underpinnings of probability distributions, it is plausible that overall inequality and subgroup inequality are connected to each other in specified ways. Accordingly, we investigate the mathematical relations between overall inequality and subgroup inequality.

A further natural question involves the connection between subgroup inequality and within-subgroup features, such as within-subgroup personal inequality and within-subgroup means. These within-subgroup features play vital parts in decomposition of overall inequality, where they appear as the basic ingredients of what are called the between-subgroup component and the within-subgroup component (Bourguignon 1979; Shorrocks 1980; Das and Parikh 1982; Champernowne and Cowell 1998). As scholarly attention expands from a single population with subgroups to comparison of many populations, each with its own set of subgroups – an enterprise "in its infancy" (Darity and Deshpande 2000:75) – as well as to comparison over time of a population with subgroups (Tomaskovic-Devey, Thomas, and Johnson 2005), clear understanding of the exact connections among all these terms and dimensions becomes all the more important. Analogously to our starting focus on the connection between overall inequality and subgroup inequality, our strategy is to investigate the mathematical relations among the terms in play.[3]

We begin with specification of a <u>general inequality parameter</u>, a property of mathematically specified distributions defined on the positive support. Intuitively, we first show

---

[3] The questions concerning the exact connections between personal inequality, subgroup inequality, subgroup-specific means and personal inequality, and between-subgroup and within-subgroup components in decomposition are indeed general questions, and these concepts and relations can be applied at any level of generality – in various situations. For example, the group and its subgroups may as easily be the world and its countries (e.g., Theil 1979; Berry, Bourguignon, and Morrisson 1983; Schultz 1998; and Firebaugh 2003) as a country and its native-born and foreign-born citizens or a social club and its male and female members. And in all these cases one can imagine the expanding question of comparisons across groups and over time – albeit only in philosophy or science fiction for comparison across worlds.

that the general inequality parameter governs four measures of overall inequality, and next we show that it also governs measures of subgroup inequality. If both overall inequality and subgroup inequality are functions of the general inequality parameter, then we may conclude that these two distinct types of inequality are presumably manifestations of the same underlying inequality in a distribution.

Next we investigate the operation of the general inequality parameter in the subgroup-specific terms and their combination into decomposition terms. Again, but in a more restricted analysis, we show that all these terms are monotonic functions of the general inequality parameter, strengthening the conclusion that all are manifestations of what we may call, using Darity and Deshpande's (2000:77) words, the "deep structure" of inequality.

In this first foray, we impose several restrictions. First, we work with specific distributions rather than following a distribution-independent approach. Second, we focus on continuous two-parameter distributions, investigating three widely used but mathematically quite simple distributions (lognormal, Pareto, and power-function). Third, the subgroups are defined as two censored subdistributions. Future work should systematically relax these restrictions, assessing, for example, distributions with more than two parameters or with more than two subgroups or with overlapping subgroups. To the extent that the nonoverlapping-subgroup case provides a model for many real-world situations – from slave and caste societies to societies in which gender, nativity, immigration status, or rank in an organizational structure predetermine minimum and maximum earnings or wealth to binational or international work settings in which different pay scales are used for locals and members of foreign organizations (such as international joint ventures as well as U.S. news media, filmmakers, universities, think tanks, and military installations abroad) to groups with differentially compensated subgroups (such as universities with differentially compensated disciplines, professional associations with members from around the world, international airlines with pilots under different contracts) -- and that, as well, it represents the extreme bound, the results reported in this paper may serve as at least a benchmark for future studies.[4]

In the analysis of the within-subgroup and decomposition terms, we focus on one inequality measure, the "second measure" proposed by Theil (1967:125-127; 1979) -- also known as the mean logarithmic deviation (MLD) – and investigate its operation in one of the three variates mentioned above, the Pareto distribution. Future work should systematically analyze the within-subgroup and decomposition terms for the MLD in other distributions and for other inequality measures.

Finally, we briefly introduce a whole-distribution graphical tool for assessing inequality across groups that have subgroups, a technique based on the quantile function and its application to inequality analysis (Jasso 1983b). This technique can be used with both overlapping and nonoverlapping subgroups; there can be any number of subgroups; and the subgroup distributions can have any mean, inequality, or variate form. This technique can be used both with mathematically specified distributions and with empirical distributions.

Section 2 introduces the general inequality parameter and reports its operation in the three variates – lognormal, Pareto, and power-function – showing that measures of overall inequality are monotonic functions of the general inequality parameter. In Section 3, we define two measures of subgroup inequality, one difference-based and the other ratio-based – yielding the absolute gap and the relative gap, respectively -- and we show that both measures are also monotonic functions of the general inequality parameter. Section 4 links the Lorenz curve to the general inequality parameter (showing that the Lorenz curve is a monotonic function of the general inequality parameter) and to the connection between overall inequality and subgroup inequality. Section 5 investigates the within-subgroup and decomposition terms in the MLD and analyzes their operation for the Pareto variate; the analysis indicates that all the within-subgroup and decomposition terms are monotonic functions of the general inequality parameter but they

---

[4] For example, on international joint ventures, see Shenkar and Zeira (1987), Leung, Smith, Wang, and Sun (1996), and Leung, Wang, and Smith (2001).

differ in their responsiveness to the proportions of the population in the two subgroups. Section 6 describes the whole-distribution graphical tool. A short note concludes the paper.[5]

## 2. THE GENERAL INEQUALITY PARAMETER

### 2.1. Two-Parameter Continuous Univariate Distributions

Consider a two-parameter continuous univariate distribution defined on the positive support. The two parameters are usually specified as a location parameter (say, the mean, median, or mode) and either a scale or a shape parameter. Table 1 presents the formulas for the three basic associated functions – the probability density function (PDF), cumulative distribution function (CDF), and quantile function (QF) – for the lognormal, Pareto, and power-function distributions; Figure 1 provides visual illustration. The location parameter is specified as the mean $\mu$. The second parameter is a shape parameter denoted $c$.[6]

– Table 1 about here –

– Figure 1 about here –

### 2.2. The Second Parameter *c* as General Inequality Parameter

We argue that the second parameter $c$ is in fact a general inequality parameter. We carry out two sets of analyses. First, we examine the formulas for the major measures of overall inequality in order to assess their relation to the general inequality parameter. If the overall inequality formulas are monotonic functions of $c$, then $c$ is a plausible candidate for general inequality parameter. Second, we examine the limit of the quantile function (QF) as $c$

---

[5] Of course, there are other dimensions along which different types of inequality can be discerned. For example, Jasso and Kotz (2007) distinguish between inequality in status and inequality in the characteristics which generate status (such as income). Moreover, Liao (2006:217) uses the phrase "two types of inequality" to refer to the between and within components in a decomposition.

[6] For detailed information on these basic distributions, see, for example, Johnson, Kotz, and Balakrishnan (1994, 1995) and, for applications pertinent to inequality, Kleiber and Kotz (2003).

approaches its low-inequality end. If the limit of the QF is the mean, then the distribution becomes degenerate, and *c* is then a fortiori a plausible candidate for general inequality parameter.[7]

### 2.3. <u>The General Inequality Parameter and Measures of Overall Inequality</u>

We focus on four major measures of overall inequality – the Gini coefficient, one of Atkinson's (1970, 1975) measures (defined as 1 minus the ratio of the geometric mean to the arithmetic mean), Theil's (1967:125-127; 1979) MLD, and Pearson's (1896) coefficient of variation (CV). Formulas for these measures are well-known (see, e.g., Cowell 1977; Kleiber and Kotz 2003). Here we provide in Table 2 some convenient formulas for these measures in both observed and mathematically specified distributions. The formulas for mathematically specified distributions are expressed in terms of the quantile function (QF) and the mean $\mu$.

– Table 2 about here –

The lognormal, Pareto, and power-function variates are quite familiar and the formulas for the four inequality measures in these variates are well-known (see, e.g., Cowell 1977; Kleiber and Kotz 2003). Tables 3.a and 3.b present formulas for Note that in all twelve formulas, there is a single factor, namely, *c*, which is the second of the two basic parameters in a two-parameter distribution and the one we are proposing as a general inequality parameter. Thus, the Gini coefficient, Atkinson's measure, Theil's MLD, and the coefficient of variation are all functions solely of *c*.

– Tables 3.a and 3.b about here –

Tables 3.a and 3.b also provide the first derivatives of each measure of overall inequality with respect to the parameter *c*. As shown, the derivatives of all four measures are in each variate of the same sign, so that the measures of overall inequality are monotonic functions of *c*.

---

[7] The idea of a general inequality parameter appears in Jasso (1987:96) and is rooted both in well-known discussions of the shape factor in the Pareto distribution (known as Pareto's constant) as a measure of inequality (e.g., Cramer 1971:51-58; Cowell 1977:95; Kleiber and Kotz 2003:78) and the visual observation that the formulas for many inequality measures in mathematically specified distributions depend on a single parameter (e.g., Jasso 1982:315-318).

Figures 2.a and 2.b present graphs of the four measures of overall inequality as functions of *c*. As expected from the derivatives, the graphs depict the measures' monotonicity with respect to *c*.

– Figures 2.a and 2.b about here –

Thus, the parameter *c* can be interpreted as a general inequality parameter.

Aside from monotonicity, *c* operates differently across the variates. In the lognormal case, inequality increases as *c* increases, while in the Pareto and power-function inequality decreases as *c* increases.[8]

It is also useful to examine the limits of the overall measures of inequality as *c* approaches its own limits. The Gini coefficient has bounds of zero and one. Accordingly, in each variate the formula for the Gini coefficient should approach its two limits as the parameter c approaches its two limits, respectively. Consider the lognormal distribution. As *c* approaches zero, the limit of the Gini coefficient is zero; and as *c* goes to infinity, the limit of the Gini coefficient is one.[9]

Proceeding in this manner, we find that for each of the three variates, the Gini coefficient approaches its two limits as the parameter *c* approaches its two limits, respectively, and so does the Atkinson measure. Theil's MLD and the CV do not have an upper bound, and for these we find that in each of the three variates, the two measures approach zero and infinity (their two limits) as the parameter *c* approaches its two limits, respectively. These results provide further indication that *c* is a general inequality parameter.

---

[8] This parameter which operates as general inequality parameter is called by different names or appears in alternate forms. For example, as mentioned above, in the Pareto case, the parameter *c* is called Pareto's constant; and the lognormal's *c* is often referred to as the standard deviation of the normal distribution obtained when the lognormal is "logged."

[9] Looking at the formula for the Gini in the lognormal case (Table 3.a), observe that as *c* goes to zero, the CDF of the unit normal goes to .5 (and .5 times 2 equals one which, after subtracting one, yields zero). Similarly, as *c* goes to infinity, the CDF of the unit normal goes to one.

### 2.4. The General Inequality Parameter and the March toward Equality[10]

Consider the expressions for the quantile functions of the three variates, shown in Table 1. Because the location parameter was specified as the arithmetic mean, the expressions for the QF contain the mean (as well as the general inequality parameter *c*). And because the arithmetic mean represents equality, the expression for the QF in an idealized Equal distribution (also known as a Dirac distribution or a degenerate distribution) should be the arithmetic mean. That is, in a perfectly equal distribution, all income amounts are equal and thus at every "relative rank" the QF is simply the mean.

It follows that if *c* is indeed a general inequality parameter, then as *c* approaches its low-inequality end or limit, the QF should approach the mean. In other words, denoting the low-inequality limit by *q*, we have:

$$\lim_{c \to q} Q(\alpha; \mu, c) = \mu . \qquad \textbf{(1)}$$

It is straightforward to show that this is indeed the case in the three variates under consideration. To illustrate, we present a formal proposition and proof for the Pareto case.

Proposition 1. Limit of the Pareto QF, as the Parameter *c* Approaches Its Low-Inequality End. In the Pareto distribution, the limit of the quantile function, as *c* goes to infinity (its low-inequality end), is equal to the arithmetic mean:

$$\lim_{c \to \infty} \frac{\mu(c-1)}{c(1-\alpha)^{1/c}} = \mu . \qquad \textbf{(2)}$$

Proof. The proof of proposition 1 is given in two steps.

1. The three constituent factors of the Pareto's QF have well-known limits:

[10] The evocative phrase "march toward equality" is in wide use. See, for example, the chronicle of milestones in the history of U.S. civil rights in the website of the U.S. Department of State (http://usinfo.state.gov/products/pubs/civilrts/march.htm).

$$\lim_{c\to\infty} \mu = \mu$$

$$\lim_{c\to\infty} \frac{c-1}{c} = 1 \qquad \textbf{(3)}$$

$$\lim_{c\to\infty} (1-\alpha)^{1/c} = 1$$

2. Using the limits in (3) and the fact that the limit of a product(quotient) equals the product(quotient) of the limits (provided, in the case of a quotient, that the divisor is not zero), we immediately obtain the limit for the Pareto case in (2).

Similarly, the limit of the power-function's QF, as *c* goes to infinity (its low-inequality end), equals the arithmetic mean; and the limit of the lognormal's QF, as *c* approaches zero (its low-inequality end), also equals the arithmetic mean.

Thus, as the inequality parameter approaches its low-inequality end or limit, the distribution collapses onto a single point, the point of equality. The parameter *c* – the second parameter in the two-parameter specifications of the variates – is indeed behaving as a general inequality parameter.

Evidently the robustness of the parameter *c* as general inequality parameter merits further elucidation.

## 3. SUBGROUP INEQUALITY AND THE GENERAL INEQUALITY PARAMETER

### 3.1. Subgroup Inequality in a Population with Two Subgroups

Consider the distribution of *X* – where *X* is a quantitative variable such as wage, earnings, or wealth – in a population. Suppose now that the persons in the population can be classified into two subgroups according to a qualitative characteristic, such as gender, race, ethnicity, nativity, or religion. Each of the two subgroups has an average amount of *X*. Whenever the two averages differ, the two subgroups will be thought to be unequal. This condition of subgroup inequality can be measured in two ways, by the difference between the two means and by the ratio of the smaller mean to the larger. While the relative gap may be the more often discussed, the absolute gap in income provides complementary information, such as a measure of the discrepancy in purchasing power (Jencks et al. 1972:216-219).

The objective is to ascertain whether subgroup inequality, measured by both the ratio and difference procedures, varies with the general inequality parameter *c* discussed in the previous section.

### 3.2. Censored Subdistribution Structure in a Population with Two Subgroups

We shall now suppose that the two subgroups are nonoverlapping in *X*, i.e., that the poorest person in the higher-average subgroup is richer than the richest person in the lower-average subgroup. This is a rather familiar situation across the social sciences, where it appears under a variety of rubrics, such as consolidation in Blau (1974:632), hierarchy and segmentation in Hechter (1978), cleavage in Jasso's (1983a, 1993) conflict model, accentuation in Hogg, Terry, and White (1995:261), and bifurcation in Ridgeway's (1996, 2001) status construction theory.

Nonoverlapping subgroups occur in a variety of contexts, not only classic slave and caste societies but also in modern times. Seven examples are: (1) societies in which the wage for rural labor is fixed at two amounts, one for men and the other for women; (2) families in which children's allowances are higher for boys than for girls, holding age constant, on the rationale that boys may have higher expenses (such as paying for dates, in certain cultural milieus); (3) firms and organizations in which paygrades are structured so that the lowest pay in one rank is higher than the highest pay in the adjacent lower rank; (4) societies in which inheritance, wealth, or access to certain resources such as land ownership or commercial radio operation are structured by gender, nationality, nativity, or religion; (5) construction and farm crews in which the highest-paid illegal-alien worker is paid less than the lowest-paid legal worker; (6) binational or international work settings in which different pay scales are used for locals and members of foreign organizations (such as international joint ventures as well as U.S. news media, filmmakers, universities, think tanks, and military installations abroad); and (7) groups with differentially compensated subgroups (such as universities with differentially compensated disciplines, professional associations with members from around the world, international airlines with pilots under different contracts). Indeed, when one considers the full scope of quantitative variables in which inequality may be of interest – not only wages, earnings, income, and wealth but also schooling, mentoring time received, length of prison sentence, grades received in school – the examples of nonoverlapping subgroups multiply. Moreover, nonoverlapping subgroups arise quickly in immigration contexts; to illustrate, except in the case where both origin and destination country speak the same language, any function of fluency in the destination-country language (wages, tutoring time, time spent on homework, experience interpreting) is likely to generate nonoverlapping subgroups of immigrants and natives.[11, 12]

In this case of nonoverlapping subgroups, the distribution generates a censored subdistribution structure, in which the censoring point *p* corresponds to the boundary between the two subgroups. The proportions in the censored subdistributions are called the subgroup split. The bottom subgroup contains *p* proportion of the population, and the top subgroup has (1 - *p*) proportion.[13] The phrase "subgroup split *p*" is used as shorthand for "subgroup split's censoring point *p*".

To construct the measures of subgroup inequality, it is necessary to derive expressions for

---

[11] An early example of disjoint monetary value by gender appears in Leviticus (27:3-4), which prescribes the value of a male slave (age 20-60) at fifty shekels and of a female slave (also 20-60) at thirty shekels.

[12] Note the intriguing contrast between immigration contexts and international joint ventures; in the first setting, the natives earn more, while in the second, the natives earn less. For further discussion of international joint ventures, see Shenkar and Zeira (1987), Leung et al. (1996), and Leung et al. (2001).

[13] Using by now standard terminology (see, e.g., Gibbons 1988:355), let censoring refer to selection of units by their ranks or percentage (probability) points; and let truncation refer to selection of units by values of the variate. Thus, the truncation point is the value *x* separating the subdistributions; the censoring point is the percentage point *p* separating the subdistributions. For example, the subgroups with incomes less than $20,000 or greater than $80,000 each form a truncated subdistribution; the top 25 percent and the bottom 75 percent of the income distribution each form a censored subdistribution.

the arithmetic means of the censored subdistributions and then embed these into expressions for the ratio and the difference.

### 3.3. Measures of Subgroup Inequality

Table 4 presents the ratio and difference measures of subgroup inequality for the three variates under consideration in this paper. As shown, all the measures are functions of two quantities, the general inequality parameter $c$ and the censoring point $p$; the difference-based measures are also functions of the arithmetic mean $\mu$. The ratio-based measure has bounds of 0 and 1 (open at 0, closed at 1), attaining the value of 1 when the two subgroups have identical means.

– Table 4 about here –

It is easy to see that the difference-based measures of subgroup inequality are ceteris paribus increasing in the arithmetic mean $\mu$. In the social sciences it is also important to investigate the direction of the effects of the proportions in the two subgroups. We shall turn to the subgroup split in section 3.5 below (after establishing the effect of the general inequality parameter on subgroup inequality).

### 3.4. Effect of the General Inequality Parameter on Subgroup Inequality

Table 4 also provides the first partial derivatives of the two measures of subgroup inequality with respect to the general inequality parameter $c$ for the three basic variates we examine; Figure 3 presents graphs of the ratio-based measures of subgroup inequality on the general inequality parameter $c$ for three values of the subgroup split $p$ in each of the variates.

– Figure 3 about here –

Recall now that $c$ operates differently across the three variates – higher $c$ associated with higher inequality in the lognormal and with lower inequality in the Pareto and power-function. Moreover, in the difference-based measure of subgroup inequality, the larger the measure, the greater the inequality, while the behavior is opposite for the ratio-based measure of subgroup inequality. The pattern of signs – (1) opposite for the ratio and difference measures, and (2) opposite for the lognormal, on the one hand, and the Pareto and power-function on the other –

indicates a consistent effect of *c*, as is evident from the graphs in Figure 3. That is, subgroup inequality seems to be governed by the general inequality parameter. Moreover, the effects are exactly in the same direction as in the case of overall inequality, namely: as the general inequality parameter moves in the direction of greater inequality, subgroup inequality increases.

As above, the robustness of the parameter *c* as general inequality parameter merits further investigation.

### 3.5. Effect of the Subgroup Split on Subgroup Inequality

As shown above, measures of overall inequality are functions of the general inequality parameter *c* alone. In contrast, measures of subgroup inequality are functions not only of *c* but also of the subgroup split – the proportions in the two subgroups, specified by the censoring point *p* (as shown in Tables 3.a, 3.b, and 4). A natural question arises: For given magnitude of *c*, how does the subgroup split affect subgroup inequality?

Figure 4 presents graphs of the ratio-based measures of subgroup inequality on the subgroup split *p* for three values of *c* in each of the three variates. We observe that the effect of the subgroup split on subgroup inequality differs across the three variates in two easily noticeable ways. First, the effect of the subgroup split is monotonic in the Pareto and power-function but not in the lognormal. Second, the monotonic effect of the subgroup split in the Pareto and power-function operates in opposite directions – in the Pareto, the smaller the proportion in the bottom subgroup, the lower the subgroup inequality (higher ratio of bottom-subgroup mean to top-subgroup mean), while in the power-function, the effect is reversed. In the lognormal, the effect of the subgroup split traces an inverted-U-shaped curve, increasing to a peak in the upper part of the range, and subsequently decreasing.

– Figure 4 about here –

These results make sense if one calls to mind the densities of the three variates (Figure 1). The Pareto and power-function distributions (of *c* greater than 1) are almost mirror images of each other, the Pareto with its single mode at the bottom of the range and a long right tail and the power-function with its single mode at the top of the range and a left tail extending to zero.

Thus, while in the Pareto case the presence of large chunks of the population in the top subgroup tempers the effect of the long right tail, making the two subgroup means more similar, in the power-function case, it is the opposite, namely, the presence of large chunks of the population in the bottom subgroup tempers the effect of the left tail and results in the two subgroup means becoming more similar. The lognormal has a mode toward the middle of the range and both a left tail going to zero and a very long right tail, resulting in the nonmonotonicity of the ratio of the bottom-subgroup mean to the top-subgroup mean relative to the subgroup split *p*. (We encourage our readers to provide graphical representation of the behavior just described.)[14]

### 3.6. Remark on Subgroup Inequality

It follows from the discussion above that subgroup inequality is shaped by three things: the form of the distribution, the distribution's general inequality parameter *c*, and the subgroup split. The operation of these three factors has been long known in the study of overall inequality. Now this paper extends the classical work to subgroup inequality, in particular, to the absolute gap and relative gap measures. Explicit attention to the three factors will assist in interpreting differences across countries and changes over time in subgroup inequality (as in the questions addressed by Tomaskovic-Devey, Thomas and Johnson 2005, and Darity and Deshpande 2000).

## 4. THE GENERAL INEQUALITY PARAMETER, THE LORENZ CURVE, AND THE LINK BETWEEN OVERALL INEQUALITY AND SUBGROUP INEQUALITY IN THE LOGNORMAL, PARETO, AND POWER-FUNCTION VARIATES

Lorenz curves introduced some 100 years ago (Lorenz 1905) provide a convenient graphical tool for inspecting the amount of inequality in a distribution and comparing it across two or more distributions. The Lorenz curve expresses the proportion of the total amount of *X* held by the bottom α proportion of the population as a function of α (α is the same α encountered

---

[14] Analysis of this variate behavior may be useful in studies of the sensitivity of different inequality measures to different regions of the distribution as well as in analyses of "latent classes" based on values of *X*, as in Liao (2007).

above as the argument of the quantile function):[15]

$$\frac{1}{\mu}\int_0^\alpha Q(\alpha)\, d\alpha\,. \qquad \textbf{(4)}$$

The three variate families described in this paper have the property that their members possess nonintersecting Lorenz curves. Indeed, the formulas for the Lorenz curve (reported in Table 5) show that, analogously to the overall inequality and subgroup inequality, the Lorenz curve is also governed by the general inequality parameter. As the Lorenz curve is a monotonic function of *c*, within each variate family different members – which differ in *c* -- possess nonintersecting Lorenz curves.[16]

– Table 5 about here –

To visually represent the operation of the general inequality parameter on Lorenz curves, two sets of graphs are provided. Figure 5 presents the usual Lorenz curve representations, for the three variates under study, and, as before, with three cases for each variate. The members' graphs are nonintersecting, and the higher the vertical placement, the lower the inequality – a higher curve Lorenz-dominates a lower curve. For further illumination, we show in Figure 6 graphs of the Lorenz curve formulas plotted on the general inequality parameter *c*, at three symmetric points in the domain of the Lorenz curve: $\alpha = .25$, .5, and .75. These graphs show how the Lorenz curve, analogously to the overall inequality and subgroup inequality, is governed by the general inequality parameter.

– Figures 5 and 6 about here –

Equivalently, holding the arithmetic mean constant, the quantile functions for any pair of distributions from each family will intersect only once, revealing who gains and who loses in a

---

[15] The quantity $\alpha$ can be interpreted in various equivalent ways, depending on the context. For example, as the outcome of the CDF, it is a probability level; as the argument of the QF, it is a relative rank; and as the argument of the Lorenz curve, it is the bottom proportion of the population.

[16] Table 5 builds on Gastwirth (1972:307) who provides a table with the formulas for the Lorenz curve in five distributions, including the Pareto and the benchmark Equal (Dirac).

shift from one member to the other. Specifically, everyone to the left of the intersection is better-off in the lower-inequality distribution, and everyone to the right of the intersection is better-off in the higher-inequality distribution.[17]

For example, for the Pareto case, the intersection of the two QFs is obtained by solving for the difference between two quantile functions, yielding, as shown in Jasso (1983b:291-293):

$$\alpha_0 = 1 - \left[ \frac{c_A(c_B - 1)}{c_B(c_A - 1)} \right]^{\frac{c_A c_B}{c_A - c_B}}, \qquad \textbf{(5)}$$

where A and B denote the two Pareto distributions and $\alpha_0$ denotes the relative rank (or the probability level) at which the intersection occurs. It can be shown that $\alpha_0$ can occur anywhere in the interval between $(1 - e^{-1})$ and 1, or approximately between .632 and 1. In a shift from A to B, where B is the lower-inequality distribution, everyone to the left of $\alpha_0$ becomes better-off and everyone to the right becomes worse-off. And vice-versa in a shift from B to A.

That the Gini and the Lorenz curve are both governed by *c* and that the Lorenz curve and the quantile function are tightly linked is not surprising, for they are clearly mathematically connected: Indeed, the Lorenz curve is the integral of the quantile function (divided by the mean), and the Gini coefficient is (1 minus 2 × [the integral of the Lorenz curve]), as shown in expression (4) and Table 2, respectively.[18]

Now suppose that, as in Section 3 above, there are two nonoverlapping subgroups. It

---

[17] Thus, in comparisons of distributions of relative income, the condition of nonintersecting Lorenz curves is also equivalent to the so-called <u>second-order stochastic dominance</u>. We express second-order stochastic dominance in terms of the quantile function (QF) rather than the usual cumulative distribution function, because the QF has a direct interpretation as the income corresponding to a person of given relative rank, leading to further results concerning who wins (or loses) in a shift from one distribution to another.

[18] The connection between the quantile function and the Lorenz curve was already known to Pietra (1915), and re-appears in Schutz (1951) and Gastwirth (1971). The connection between the Lorenz curve and the Gini coefficient was established by Gini (1914). For further details, see Kleiber and Kotz (2003).

follows that in the within-variate shift from a higher-inequality distribution A to a lower-inequality distribution B, the mean of the upper subgroup (the upper censored subdistribution) decreases and the mean of the bottom subgroup (the bottom censored subdistribution) is increased, reducing the subgroup inequality. Hence, the conditions of nonoverlapping subgroups and nonintersecting Lorenz curves are jointly sufficient for the link between overall inequality and subgroup inequality.

Accordingly, in certain mathematically specified distributions, there exists a general inequality parameter which determines all inequality-related aspects, including overall inequality and subgroup inequality as well as Lorenz dominance and stochastic dominance.

## 5. THE CONNECTION BETWEEN THE GENERAL INEQUALITY PARAMETER, SUBGROUP INEQUALITY, AND DECOMPOSITION OF OVERALL INEQUALITY: AN ANALYSIS BASED ON THEIL'S MLD FOR THE PARETO VARIATE

### 5.1. Introducing Decomposition into the Analysis

Our discussion to this point has focused on two types of inequality, inequality between persons – personal inequality (or, equivalently, when defined on all persons in a group or population, overall inequality) – and inequality between subgroups – subgroup inequality. We have shown that in certain basic mathematically specified distributions both personal inequality and subgroup inequality are monotonic functions of the general inequality parameter and that as general inequality increases, both personal inequality and subgroup inequality increase. We have also confirmed that Lorenz curve dominance is a monotonic function of the general inequality parameter.

We now introduce into the discussion decomposition of overall inequality into a between-subgroup component and a within-subgroup component. Such decomposition holds the promise of gauging how much of overall inequality is due to inequality within subgroups and how much to inequality between subgroups (Theil 1967, 1979; Bourguignon 1979; Shorrocks 1980; Das and Parikh 1982; Jasso 1982:321-323; Berry, Bourguignon, and Morrisson 1983; Champernowne and

Cowell 1998; Schultz 1998; Firebaugh 1999, 2003; Liao 2006). More generally, one would expect to find a link between one or more elements of the decomposition and the difference-based or ratio-based measures of subgroup inequality analyzed in this paper. Accordingly, we now investigate the mathematical relations among the expanded set of terms encompassing not only the overall inequality and subgroup inequality but also new decomposition-specific terms.

We shall focus on one well-known and appealing inequality measure, Theil's MLD (which combines several useful properties – additive decomposability, scale invariance, and sensitivity to population shares (Theil 1979; Bourguignon 1979; Shorrocks 1980)) – and on one widely used distributional family, the Pareto.[19]

Note that assessing the mathematical relations among the newly expanded set of terms contributes to addressing what Darity and Deshpande (2000:77) call the "grand questions," such as the relation between overall inequality and subgroup inequality and the relation between subgroup inequality and within-subgroup personal inequality.

### 5.2. Theil's MLD, Elements of the Decomposition, and the Link to Subgroup Inequality

We begin by collecting in Table 6 the formulas for the MLD and for all the constituent elements of its decomposition. The MLD for a given quantitative variable in a group (or population) is defined as the average of the log of the ratio of the group mean μ to each unit's amount *x*, which for simplicity can be expressed as an expectation and which yields the log of the ratio of the group mean μ to the group geometric mean, denoted *G*:

$$E\left[\ln\left(\frac{\mu}{X}\right)\right] = \ln\left(\frac{\mu}{G}\right) . \qquad \textbf{(6)}$$

In the case investigated in this paper of two nonoverlapping subgroups, each subgroup also has

---

[19] Theil's MLD has an additional interesting property. It is the negative of the special justice index JI1* which summarizes the experience of justice and injustice in a group or population in the special case where each individual's idea of the just reward is perfect equality. This special case, known as the "justice is equality" case, can be traced to Plato's Gorgias (Jasso 1980, 1999).

its own $\mu$, $G$, and MLD. These are indicated by the two subscripts, $B$ for the bottom subgroup and $T$ for the top subgroup. As before, $p$ denotes the proportion in the bottom subgroup.[20]

– Table 6 about here –

Because $X$ is defined on the positive support, its mean and the two subgroup means are positive. The overall MLD and the two subgroup MLDs are also always positive, a result established by classical theorems relating the arithmetic mean to the geometric mean (Hardy, Littlewood, and Pólya 1952:16-18). For a positive variable the geometric mean is always less than or equal to the arithmetic mean, with equality obtained when the distribution is Equal (equivalently, Dirac or degenerate).

Table 6 reports the general formula for each subgroup's mean and MLD and the general formulas for the between component and the within component in the MLD decomposition. It is straightforward to show, using properties of logarithms and geometric means, that the between component and the within component sum up to the overall MLD.

Two formulas are given for the between and within components, the definitional formulas to the left of the equals sign and to the right of the equals sign new formulas obtained algebraically which highlight the ratio-based measure of subgroup inequality analyzed in section 3. These formulas show that the subgroup inequality is embedded in both the between component and the within component of the MLD.

The between component has occasionally been used as a proxy for subgroup inequality. Indeed, it is designed to capture the portion of the MLD attributable to the inequality between subgroups. However, as Darity and Deshpande (2000:76-77) point out, the between component

[20] The argument of the logarithmic function on the lefthand side of the equals sign is the reciprocal of the pivotal quantity $x/\mu$, the quantity known as the relative amount which arises in statistics and which plays many parts in all the sciences. In applications to income, it is called the relative income (Jasso 1983b) and the income ratio (Firebaugh 1999, 2003); in justice theory it is the comparison ratio under the primitive restrictions (Jasso 1980); and in inequality analysis it appears explicitly in the definitional formulas of many measures of inequality (such as the Atkinson family and the Theil family – see, for example, Cowell 1977:155 and Jasso 1982:306, 311) and is embedded in still other inequality measures (Firebaugh 1999, 2003).

in decompositions usually uses the ratio of the subgroup mean to the overall mean when in fact the "socially relevant anchor" for the bottom subgroup is not the overall mean but rather the top-subgroup mean. These authors are thus led to the ratio-based measure of subgroup inequality as a more exact measure of disparity between subgroups than the between component of a decomposition.

### 5.3. <u>Theil's MLD in the Pareto Variate</u>

Our first task is to derive for the Pareto variate all the quantities associated with the MLD decomposition (shown in Table 6). We already have in hand the overall MLD (Table 3.b), and now obtain mathematical expressions for the other terms, collecting them in Table 7. There is one fundamental result in Table 7, along with several surprises.

– Table 7 about here –

The fundamental result is that all the quantities are functions of the general inequality parameter *c* and all but one of the subgroup and decomposition quantities are also functions of the subgroup split *p*. The main surprises are two: First, the top-subgroup MLD is identical to the overall MLD (and independent of the subgroup split *p*). Second, the expression for the ratio-based subgroup inequality in the Pareto case (Table 4),

$$\frac{(1-p)^{1/c} - (1-p)}{p}, \qquad \textbf{(7)}$$

is embedded not only in the between and within components (as expected from Table 6) but also in the bottom-subgroup MLD.

Actually, the fact that the top-subgroup MLD is the same as the overall MLD should not be too surprising, given that the Pareto curve is known to possess the property that the top-subgroup mean is a constant multiple of the *x* value at the censoring point *p* (Allen 1938:407-408).

Finally, we observe that the formulas for the between and within components sum to the overall MLD.

## 5.4. The General Inequality Parameter and the MLD Decomposition in the Pareto Case[21]

We turn now to examine each of the quantities associated with decomposition of the MLD and their connection to the general inequality parameter *c*.

Subgroup Means and *c*. Partial differentiation of the two subgroup means with respect to *c* indicates that, holding constant the subgroup split, the bottom(top)-subgroup mean is an increasing(decreasing) function of *c*. Thus, as general inequality increases, the bottom(top)-subgroup mean decreases(increases), as shown in panels A and B of Figure 7.

– Figure 7 about here --.

Subgroup MLDs and *c*. Given that the top-subgroup MLD is in this case the same as the overall MLD, it follows that it is a decreasing function of *c*. Partial differentiation of the bottom-subgroup MLD with respect to *c* indicates that it is also a decreasing function of *c* (panels C and D of Figure 7). Thus, as general inequality increases, both subgroup MLDs increase.

Between Component and *c*. Partial differentiation of the between component shows that it is a decreasing function of *c*. Consequently, as general inequality increases, so does the between component. This is indicated in Figure 8, which presents graphs of the overall MLD and the two components for three subgroup splits.

– Figure 8 about here –

Within Component and *c*. Partial differentiation of the within component also indicates that it is a decreasing function of *c*. Thus, as general inequality increases, both the within and between components increase (Figure 8).

Between Component as a Percentage of Overall MLD. It is of interest to consider the relative sizes of the between and within components. Given that they sum up to the overall MLD, either one of them can be expressed as a fraction of the overall MLD. We examine the

---

[21] Sections 5.4 and 5.5 ought to be read, and analyzed, most carefully to arrive at a clear picture of the subgroup split and MLD decomposition in the important (pivotal) Pareto family.

between component as a percentage of overall MLD.[22] Partial differentiation of this new measure implies that it is a decreasing function of *c*; moreover, inspection of values and graphs, provided in Figure 9, shows that this measure is almost constant. Thus, as general inequality decreases (as *c* increases), the relative sizes of the between and within components change trivially, with the relative size of the between component decreasing steadily (if mildly).

– Figure 9 about here –

These results provide further evidence of the operation of the general inequality parameter. In a nutshell, the general inequality parameter governs overall inequality and subgroup inequality as well as the two subgroup means and (at least for the case of the MLD for the Pareto variate) governs also the two subgroups' MLDs, the between and within components, and the relative sizes of the two components.

Recall that the general inequality parameter may encapsulate the "deep structure" of inequality envisaged by Darity and Deshpande (2000:77).

### 5.5. The Subgroup Split and the MLD Decomposition in the Pareto Case

As we have seen, all the subgroup-specific quantities (except the top-subgroup MLD) depend also on the subgroup split, as do the between and within components. We shall therefore examine here the behavior of the relative sizes of the two subgroups.

Subgroup Means and *p*. It would seem that both subgroup means should be increasing functions of the proportion in the bottom subgroup, given that as the proportion in the bottom subgroup increases, the bottom(top) subgroup acquires(loses) units with higher(lower) values of the quantitative variable. Partial differentiation of the two subgroup means with respect to *p* confirms this conjecture. This operation is depicted in panels A and B of Figure 10, which provides graphs for the subgroup means and MLDs, expressed as functions of the subgroup split (complementing Figure 7 which deals with functions of the general inequality parameter *c*).

[22] In a decomposition of the Gini coefficient, Liao (2006:217-218) uses this measure – the ratio of the between component to overall inequality – as an index of relative stratification.

– Figure 10 about here –

Subgroup MLDs and *p*. The top-subgroup MLD is independent of the subgroup split while the bottom-subgroup MLD is an increasing function of *p*. Thus, as the bottom subgroup gets larger, its inequality increases, as one would expect in the Pareto case. Graphs of the subgroup MLDs, as functions of *p*, appear in panels C and D of Figure 10.

Between Component and *p*. Partial differentiation of the between component indicates that it is a nonmonotonic function of *p*. Figure 11, complementing Figure 8, presents graphs of the overall MLD and the two components, expressed now as functions of *p*. The curves for the between component indicate that as the proportion in the bottom subgroup increases, the between component first increases, reaches a peak, and then declines.

– Figure 11 about here –

Within Component and *p*. The within component is a mirror image of the between component. Consequently, partial differentiation with respect to *p* confirms that it is a nonmonotonic function of *p*, as shown in Figure 11. As the proportion in the bottom subgroup increases, the within component first decreases, reaches its nadir, and then increases.

Between Component as a Percentage of Overall MLD. Analogously to the between component on which it is based, this measure is also a nonmonotonic function of *p*. As the proportion in the bottom subgroup increases, this measure increases, reaching a peak that is a function of the general inequality parameter *c*, and then declines (see Figure 12, complementing Figure 9).

– Figure 12 about here –

### 5.6. Other Features of the MLD Decomposition in the Pareto Case

Relative Sizes of Bottom-Subgroup and Top Subgroup MLDs. The top-subgroup MLD is always larger than the bottom-subgroup MLD, but the difference decreases as the subgroup split *p* increases. The limits of this difference are the overall MLD (as *p* approaches zero from the right) and zero (as *p* approaches 1 from the left). Figure 13 presents graphs of the top-subgroup and bottom-subgroup MLDs for three members of the Pareto family (with $c = 1.5$, 2, and 2.5).

As already noted, in the Pareto case the top-subgroup MLD is the same as the overall MLD, and (as shown in Figure 2.b and discussed earlier) is always a monotonic function of the general inequality parameter *c*.

– Figure 13 about here –

Relative Sizes of the Two Components. It follows from Figures 8, 9, 11, and 12 that the relative size of the components depends on both the general inequality parameter *c* and the subgroup split *p*. For subgroup splits roughly between 51-52% and points at the very upper end in the range of approximately 95-99%, the between component is always larger (for all values of *c*). As shown in Figure 12, the between component expressed as a percentage of overall MLD is of inverted-U shape. For example, the curve corresponding to *c* of 2 crosses .5 at approximately $p$ = .514, reaches a peak of about .706 at approximately $p$ = .851, then crosses .5 again at approximately $p$ = .975.

### 5.7. Head-to-Head Contrast of the Between Component and Subgroup Inequality in the Pareto Case

We mentioned above that the between component is sometimes thought of as a gauge of inequality between subgroups, and observed the differences in the formulas of the between component and the ratio-based subgroup inequality. In particular, as noted by Darity and Deshpande (2000:76), the between component lacks the specificity of the ratio of the bottom-subgroup mean to the top-subgroup mean as a measure of disparity between subgroups. We now present a "head-to-head" contrast.

To begin, we obtain a new measure of subgroup inequality defined as one minus the ratio of the bottom-subgroup mean to the top-subgroup mean:

$$1 - \left( \frac{\mu_B}{\mu_T} \right) . \qquad \textbf{(8)}$$

Thus subgroup inequality and the between component are now defined in the same direction – namely, the high-inequality end at the left (as *c* approaches 1 from the right) and the low-inequality end at the right (as *c* grows larger, tending to infinity). Of course, the new measure of subgroup inequality retains its bounds of 0 and 1.

Figure 14 depicts graphs of the between component and the new measure of subgroup inequality – labeled "BetComp" and "SubIneq" – for three subgroup splits, with $p$ = .25, .5, and .75. A careful visual analysis shows that, while the between component can be much larger than subgroup inequality in a small region of very high inequality (as $c$ approaches 1 from the right), for the bulk of the domain of $c$, subgroup inequality is larger than the between component. The discrepancy between them is a function of the subgroup split. Both the between component and the new measure of subgroup inequality have similar convexity.

– Figure 14 about here –

On the basis of Figure 14, neglecting the difference in the two measures' numerical values, one would have to conclude that both measures capture the essential nature of inequality between subgroups and, as pointed out above, both move in the same direction with the general inequality parameter $c$.

Evidently, both measures also vary with the subgroup split, and thus, before any conclusion is reached about their relative performance, they should be examined as functions of $p$. Figure 15 indicates that the two measures now differ substantially. While subgroup inequality is monotonic, the between component is nonmonotonic (as follows from our earlier analysis). The two measures are therefore <u>not</u> interchangeable. It would appear that, as Darity and Deshpande (2000:76) observe, the relative gap is a sharper measure of subgroup inequality than the between component and thus is a serious contender for the measure of choice.

– Figure 15 about here –

Further research on other variates (lognormal, power-function, etc.) would hopefully elucidate the relation between the between component and the ratio measure of subgroup inequality.

## 6. WHOLE-DISTRIBUTION GRAPHICAL TOOL FOR ASSESSING INEQUALITY

So far we have focused on a variety of scalar measures for assessing inequality. It is of interest and importance to also visualize the distribution as a whole, and for this purpose the well-known quantile function (QF) becomes handy. The quantile function represents the amount of a quantitative variable as a function of the relative rank. Figure 16 illustrates use of the QF for the simple case of the Pareto distribution with $c = 2$. Here there are two nonoverlapping subgroups each with half of the population. Horizontal lines indicate the overall mean (fixed at 10) and the two subgroup means. Using the formulas in Table 7 we easily arrive at the values of 5.858(14.142) for the bottom(top)-subgroup means, respectively.[23]

– Figure 16 about here –

It is highly instructive to visually assess all the important quantities and their interrelations. For example, subgroup inequality is the ratio of the lower half-horizontal line to the upper half-horizontal line. The flatness of the curve provides an assessment of the degree of inequality (the flatter the curve the less the inequality), and consequently the QF provides a gauge not only of overall inequality but also of inequality within each subgroup.

This graphical device is also evidently useful in the general case, including overlapping subgroups as well as subgroup distributions differing in mean, general inequality parameter, or the underlying variate. Here each subgroup is graphed as though it were the entire distribution and all subgroup distributions are superimposed on each other. Figure 17 illustrates such a case, where both distributions are Paretos; the bottom(top) subgroup has a mean of 8(12) and a $c$ of 2.5(1.5). Observe that the top subgroup, as measured by the mean, has poorer persons than the bottom subgroup. (The poorest person in the top subgroup possesses an amount of 4, smaller than the 4.8 of the poorest person in the bottom subgroup). This simple example underscores the

[23] Using the quantile function to assess both personal and subgroup inequality builds on work carried out over fifty years ago using the QF to assess personal inequality (see Schutz 1951 and, for overviews and additional references, Jasso 1983b and Chipman 1985).

critical importance of whole-distribution methods.

– Figure 17 about here –

These simple graphs can be refined in several ways. For example, the curves could cover an area proportional to population. Indeed, this graphical tool can be used with any combination of variates and any number of subgroups (overlapping or not). It can as well be used with empirical data.

Note that if different mechanisms shape the *X* distribution within subgroups, the group's distribution will then be a mixture of distributions and its form may obscure the mathematical representation. Fortunately, in recent years substantial advances have been made in the theory and application of mixtures of continuous distributions.

## 7. CONCLUDING NOTE

This paper attempts to clarify the mathematical connection between personal (overall) inequality and subgroup inequality. Using three widely-used two-parameter continuous univariate distributions and four measures of personal inequality, and restricting to the special case of two nonoverlapping subgroups, we find that for the three variates examined, both personal inequality and subgroup inequality are governed by one of the variate's two parameters, a parameter to be called the general inequality parameter. This same general inequality parameter also governs Lorenz dominance. Further, we find (in a restricted analysis involving one measure of inequality and one variate) that the general inequality parameter governs all the "statistics" in a group with subgroups – the arithmetic means and the measures of personal inequality within the subgroups – as well as the between and within components in the highly useful decomposition analysis. Thus, the general inequality parameter seems to capture the "deep structure" of inequality.

Additional work remains to be done – first and foremost finding a distribution-independent ("nonparametric") relation between personal inequality and subgroup inequality (beyond the connection made in the paper between the relative gap and the "statistics" of Theil's

MLD) and also relaxing the restriction of nonoverlapping subgroups and broadening the analysis to incorporate also empirical distributions. Nevertheless, the results reported in this paper are quite revealing and point in the direction of a unified understanding of the operation of inequality. We arrive at the tentative important conclusion that in a subset of widely used distributions, the same inequality that affects relations between individuals also affects relations between subgroups.

This unitary operation, however, may be restricted to certain kinds of distributions, exemplified by the mathematically specified two-parameter distributions (examined in this paper) and by distributions with nonintersecting Lorenz curves. For example, it is intriguing to consider that in empirical distributions, transfers within subgroup will evidently alter overall (personal) inequality but leave intact subgroup inequality. Such transfers unfortunately violate the essential element in transfers associated with two-parameter distributions and nonintersecting Lorenz curves (namely, the transfers involve persons at the very ends of the distribution, e.g., from the richest to the poorest). The tasks ahead are to establish conditions, in both mathematically specified and empirical distributions, for the monotone connection between personal inequality and subgroup inequality and to explore how societal income distributions "jump" from one variate family to another, breaking the connection between personal inequality and subgroup inequality and generating intersecting Lorenz curves.

At this point, we can conclude that in at least some societies, characterized by special income distributions, whenever a population possesses qualitative characteristics which can generate subgroups -- such as gender (following even the most liberal classifications) and some other characteristics, such as race, nativity, language, or religion – increases in inequality may operate not only on individuals but also on subgroups. It is therefore plausible that in such societies, whenever inequality disrupts social harmony and social cohesion, it attacks on two fronts, via subgroup inequality as well as personal inequality.

## REFERENCES


Allen, Roy G. D. 1938. Mathematical Analysis for Economists. New York: St. Martin's Press.

Atkinson, Anthony B. 1970. "On the Measurement of Inequality." Journal of Economic Theory 2:244-263.

______. 1975. The Economics of Inequality. London: Oxford.

Berry, Albert, Francois Bourguignon, and Christian Morrisson. 1983. "The Level of World Inequality: How Much Can One Say?" Review of Income and Wealth 29:217-243.

Blau, Francine D., and Lawrence M. Kahn. 2000. "Gender Differences in Pay." Journal of Economic Perspectives 14:75-99.

Blau, Peter M. 1974. "Presidential Address: Parameters of Social Structure." American Sociological Review 39:615-635.

Bourguignon, Francois. 1979. "Decomposable Income Inequality Measures." Econometrica 47:901-920.

Champernowne, David G., and Frank A. Cowell. 1998. Economic Inequality and Income Distribution. Cambridge, UK: Cambridge University Press.

Chipman, John. 1985. "The Theory and Measurement of Income Distribution." Advances in Econometrics 4:135-165.

Cowell, Frank A. 1977. Measuring Inequality: Techniques for the Social Sciences. New York: Wiley.

Cramer, Jan S. 1971. Empirical Econometrics. New York: Elsevier North-Holland.

Darity, William A., Jr., and Ashwini Deshpande. 2000. "Tracing the Divide: Intergroup Disparity Across Countries." Eastern Economic Journal 26:75-85.

______ and Patrick L. Mason. 1998. "Evidence on Discrimination in Employment: Codes of Color, Codes of Gender." Journal of Economic Perspectives 12:63-90.

Das, T., and Ashok Parikh. 1982. "Decomposition of Inequality Measures and a Comparative Analysis." Empirical Economics 7:23-48.

Firebaugh, Glenn. 1999. "Empirics of World Income Inequality." American Journal of

Sociology 104:1597-1630.

______. 2003. The New Geography of Global Income Inequality. Cambridge, MA: Harvard University Press.

Gastwirth, Joseph L. 1971. “A General Definition of the Lorenz Curve.” Econometrica 39:1037-1039.

______. 1972. The Estimation of the Lorenz Curve and Gini Index.” The Review of Economics and Statistics 54:306-316.

Gibbons, Jean Dickinson. 1988. "Truncated Data." P. 355 in Samuel Kotz, Norman L. Johnson, and Campbell B. Read (eds.), Encyclopedia of Statistical Sciences, Volume 9. New York: Wiley.

Gini, Corrado. 1914. “Sulla Missura della Concentrazione e della Variabilità dei Caratteri.” Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti 73:1203-1248.

Goldin, Claudia. 1990. Understanding the Gender Gap: An Economic History of American Women. New York, NY: Oxford University Press.

______. 2006. “Gender Gap.” The Concise Encyclopedia of Economics. Library of Economics and Liberty. Retrieved on June 22, 2006 from the World Wide Web: http://www.econlib.org/LIBRARY/Enc/GenderGap.html .

Harkness, Susan. 1996. “The Gender Earnings Gap: Evidence from the UK.” Fiscal Studies 17:1-36.

Hogg, Michael A., Deborah J. Terry, and Katherine M. White. 1995. “A Tale of Two Theories: A Critical Comparison of Identity Theory with Social Identity Theory.” Social Psychology Quarterly 58:255-269.

Jasso, Guillermina. 1980. "A New Theory of Distributive Justice." American Sociological Review 45:3-32.

______. 1982. "Measuring Inequality by the Ratio of the Geometric Mean to the Arithmetic Mean." Sociological Methods and Research 10:303-326.

______. 1983a. "Social Consequences of the Sense of Distributive Justice: Small-Group

Applications." Pp. 243-294 in David M. Messick and Karen S. Cook (eds.), Theories of Equity: Psychological and Sociological Perspectives. New York: Praeger.

______. 1983b. "Using the Inverse Distribution Function to Compare Income Distributions and Their Inequality." Research In Social Stratification and Mobility 2:271-306.

______. 1987. "Choosing a Good: Models Based on the Theory of the Distributive-Justice Force." Advances in Group Processes: Theory and Research 4:67-108.

______. 1993. "Analyzing Conflict Severity: Predictions of Distributive-Justice Theory for the Two-Subgroup Case." Social Justice Research 6:357-382.

______. 1999. "How Much Injustice Is There in the World? Two New Justice Indexes." American Sociological Review 64:133-168.

______ and Samuel Kotz. 2007. "A New Continuous Distribution and Two New Families of Distributions Based on the Exponential." Statistica Neerlandica 61:.

Jencks, Christopher, Marshall Smith, Henry Acland, Mary Jo Bane, David Cohen, Herbert Gintis, Barbara Heyns, and Stephan Michelson. 1972. Inequality: A Reassessment of the Effect of Family and Schooling in America. New York, NY: Basic Books.

Johnson, Norman L., Samuel Kotz, and N. Balakrishnan. 1994 (1995). Continuous Univariate Distributions, Volume 1 (2). Second Edition. New York, NY: Wiley.

Karoly, Lynn A., and Gary Burtless. 1995. "Demographic Change, Rising Earnings Inequality, and the Distribution of Personal Well-Being, 1959-1989." Demography 32:379-405.

Kleiber, Christian, and Samuel Kotz. 2003. Statistical Size Distributions in Economics and Actuarial Sciences. Hoboken, NJ: Wiley.

Hardy, Godfrey H., John E. Littlewood, and George Pólya. [1934] 1952. Inequalities. Second Edition. Cambridge, UK: Cambridge University Press.

Liao, Tim Futing. 2006. "Measuring and Analyzing Class Inequality with the Gini Index Informed by Model-Based Clustering." Sociological Methodology 36:201-224.

Leung, Kwok, Peter B. Smith, Zhongming Wang, and Haifa F. Sun. 1996. "Job Satisfaction in Joint Venture Hotels in China: An Organizational Justice Analysis." Journal of

International Business Studies 27:947-962.

Leung, Kwok, Peter B. Smith, and Zhongming Wang. 2001. "Job Attitudes and Organizational Justice in Joint Venture Hotels in China: The Role of Expatriate Managers." International Journal of Human Resource Management 12:926-945.

Lorenz, Max Otto. 1905. "Methods of Measuring the Concentration of Wealth." Journal of the American Statistical Association 9:209-219.

O'Neill, June. 2003. "The Gender Gap in Earnings: Circa 2000." American Economic Review, Papers and Proceedings.

Pearson, Karl. 1896. "Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia." Philosophical Transactions of the Royal Society of London, Series A 187:253-318.

Pietra, Gaetano. 1915. "Delle Relazioni fra Indici de Variabilità, Note I e II." Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti 74:775-804.

Reskin, Barbara F., and Denise D. Bielby. 2005. "A Sociological Perspective on Gender and Career Outcomes." Journal of Economic Perspectives 19:71-86.

Ridgeway, Cecilia L. 1991. "The Social Construction of Status Value: Gender and Other Nominal Characteristics." Social Forces 70:367-86.

______. 2001. "Inequality, Status, and the Construction of Status Beliefs." Pp. 323-340 in Jonathan H. Turner (ed.), Handbook of Sociological Theory. New York: Kluwer Academic/Plenum Press.

Schultz, T. Paul. 1998. "Inequality in the Distribution of Personal Income in the World: How It Is Changing and Why." Journal of Population Economics 11:307-344.

Schutz, Robert R. 1951. "On the Measurement of Income Inequality." American Economic Review 41:107-22.

Shenkar, Oded, and Yoram Zeira. 1987. "Human Resource Management in International Joint Ventures: Directions for Research." Academy of Management Review 12:546-557.

Shorrocks, Anthony F. 1980. "The Class of Additively Decomposable Inequality Measures."

Econometrica 48:613-626.

Theil, Henri. 1967. Economics and Information Theory. Amsterdam: North-Holland.

______. 1979. "World Income Inequality and Its Components." Economic Letters 2:99-102.

Tomaskovic-Devey, Donald, Melvin Thomas, and Kecia Johnson. 2005. "Race and the Accumulation of Human Capital across the Career: A Theoretical Model and Fixed-Effects Application." American Journal of Sociology 111:58-89.

U.S. Council of Economic Advisers. 1998a. "Annual Report of the Council of Economic Advisers." Chapter 4: "Economic Inequality Among Racial and Ethnic Groups." Available at http://www.gpoaccess.gov/usbudget/fy99/pdf/erp.pdf .

______. 1998b. "Explaining Trends in the Gender Wage Gap." Available at http://clinton4.nara.gov/WH/EOP/CEA/html/gendergap.html .

U.S. Department of State. 1998. "March Toward Equality: Significant Moments in the Civil Rights Movement." Available at http://usinfo.state.gov/products/pubs/civilrts/march.htm .

**Table 1. Three Continuous Univariate Distributions and Associated Functional Characteristics**

| Variate | Cumulative Distribution Function | Probability Density Function | Quantile Function |
|---|---|---|---|
| Lognormal $x>0, c>0$ | $F_N\left\{\dfrac{\left[\ln\left(\dfrac{x}{\mu}\right)+\dfrac{c^2}{2}\right]}{c}\right\}$ | $\dfrac{1}{xc\sqrt{2\pi}}\exp\left\{-\dfrac{\left(\dfrac{c^2}{2}+\ln\dfrac{x}{\mu}\right)^2}{2c^2}\right\}$ | $\mu\exp\left[cQ_N(\alpha)-\dfrac{c^2}{2}\right]$ |
| Pareto $x>\dfrac{\mu(c-1)}{c}, c>1$ | $1-\left[\dfrac{\mu(c-1)}{cx}\right]^c$ | $\left[\dfrac{\mu(c-1)}{c}\right]^c cx^{-c-1}$ | $\dfrac{\mu(c-1)}{c(1-\alpha)^{1/c}}$ |
| Power-Function $0<x<\dfrac{\mu(c+1)}{c}, c>0$ | $\left[\dfrac{xc}{\mu(c+1)}\right]^c$ | $\left[\dfrac{c}{\mu(c+1)}\right]^c cx^{c-1}$ | $\dfrac{\mu(c+1)\alpha^{1/c}}{c}$ |

*Notes*: For all variates, $x>0$; other restrictions as indicated. The expressions $F_N(\cdot)$ and $Q_N(\cdot)$ denote the cumulative distribution function and the quantile function, respectively, of the standard normal variate:

$$F_N(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt\,;$$

$$Q_N(\alpha) = \sqrt{2}\,\mathrm{erf}^{-1}(2\alpha-1)\,,$$

where erf denotes the error function. Inequality is a decreasing function of $c$ for the Pareto and the power-function variates and an increasing function of $c$ for the lognormal distribution.

**Table 2. Some Convenient Formulas for Four Inequality Measures in Observed and Mathematically Specified Distributions**

| Inequality Measure | Observed Distributions | Mathematically Specified Distributions |
|---|---|---|
| Gini Coefficient | $\frac{2\sum_{i=1}^{N} i x_i}{\mu N(N-1)} - \frac{N+1}{N-1}$ | $1 - \frac{2}{\mu}\int_0^1 \int_0^\alpha Q(\alpha)\, d\alpha\, d\alpha$ |
| Atkinson's Measure | $1 - \left(\frac{G}{\mu}\right)$ | $1 - \frac{\exp\left\{\int_0^1 \ln[Q(\alpha)]\, d\alpha\right\}}{\mu}$ |
| Theil's MLD | $\ln\left(\frac{\mu}{G}\right)$ | $\ln(\mu) - \int_0^1 \ln[Q(\alpha)]\, d\alpha$ |
| Pearson's Coefficient of Variation | $\frac{\sigma}{\mu}$ | $\left\{\int_0^1 [Q(\alpha)]^2 d\alpha - \mu^2\right\}^{1/2} \Big/ \mu$ |

*Notes*: The letters $x$, $i$, and $N$ denote the magnitude of $X$, the absolute rank (arranged in ascending order), and the population size, respectively. The letters $\mu$, $G$, and $\sigma$ denote the arithmetic mean, geometric mean, and standard deviation, respectively. Formulas for mathematically specified distributions are expressed in terms of the quantile function (QF) and the mean $\mu$.

**Table 3.a. Personal Inequality as a Function of the General Inequality Parameter *c*: Gini Coefficient and Atkinson's Measure**

| Variate | Measures of Personal Inequality | | | |
|---|---|---|---|---|
| | Gini Coefficient | | Atkinson's Measure | |
| | Measure | Effect of *c* | Measure | Effect of *c* |
| Lognormal | $2F_N\left(\frac{c}{\sqrt{2}}\right) - 1$ | $\frac{2}{\sqrt{2}} f_N\left(\frac{c}{\sqrt{2}}\right) > 0$ | $1 - \exp(-c^2/2)$ | $c\left[\exp\left(-\frac{c^2}{2}\right)\right] > 0$ |
| Pareto | $\frac{1}{2c-1}$ | $-\frac{2}{(2c-1)^2} < 0$ | $1 - \frac{(c-1)\exp(1/c)}{c}$ | $-\frac{\exp(1/c)}{c^3} < 0$ |
| Power-Function | $\frac{1}{2c+1}$ | $-\frac{2}{(2c+1)^2} < 0$ | $1 - \frac{(c+1)\exp(-1/c)}{c}$ | $-\frac{\exp(-1/c)}{c^3} < 0$ |

*Notes*: The expressions $F_N(\cdot)$ and $f_N(\cdot)$ denote the cumulative distribution function (CDF) and the probability density function (PDF), respectively, of the standard normal variate. The PDF is given by:

$$f_N(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty .$$

The columns headed "Effect of *c*" provide the first derivative of the inequality measure with respect to *c* and its sign. Inequality is a decreasing function of *c* for the Pareto and the power-function variates and an increasing function of *c* for the lognormal case. Thus, both the Gini and Atkinson measures of inequality are increasing functions of the variate's general inequality parameter.

**Table 3.b. Personal Inequality as a Function of the General Inequality Parameter *c*: Theil's MLD and the Coefficient of Variation**

| Variate | Measures of Personal Inequality | | | |
|---|---|---|---|---|
| | Theil's MLD | | Coefficient of Variation | |
| | Measure | Effect of *c* | Measure | Effect of *c* |
| Lognormal | $\frac{c^2}{2}$ | $c > 0$ | $\sqrt{[\exp(c^2)]-1}$ | $c\left[\exp\left(-\frac{c^2}{2}\right)\right] > 0$ |
| Pareto | $\ln\left(\frac{c}{c-1}\right) - \frac{1}{c}$ | $-\frac{1}{c^2(c-1)} < 0$ | $\frac{1}{\sqrt{c(c-2)}},\ c>2$ | $-\frac{c-1}{c(c-2)^{3/2}} < 0$ |
| Power-Function | $\ln\left(\frac{c}{c+1}\right) + \frac{1}{c}$ | $-\frac{1}{c^2(c+1)} < 0$ | $\frac{1}{\sqrt{c(c+2)}}$ | $-\frac{c+1}{c(c+2)^{3/2}} < 0$ |

*Notes*: The columns headed "Effect of *c*" provide the first derivative of the inequality measure with respect to *c* and its sign. Inequality is a decreasing function of *c* for the Pareto and the power-function variates and an increasing function of *c* for the lognormal case. Thus, both Theil's MLD and the coefficient of variation are increasing functions of the variate's general inequality parameter.

**Table 4. Subgroup Inequality as a Function of the General Inequality Parameter *c***

| Variate | Measures of Subgroup Inequality | | | |
|---|---|---|---|---|
| | Difference<br>Top Subgroup Mean - Bottom Subgroup Mean | | Ratio<br>Bottom Subgroup Mean / Top Subgroup Mean | |
| | Measure | Effect of *c* | Measure | Effect of *c* |
| Lognormal | $\left(\frac{\mu}{1-p}\right)\left\{1 - \frac{F_N[Q_N(p)-c]}{p}\right\}$ | $\left(\frac{\mu}{1-p}\right)\left\{\frac{f_N[Q_N(p)-c]}{p}\right\} > 0$ | $\frac{(1-p)F_N[Q_N(p)-c]}{p\{1-F_N[Q_N(p)-c]\}}$ | $-\left(\frac{1-p}{p}\right)\left\{\frac{f_N[Q_N(p)-c]}{\{1-F_N[Q_N(p)-c]\}^2}\right\} < 0$ |
| Pareto | $\frac{\mu[(1-p)^{-1/c}-1]}{p}$ | $\frac{\mu[(1-p)^{-1/c}\ln(1-p)]}{pc^2} < 0$ | $\frac{(1-p)^{1/c}-(1-p)}{p}$ | $-\frac{(1-p)^{1/c}\ln(1-p)}{pc^2} > 0$ |
| Power-Function | $\frac{\mu(1-p^{1/c})}{1-p}$ | $\frac{\mu[p^{1/c}\ln(p)]}{c^2(1-p)} < 0$ | $\frac{p^{1/c}(1-p)}{1-p^{(c+1)/c}}$ | $\frac{p^{1/c}(p-1)\ln(p)}{c^2[p^{(c+1)/c}-1]^2} > 0$ |

*Notes*: The terms $F_N(\cdot)$, $f_N(\cdot)$, and $Q_N(\cdot)$ denote the cumulative distribution function, the probability density function, and the quantile function, respectively, of the standard normal variate. The subgroup split is represented by the censoring point $p$, where $0 < p < 1$; accordingly, the proportion in the bottom subgroup is given by $p$, and the proportion in the top subgroup is $(1 - p)$. The columns headed "Effect of $c$" provide the first partial derivative of the inequality measure with respect to $c$ and its sign. Inequality is a decreasing function of $c$ for the Pareto and the power-function variates and an increasing function of $c$ for the lognormal variate. In the difference-based measure of subgroup inequality, the larger the measure, the greater the inequality; in the ratio-based measure of subgroup inequality, the larger the measure, the smaller the inequality. Thus, for both measures, subgroup inequality is an increasing function of the variate's general inequality parameter.

**Table 5. Lorenz Curve as a Function of the General Inequality Parameter *c***

| Variate | Lorenz Curve | Effect of *c* |
|---|---|---|
| Lognormal | $F_N\left[Q_N(\alpha) - c\right]$ | $-f_N\left[Q_N(\alpha) - c\right] < 0$ |
| Pareto | $1 - (1 - \alpha)^{\frac{c-1}{c}}$ | $-\dfrac{(1-\alpha)^{\frac{c-1}{c}} \ln(1-\alpha)}{c^2} > 0$ |
| Power-Function | $\alpha^{\frac{c+1}{c}}$ | $-\dfrac{\alpha^{\frac{c+1}{c}} \ln(\alpha)}{c^2} > 0$ |

*Notes*: The terms $F_N(\cdot)$, $f_N(\cdot)$, and $Q_N(\cdot)$ denote the cumulative distribution function, the probability density function, and the quantile function, respectively, of the standard normal variate. The columns headed "Effect of *c*" provide the first partial derivative of the Lorenz curve with respect to *c* and its sign. Inequality is a decreasing function of *c* for the Pareto and the power-function variates and an increasing function of *c* for the lognormal. Thus, as the general inequality parameter moves in the direction of more inequality, the Lorenz curve is lower on the grid.

**Table 6. Theil's MLD and Its Decomposition**

| Feature | Bottom Subgroup | Top Subgroup |
|---|---|---|
| Subgroup Mean | $\mu_B$ | $\mu_T$ |
| Subgroup MLD | $E\left[\ln\left(\frac{\mu_B}{X}\right)\right] = \ln\left(\frac{\mu_B}{G_B}\right)$ | $E\left[\ln\left(\frac{\mu_T}{X}\right)\right] = \ln\left(\frac{\mu_T}{G_T}\right)$ |
| Overall MLD | $E\left[\ln\left(\frac{\mu}{X}\right)\right] = \ln\left(\frac{\mu}{G}\right)$ | |
| Between Component | $p\ln\left(\frac{\mu}{\mu_B}\right) + (1-p)\ln\left(\frac{\mu}{\mu_T}\right) = -p\ln\left(\frac{\mu_B}{\mu_T}\right) + \ln\left(\frac{\mu}{\mu_T}\right)$ | |
| Within Component | $p\ln\left(\frac{\mu_B}{G_B}\right) + (1-p)\ln\left(\frac{\mu_T}{G_T}\right) = p\ln\left(\frac{\mu_B}{\mu_T}\right) - p\ln\left(\frac{G_B}{G_T}\right) + \ln\left(\frac{\mu_T}{G_T}\right)$ | |

*Notes*: The subgroup split is represented by the censoring point $p$, where $0 < p < 1$; accordingly, the proportion in the bottom subgroup is given by $p$, and the proportion in the top subgroup is $(1 - p)$. The letters $\mu$ and $G$ denote the arithmetic mean and geometric mean, respectively; and $B$ and $T$ denote the bottom and top subgroups, respectively; and the operator $E(\cdot)$ denotes the expected value.

**Table 7. Theil's MLD and Its Decomposition in the Pareto Variate**

| Feature | Bottom Subgroup | Top Subgroup |
|---|---|---|
| Subgroup Mean | $\frac{\mu - \mu(1-p)^{(c-1)/c}}{p}$ | $\mu(1-p)^{-1/c}$ |
| Subgroup MLD | $\ln\left(\frac{(1-p)^{1/c} - (1-p)}{p}\right) - \frac{\ln(1-p)}{cp} + \ln\left(\frac{c}{c-1}\right) - \frac{1}{c}$ | $\ln\left(\frac{c}{c-1}\right) - \frac{1}{c}$ |
| Overall MLD | $ln\left(\frac{c}{c-1}\right) - \frac{1}{c}$ | |
| Between Component | $-p\ln\left(\frac{(1-p)^{1/c} - (1-p)}{p}\right) + \frac{\ln(1-p)}{c}$ | |
| Within Component | $p\ln\left(\frac{(1-p)^{1/c} - (1-p)}{p}\right) - \frac{\ln(1-p)}{c} + \ln\left(\frac{c}{c-1}\right) - \frac{1}{c}$ | |

*Notes*: Inequality is a decreasing function of $c$ for the Pareto. As inequality declines (i.e., as $c$ increases), both the between component and the within component decline, as does the top subgroup mean, but the bottom subgroup mean increases.

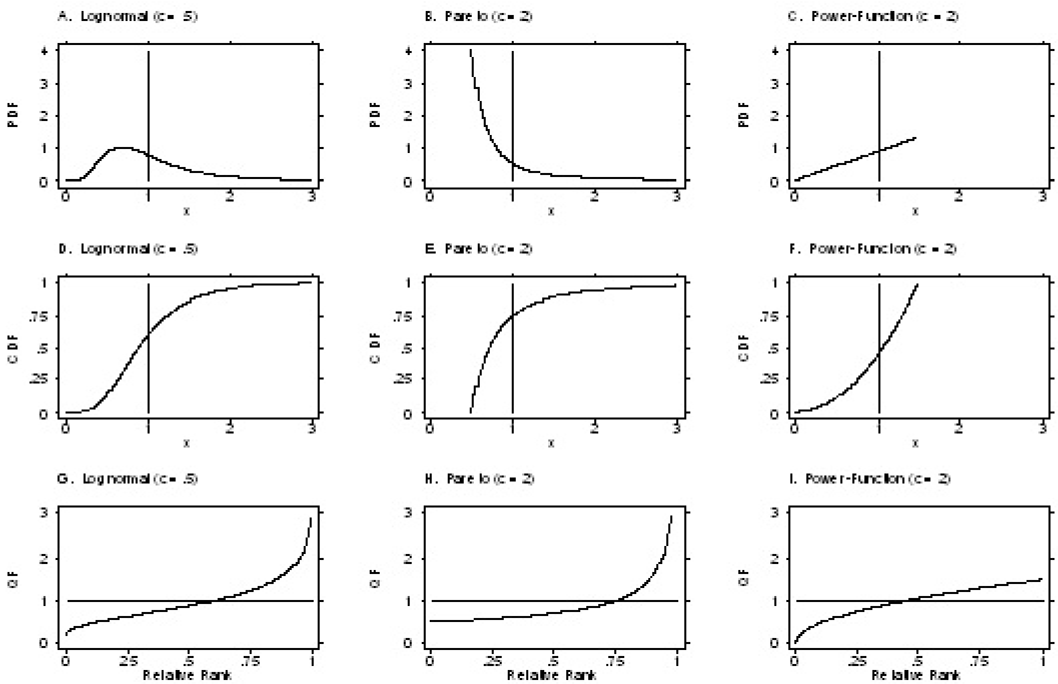


Figure 1. Basic Associated Functions (PDF, CDF, and QF) in Three Continuous Univariate Distributions. Arithmetic mean fixed at unity and displayed as either a vertical line (PDF and CDF) or horizontal line (QF).

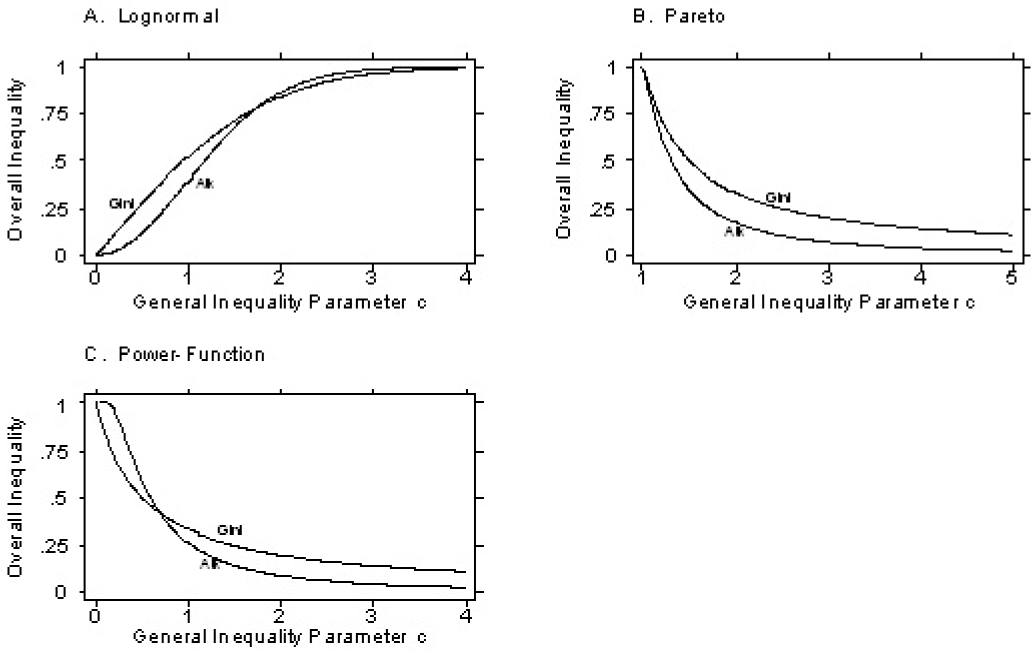


Figure 2.a. Measures of Overall Inequality, Expressed as Functions of the General Inequality Parameter *c*: Gini Coefficient and Atkinson's Measure. Inequality is a decreasing function of *c* for the Pareto and the power-function variates (plots B and C) and an increasing function of *c* for the lognormal case (plot A). Thus, both the Gini and Atkinson measures of inequality are increasing functions of the variate's general inequality parameter.

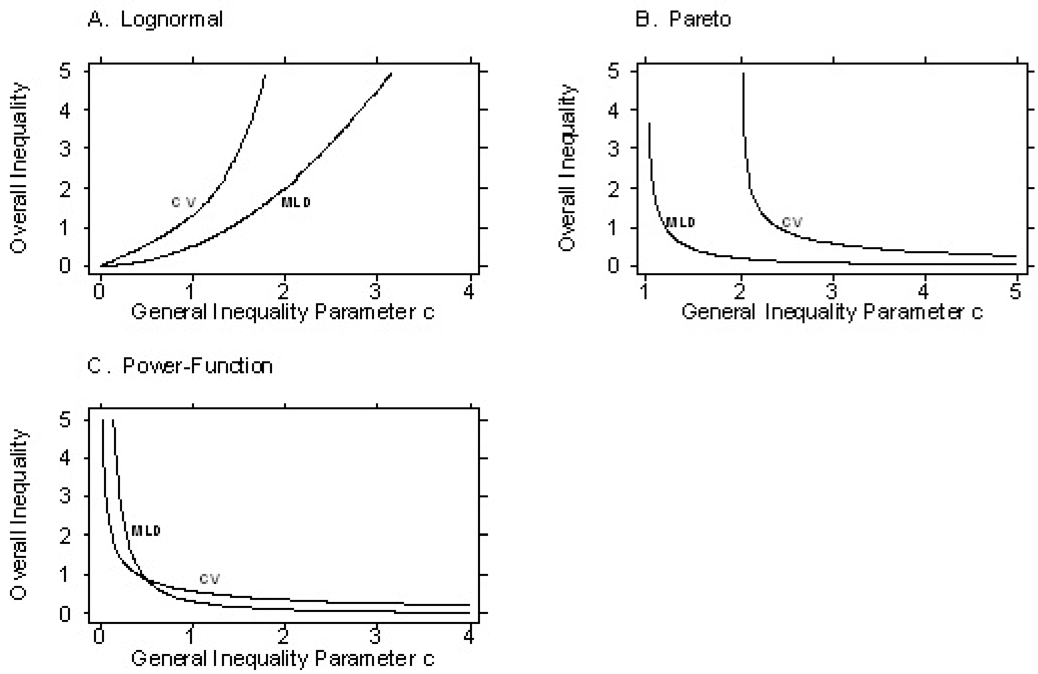


Figure 2.b. Measures of Overall Inequality, Expressed as Functions of the General Inequality Parameter *c*: Theil's MLD and the Coefficient of Variation. Inequality is a decreasing function of *c* for the Pareto and the power-function variates (plots B and C) and an increasing function of *c* for the lognormal case (plot A). Thus, both Theil's MLD and the coefficient of variation are increasing functions of the variate's general inequality parameter.

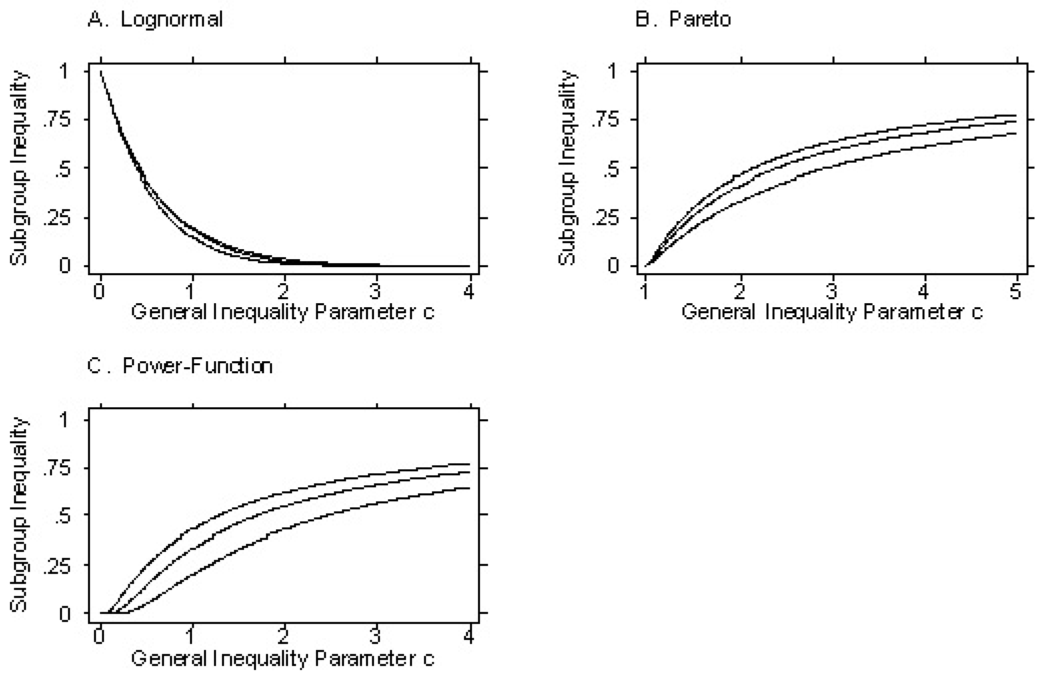


Figure 3. Subgroup Inequality (Relative Gap: Ratio of the Bottom-Subgroup Mean to the Top-Subgroup Mean), Expressed as Function of the General Inequality Parameter *c*. Each plot presents graphs for three cases, defined by three subgroup splits: .25, .5, and .75. The higher the vertical placement of the graph, the higher the ratio of the bottom-subgroup mean to the top-subgroup mean – that is, the lower the subgroup inequality. In the lognormal plot, the three curves are very close and they intersect across the range. In the Pareto plot, the subgroup splits are, from lowest to highest, .75, .5, and .25. In the power-function plot, the subgroup splits are, from lowest to highest, .25, .5, .75. Thus, as will become clearer in Figure 4, the effect of the subgroup split on subgroup inequality differs across the three variates. But the effect of the general inequality parameter *c* is invariant: As general inequality increases, so does subgroup inequality.

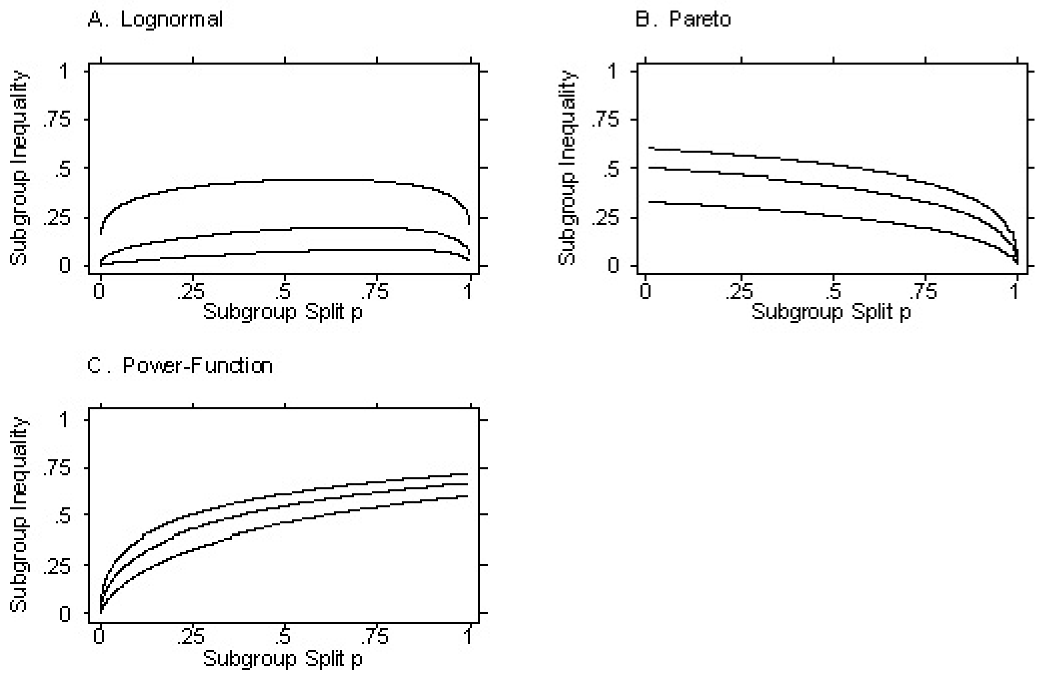


Figure 4. Subgroup Inequality (Relative Gap: Ratio of the Bottom-Subgroup Mean to the Top-Subgroup Mean), Expressed as Function of the Subgroup Split *p*. Each plot presents graphs for three cases, defined by three values of the general inequality parameter: $c$ = 1.5, 2, and 2.5. As shown also in Figure 3, the greater the general inequality, the greater the subgroup inequality. But, more pointedly in these plots, the effect of the subgroup split on subgroup inequality differs across the three variates. In the Pareto and power-function, the effect of the subgroup split is monotonic but in opposite directions – in the Pareto, the smaller the proportion in the bottom subgroup, the lower the subgroup inequality (higher ratio of bottom-subgroup mean to top-subgroup mean), while in the power-function, the smaller the proportion in the bottom subgroup, the higher the subgroup inequality. In the lognormal, the effect of the subgroup split is nonmonotonic, increasing to a peak in the upper part of the range, subsequently decreasing.

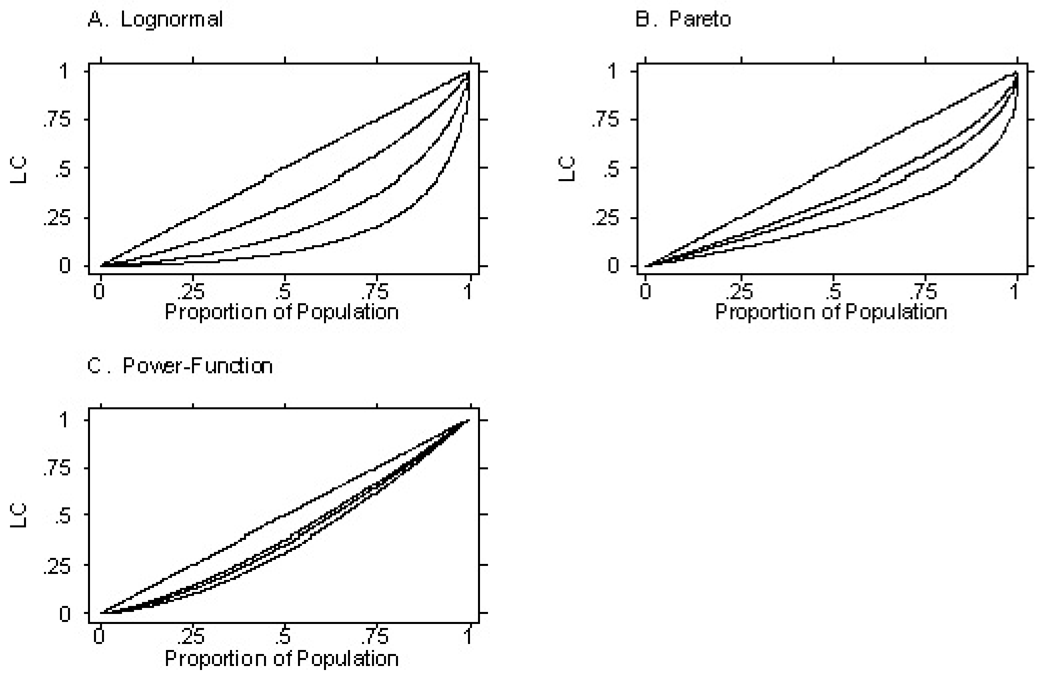


Figure 5. Lorenz Curves in Three Variates. Each plot presents graphs for three members of each variate family, defined by three values of the general inequality parameter: $c$ = 1.5, 2, and 2.5 for the Pareto and power-function variates and .5, 1, and 1.5 for the lognormal variate.

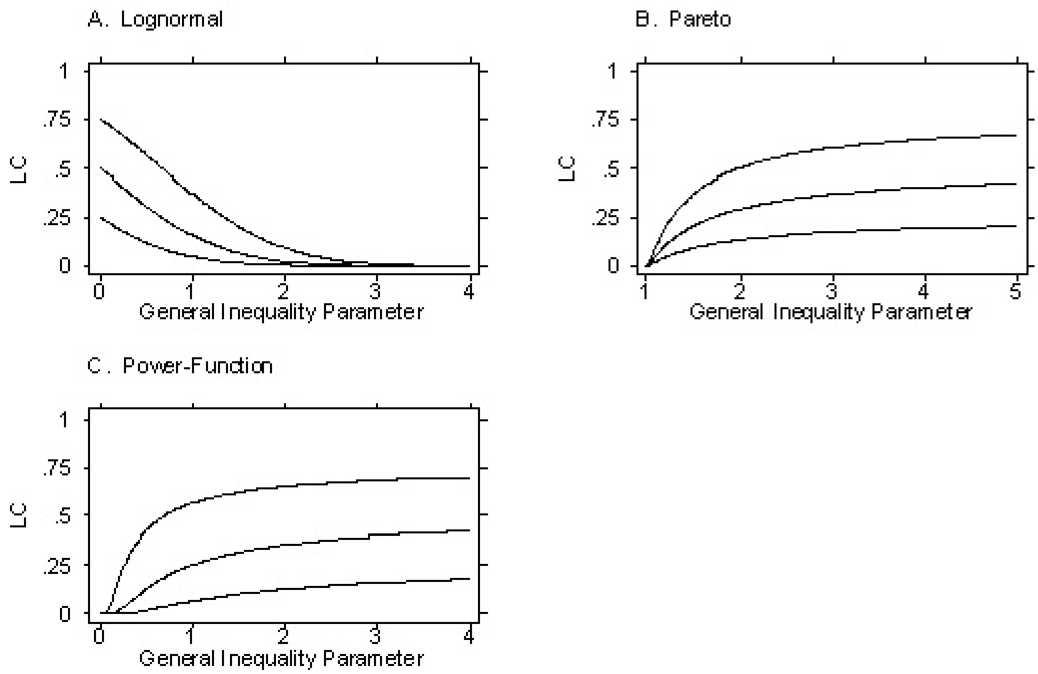


Figure 6. Lorenz Curve Formulas in Three Variates, Expressed as Functions of General Inequality Parameter *c*. Each plot presents graphs for three members of each variate family, defined by three values in the domain of the Lorenz curve: .25, .5, and .75.

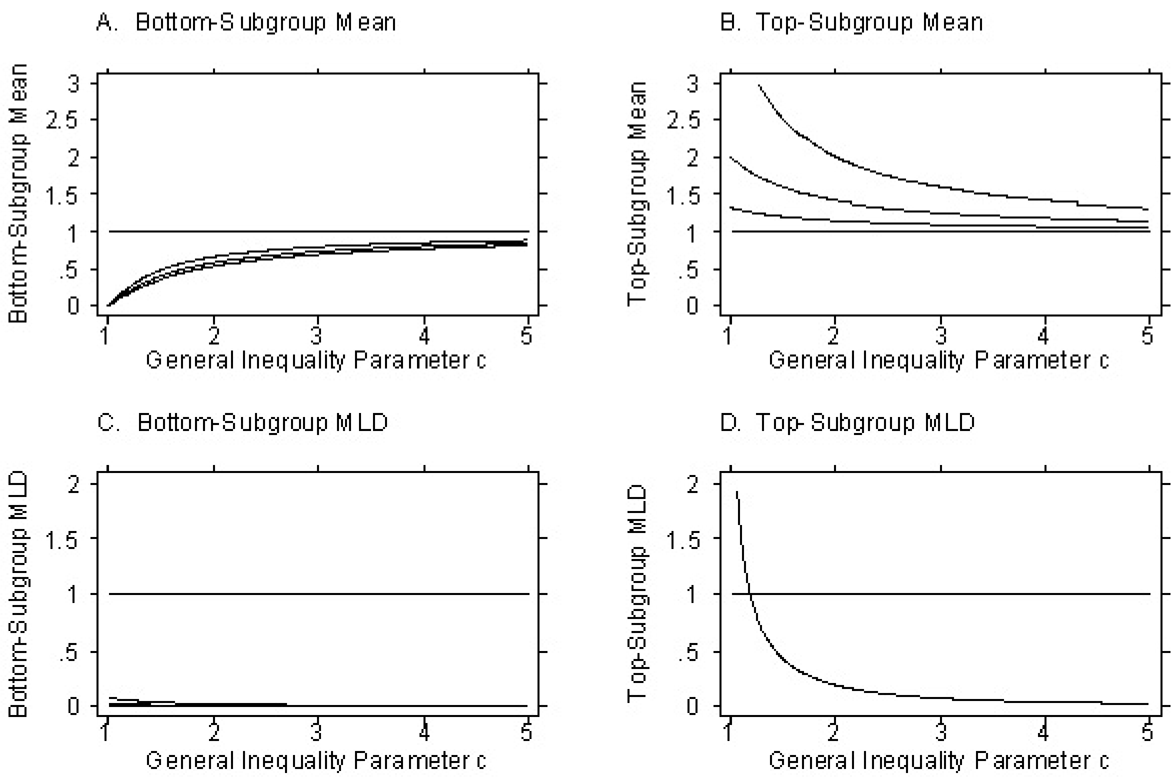


Figure 7. Mean and MLD in the Bottom and Top Subgroups of the Pareto Distribution, Expressed as Functions of General Inequality Parameter *c*. Plots for the two means and the bottom-subgroup MLD each show three curves, corresponding to subgroups splits of .25, .5, and .75. The top-subgroup MLD is a function solely of *c*; hence a single curve appears in panel D. The overall mean μ, which appears in the subgroup means, is set to unity.

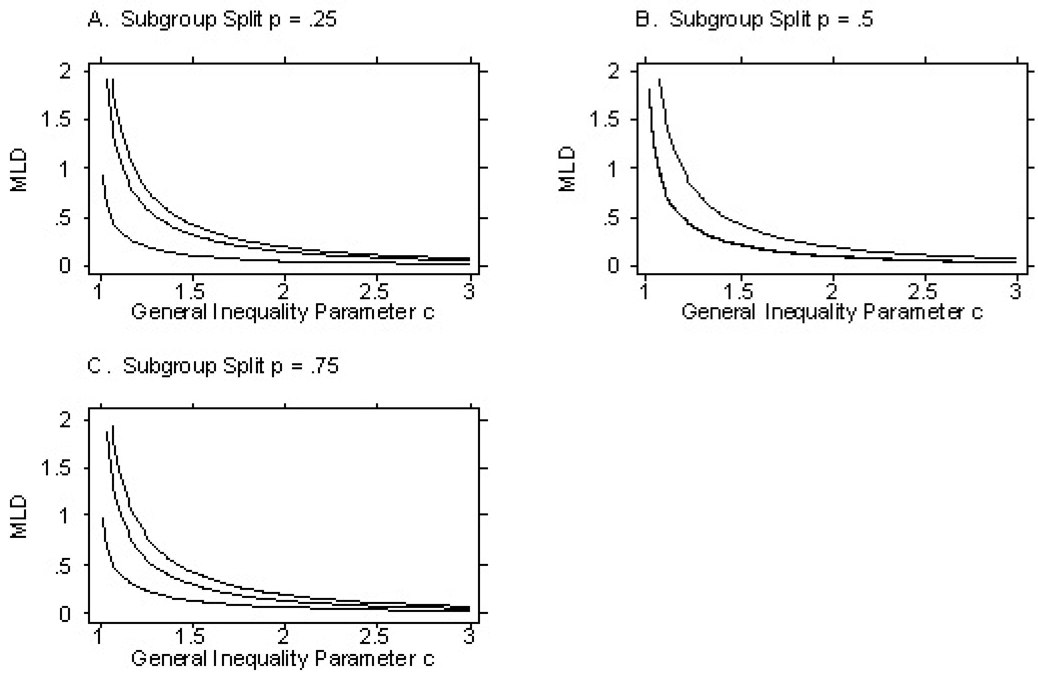


Figure 8. Overall MLD, Between Component, and Within Component in the Pareto Distribution, Expressed as Functions of General Inequality Parameter *c*. The overall MLD has the highest curve in each plot. The two components are very similar at $p = .5$. The between component is smaller than the within component at $p = .25$ and larger at $p = .75$.

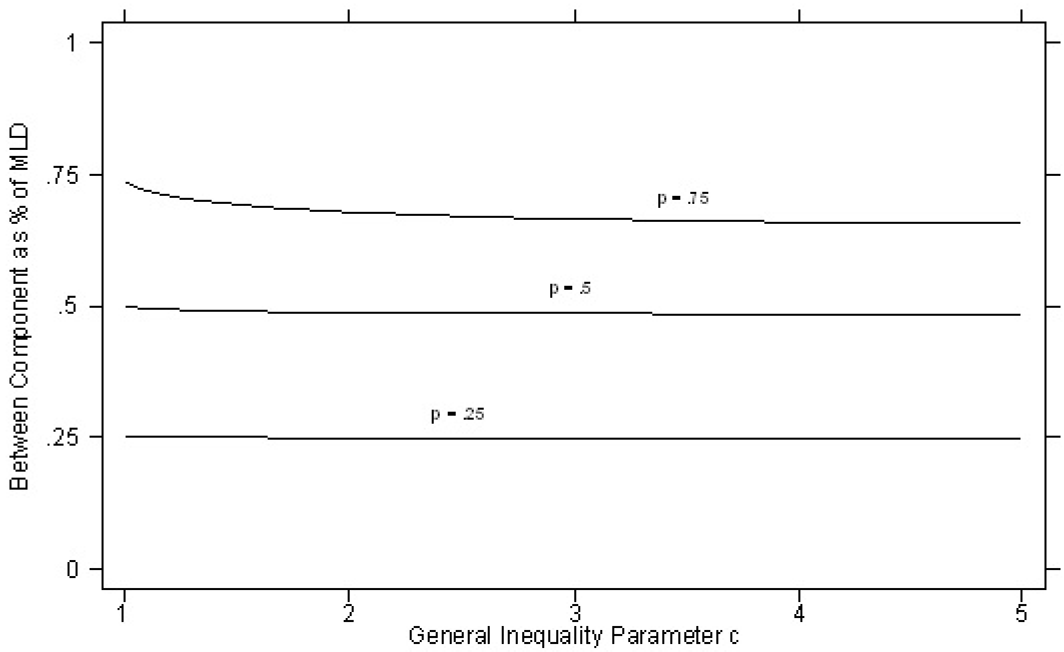


Figure 9. Between Component as Percentage of Overall MLD in the Pareto Distribution, Expressed as Function of General Inequality Parameter *c*, for Three Subgroup Splits. Values of the subgroup split *p* are .25, .5, and .75. All three curves are monotonically decreasing in *c*, though it is difficult for the naked eye to see the decrease in the bottom curve.

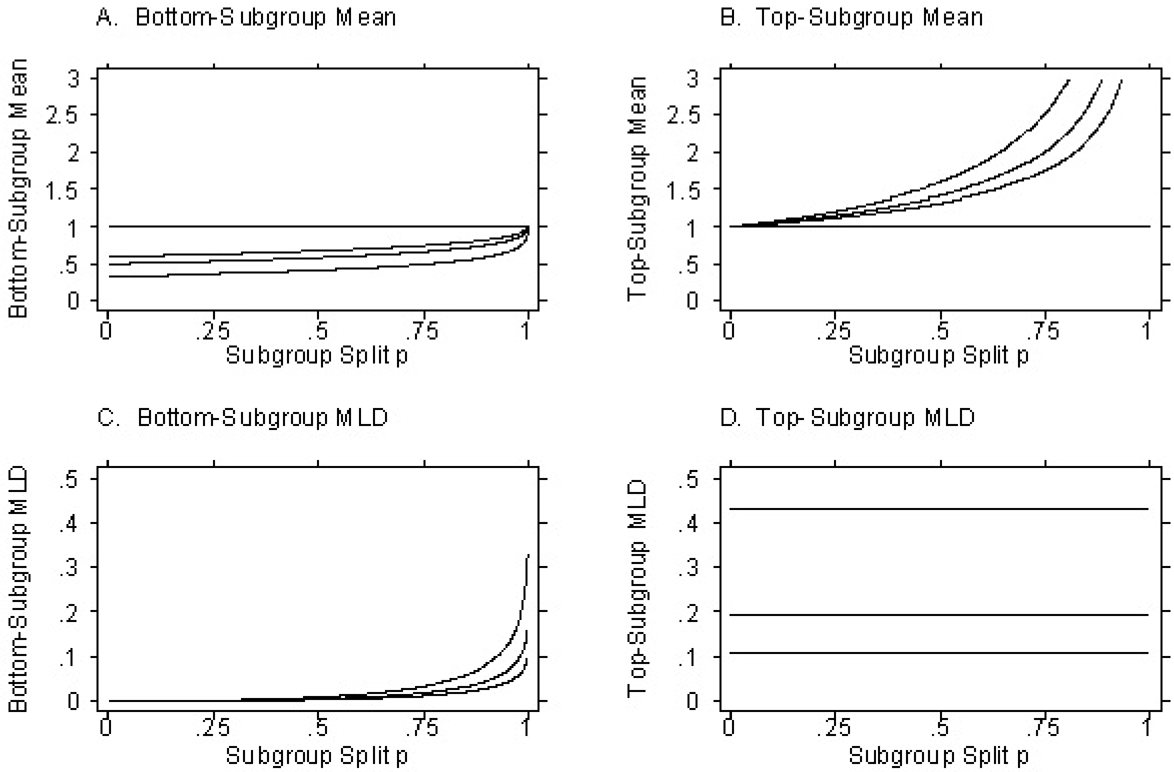


Figure 10. Mean and MLD in the Bottom and Top Subgroups of the Pareto Distribution, Expressed as Functions of Subgroup Split *p*. Plots each show three curves, corresponding to values of *c* of 1.5, 2, and 2.5. The overall mean μ, which appears in the subgroup means, is set to unity. The three graphs for the top-subgroup MLD (panel D) are perfectly flat, as that measure is invariant over *p*.

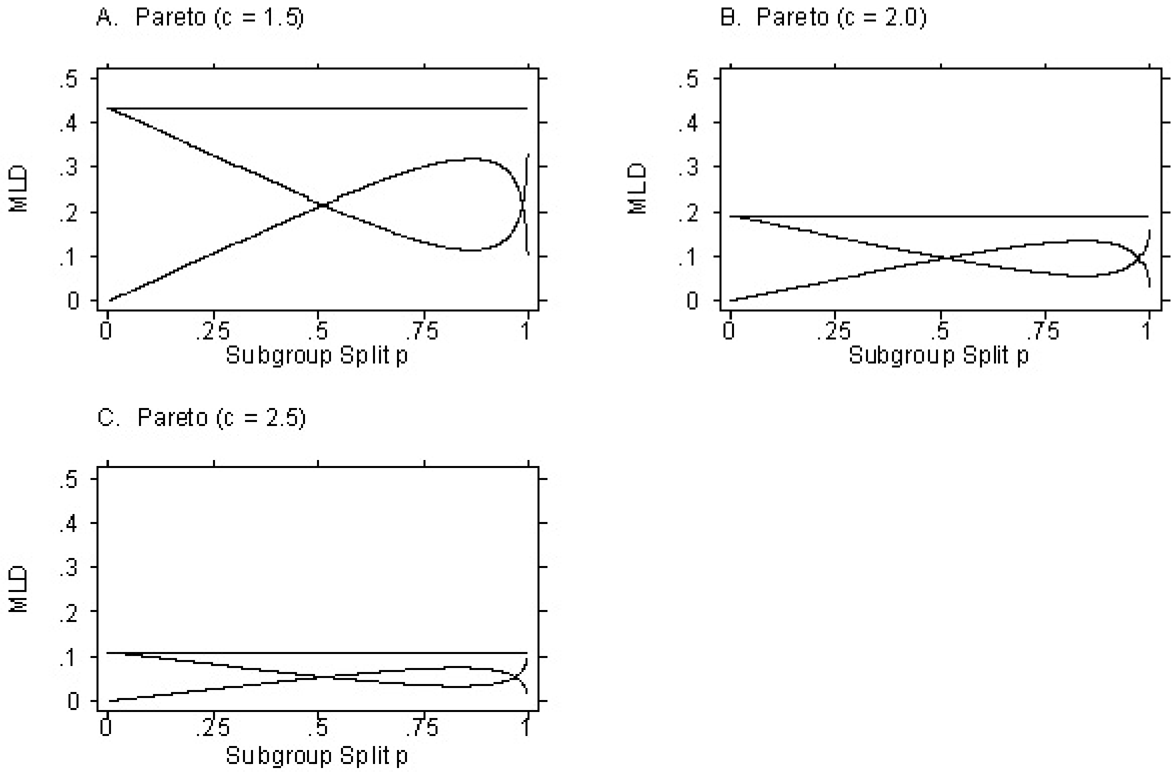


Figure 11. Overall MLD, Between Component, and Within Component in the Pareto Distribution, Expressed as Functions of Subgroup Split *p*. The overall MLD has the highest curve in each plot. The between and within components intersect twice, with the between component smaller than the within component to the left of *p* approximately equal to .5 and again to the right of a high value of *p*.

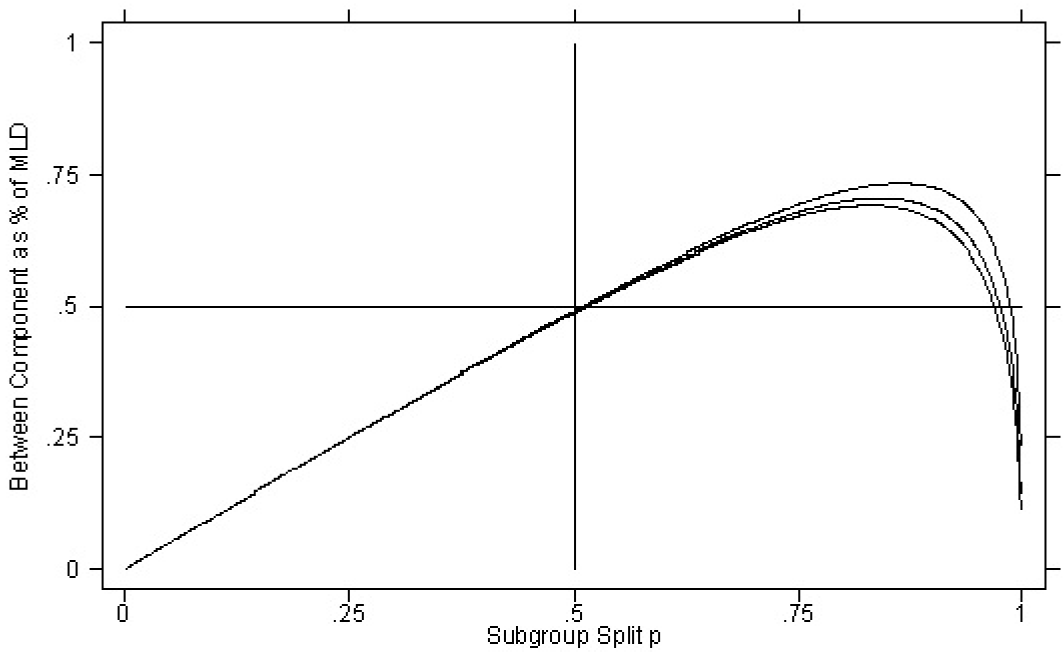


Figure 12. Between Component as Percentage of Overall MLD, in Three Members of the Pareto Variate, Expressed as Function of the Subgroup Split *p*. The Paretos are defined by three values of the general inequality parameter: $c$ = 1.5, 2, and 2.5.

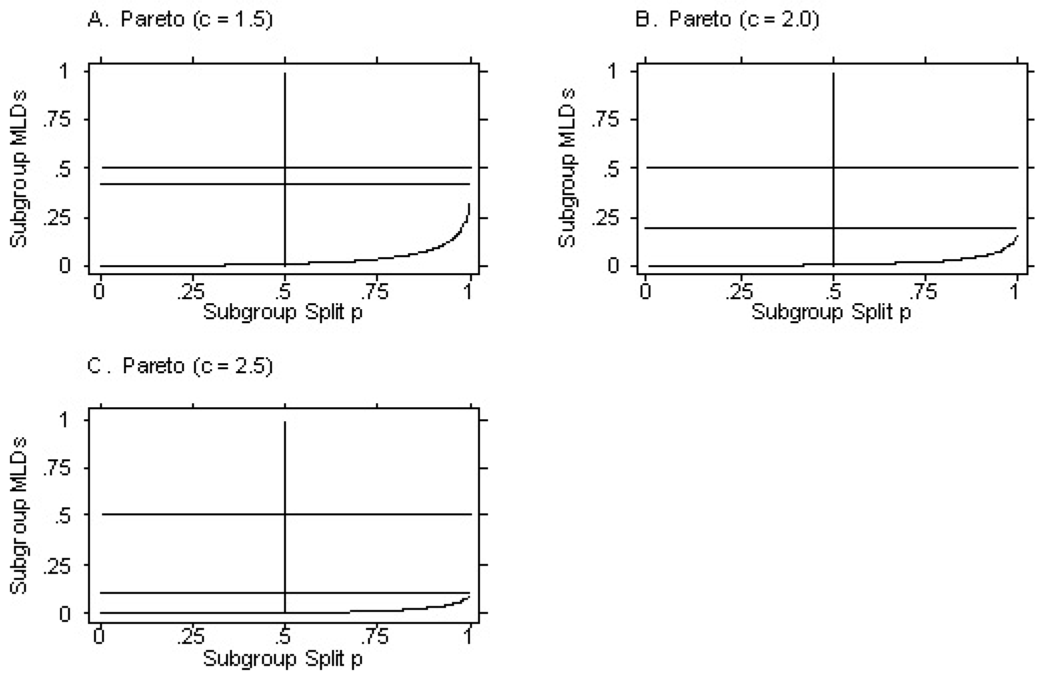


Figure 13. MLD in the Bottom and Top Subgroups, for Three Members of the Pareto Variate Family. In each plot, the top curve depicts the top-subgroup MLD and the bottom curve the bottom-subgroup MLD.

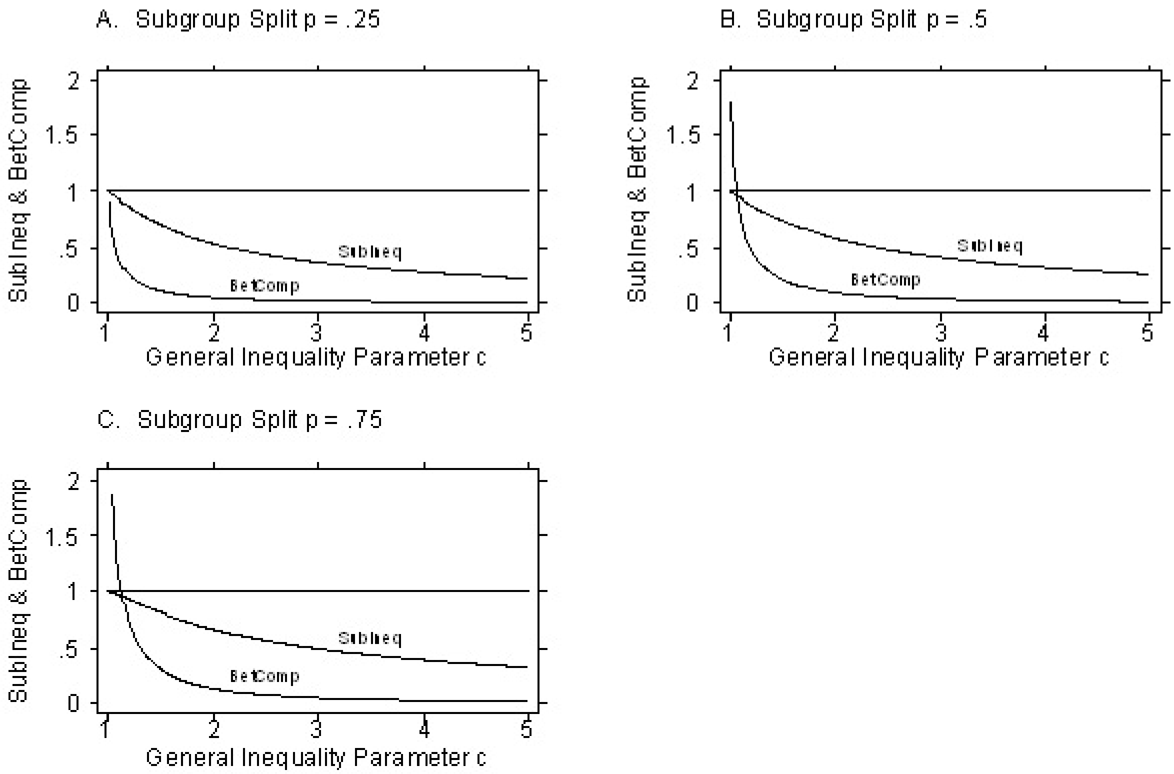


Figure 14. “Head-to-Head” Contrast of Between Component and Subgroup Inequality in the Pareto Distribution, Expressed as Functions of General Inequality Parameter *c*. Subgroup inequality is represented by 1 minus the relative gap (1 minus the ratio of the bottom-subgroup mean to the top-subgroup mean).

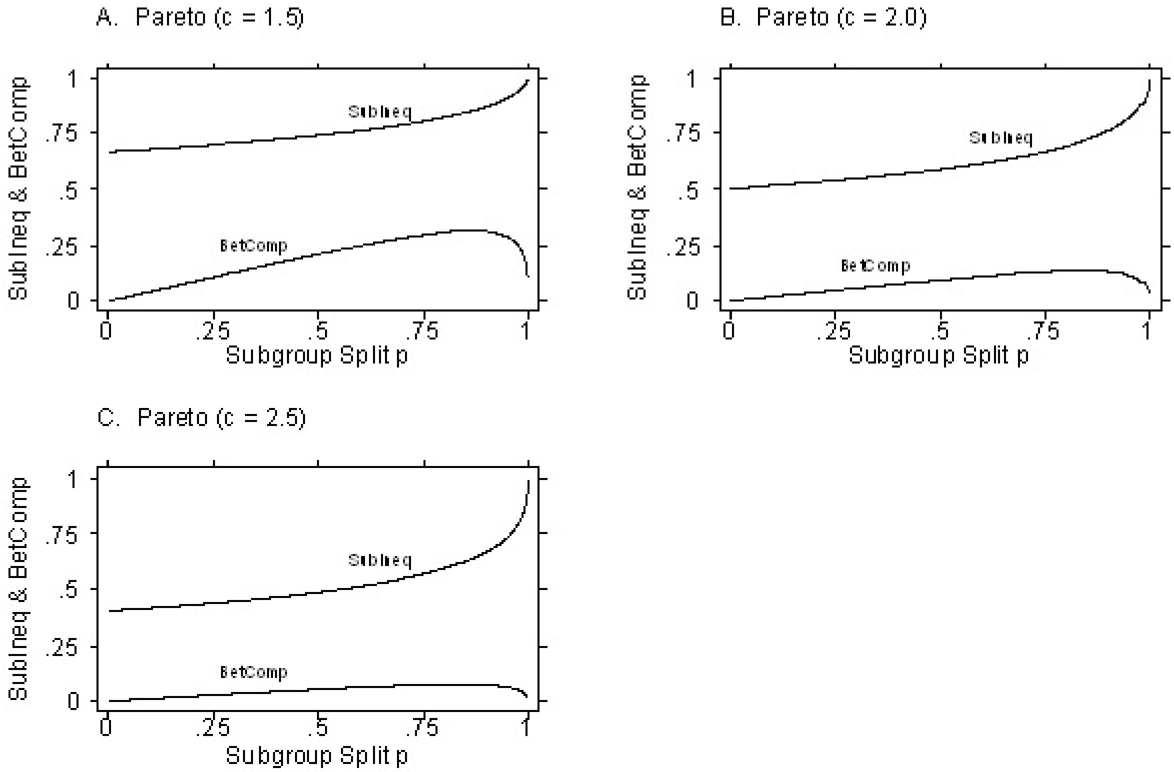


Figure 15. "Head-to-Head" Contrast of Between Component and Subgroup Inequality in the Pareto Distribution, Expressed as Functions of Subgroup Split *p*. Subgroup inequality is represented by 1 minus the relative gap (1 minus the ratio of the bottom-subgroup mean to the top-subgroup mean).

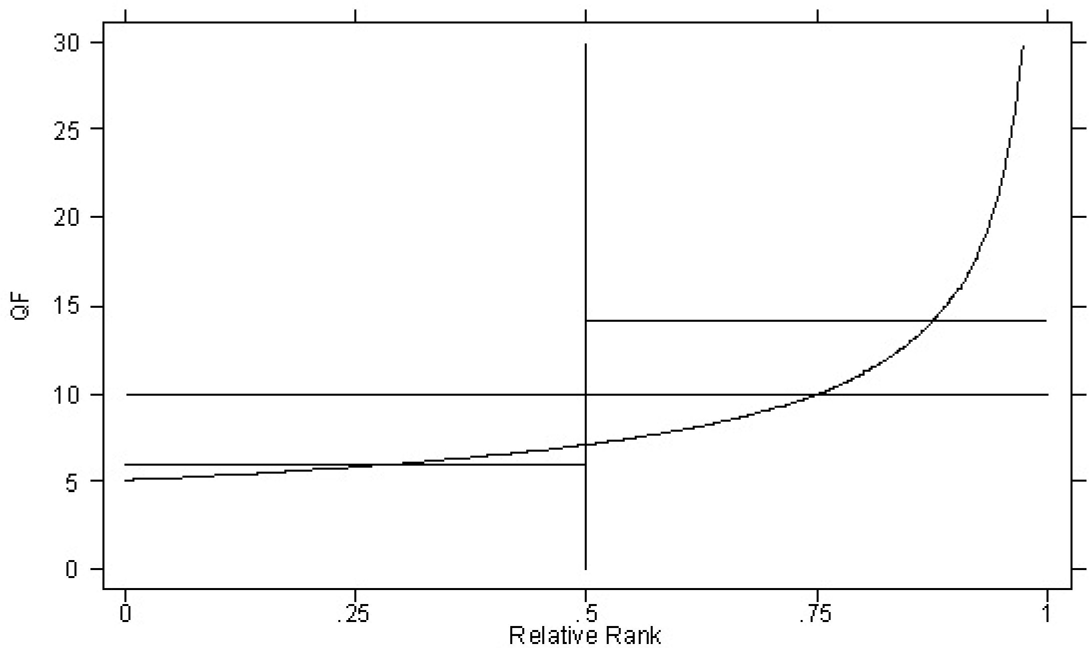


Figure 16. Pareto Distribution ($c$ = 2), with Two Equally-Sized Subgroups. Horizontal lines indicate the overall mean (10), bottom-subgroup mean (5.858), and top-subgroup mean (14.142).

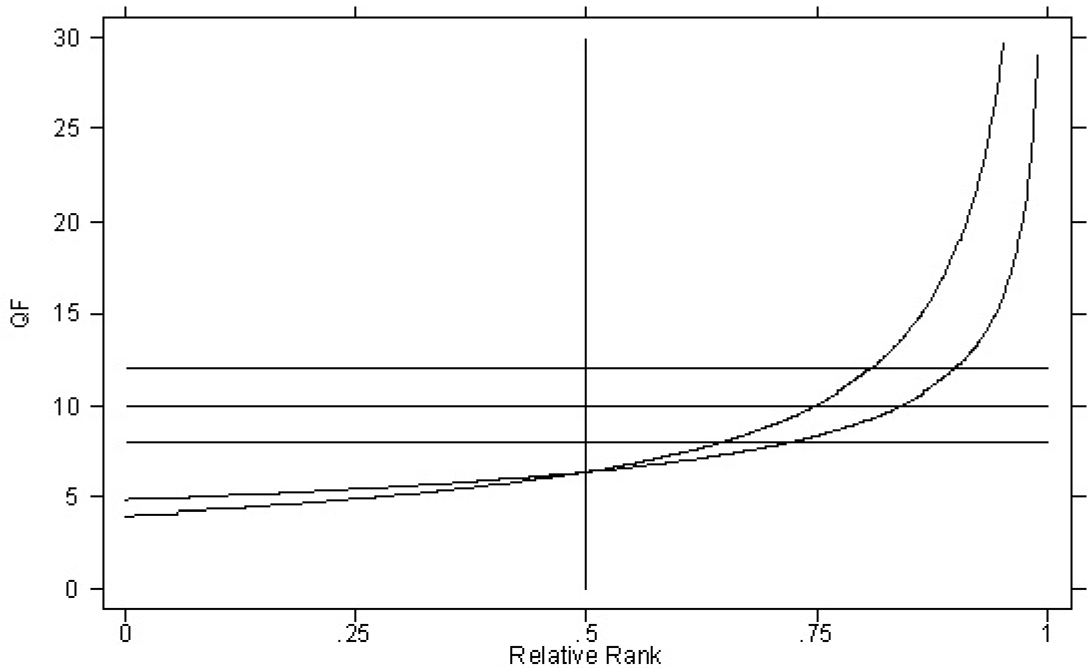


Figure 17. Two Pareto Distributions ([μ = 8, *c* = 2.5] and [μ = 12, *c* = 1.5]). Horizontal lines indicate the overall mean (10) and the two subgroup means.