

Extreme Sample Selection Bias: Conditions that Cause the Correlation Between Two Variables to Switch Signs

Tim Groseclose
Departments of Political Science and Economics
UCLA
Los Angeles, CA 90095

31 July 2006

Abstract

It is well-known that if researchers gather a truncated sample, then they will bias their regression estimates. For instance, such estimates will attenuate from their true value toward zero, if the the errors of the regression follow a normal distribution (Heckman, 1979). I analyze cases where the sample is *extremely* truncated—that is, where the threshold for selecting the sample approaches infinity, making the sample infinitesimally small relative to the population. I prove that when this happens, regression estimates do not just attenuate, they become zero or reverse signs. I illustrate this theoretical result with three empirical examples. The first examines how the personal wealth of politicians can affect their ability to win elections. While this ability is surely greater if a candidate is rich, if a sample only includes incumbents of the U.S. House (that is, the tiny portion of the population who have proven themselves to be top campaigners), then we find the opposite. Another illustration shows that, although white sprinters seem to be slower than black sprinters, white sprinters who pass a certain threshold tend to run faster than the black sprinters who pass the same threshold. A third illustration shows that, although it is reasonable to believe that people with high SAT scores will earn more income than those with low scores, if we look only at people who attended elite universities, then we find the opposite.

1. Introduction

It is well-known that if researchers gather a truncated sample, then they will bias their regression estimates. For instance, such estimates will attenuate from their true value toward zero, if the errors of the regression follow a normal distribution (Heckman, 1979). I analyze cases where the sample is *extremely* truncated—that is, where the threshold for selecting the sample approaches infinity, making the sample infinitesimally small relative to the population. I prove that when this happens, regression estimates do not just attenuate, they become zero or reverse signs. The result is very robust. That is, to prove it, essentially all I need to assume is that the distribution of the error term has an infinite upper tail.

To see the intuition of the result, suppose that (in the population) millionaires tend to be better political candidates than non-millionaires. Suppose, however, that a researcher only collected observations about incumbents. If so, then the non-millionaires in the sample would have other features that help them electorally; otherwise they would not be incumbents. Meanwhile, the millionaires do not need as many of the other features to get elected. Consequently, in a sample of incumbents, the millionaires will have less of the other features than the non-millionaires. Not only that, under a special technical condition, the non-millionaires will be so blessed with the other features that they overcome their wealth disadvantage, and on net they are more successful at reelection than millionaires.

The technical condition is that the “other features” follow a distribution that has a decreasing hazard rate. If a distribution has a decreasing hazard rate, then it will exhibit some unintuitive properties. For instance, suppose that we want to predict stopping points among a group of walkers, and suppose these stopping points exhibit a decreasing hazard rate. Now suppose that one walker joins another after the first has walked six miles. Suppose after walking an additional four miles the two walkers are still together. Who is more likely to quit first? Although intuition may suggest that it is the one who has walked ten miles, the decreasing-hazard-rate assumption implies that it is the one who has walked four.

Applied to the election example, the “other features” of the non-millionaire candidates had to be greater (i.e. to “walk longer”) to allow those candidates to win their first election. Now, which types of candidates are more likely to be just barely good enough to win election (i.e. “to quit walking” just barely after making the threshold to win office), and which ones are more likely to be significantly better than just good enough? Like the ten-mile walker,

the candidates whose “other features” are greater—i.e. the non-millionaires—are the ones more likely to be better than just good enough. Consequently, the non-millionaires will be the better candidates in a truncated sample, even though wealth actually *causes* a candidate to be better.

Next, I show that essentially all infinitely-upper-tailed distributions have a certain property. This is that at some point in the upper tail, the hazard rate of the distribution either (i) begins to decrease continually, or (ii) becomes arbitrarily close to a constant rate. If not, the hazard rate must increase continually, and if it did this it would be impossible for it to have an infinite upper tail.¹

If the hazard rate becomes continually decreasing, then the correlation between millionaire status and campaign ability becomes negative. If the hazard rate becomes constant, then the correlation between the two factors becomes zero.

To illustrate this theoretical result, I examine three empirical examples. One shows that the above intuition is not just a theoretical exercise: At least in the U.S. House, non-millionaires really do fare better at reelection than millionaires. Another example examines sports where one race or sex seems to have an advantage over another race or sex. I show, however, that if we restrict the sample only to the very elite of the sport, then the disadvantaged race or sex actually performs better than the other race or sex. For instance, as *Track and Field News* noted in July 2004, almost no white runners had qualified for the finals of a U.S. or World championship 400-meter race between 1968 and 2000. However, shortly after *Track and Field News* printed this, a white runner, Jeremy Wariner, qualified for the finals of the U.S. championship. Not only did he run faster than the average black qualifier, he won the race. Moreover, a few weeks later he won the Olympic 400m, and *Track and Field News* declared that “The Wariner era has begun.”² Tiger Woods, the golf star, and Danica Patrick, the race-car driver, are two other illustrations of the phenomenon. A third illustration involves SAT scores and income. Although we might expect those with

¹ To see this, consider the example of walkers, and consider the probability that a walker will quit during the next mile, given that he or she has already walked n miles. Note that this probability is a function of n . Suppose the probability continues to increase as n increases, and it never reaches an asymptote—i.e. the hazard rate does not approach a constant. If so, then the probability function must reach 1.0 at some point, which means that if the walker reaches a certain mile, then he or she is guaranteed to stop during the next mile. This means that the miles-walked distribution cannot have an infinite upper tail.

² “Wariner!,” Bob Hersh, *Track and Field News*, October, 2004, p. 10

high scores to be richer than those with low scores, a famous study finds the opposite. The reason, I suggest, is that the study’s sample only included individuals who had matriculated to an elite university—i.e. it was due to a truncated sample.

2. General Setup and Two Examples

Let $y = x + \epsilon$, where $y \in \mathfrak{R}$ is some observable outcome, such as a 400m time or a politician’s reelection margin of victory; $x \in \mathfrak{R}$ is an observable factor, such as the wealth of a politician or an indicator that a runner is black; and $\epsilon \in \mathfrak{R}$ comprises all other factors that contribute to the outcome y .

Define $f(z)$ and $F(z)$ respectively as the pdf and cdf of ϵ . Define μ as the expected value of ϵ . Define $\bar{\mu}(z) = E[\epsilon | \epsilon \geq z]$ and $\underline{\mu}(z) = E[\epsilon | \epsilon \leq z]$. Define the hazard rate as $h(z) = f(z)/[1 - F(z)]$. I assume the following about the distribution, expectation and conditional expectation of ϵ .

- A1: $f()$ has support $[\underline{\epsilon}, \infty]$ (where $\underline{\epsilon}$ may be $-\infty$);
- A2: $f()$ is positive over its entire support, but $\lim_{z \rightarrow \infty} f(z) = 0$.
- A3: $f()$ is everywhere differentiable.
- A4: The limits of (i) $z[1 - F(z)]$ and (ii) $\frac{d}{dz} \frac{1 - F(z)}{f(z)}$ exist (but are not necessarily finite).
- A5: μ is finite.
- A6: For all finite z , $\bar{\mu}(z)$ is finite. (This and A5 imply that $\underline{\mu}(z)$ is finite.)

Essentially all distribution functions that have an infinite right tail—including the normal, log-normal, logit, Pareto, exponential, Weibull, and Gamma—satisfy these conditions.

Let $c \in \mathfrak{R}$ be a cutoff value for selecting a sample. Assume that for any element in the sample, $y \geq c$. Note that x has a positive, causal effect on y , specifically $dy/dx = 1$. However, we want to determine if x is correlated with y after the sample is truncated. Specifically, let $\bar{y} = E[y | y \geq c]$. We want to determine $d\bar{y}/dx$.

Some simple calculations show

$$\begin{aligned}
 \frac{d\bar{y}}{dx} &= \frac{d}{dx} E[x + \epsilon | x + \epsilon \geq c] \\
 &= \frac{d}{dx} (x + E[\epsilon | \epsilon \geq (c - x)]) \\
 &= 1 + \frac{d}{dx} \bar{\mu}(c - x).
 \end{aligned} \tag{1}$$

As the following two examples show, these assumptions do not guarantee that $d\bar{y}/dx$ is positive.

Example 1: The Exponential Distribution. Suppose that ϵ is distributed exponentially with mean μ . Then $\bar{\mu}(c-x) = c-x+\mu$. (This follows from the memory-less property of the exponential distribution.) This implies $\frac{d\bar{y}}{dx} = 1 + \frac{d}{dx}(c-x+\mu) = 0$.

Example 2: The Pareto Distribution. Let $F(z) = 1 - z^{-a}$, where $z \geq 1$, (i.e., $\underline{\epsilon} = 1$) and $a > 1$. (The Pareto distribution is still defined for $a \in (0, 1]$; however it no longer has finite mean.) It is easily shown that $f(z) = az^{-(a+1)}$ and $E[\epsilon] = \frac{a}{a-1}$. It is also easily shown that $\bar{\mu}(c-x) = (c-x) \cdot \frac{a}{a-1}$. Therefore,

$$\frac{d\bar{y}}{dx} = 1 - \frac{a}{a-1} < 0.$$

3. Preliminary Results

Note that we can rewrite $\bar{\mu}(c-x)$ as

$$\bar{\mu}(c-x) = \frac{\int_{c-x}^{\infty} zf(z)dz}{1-F(c-x)} \tag{2}$$

The objective of this paper is to find conditions that make the derivative of the right-hand side of the equation less than or equal to one, as c grows large.

As Example 2 shows, \bar{y} is negatively correlated with x when ϵ follows a Pareto distribution. A key property of the Pareto distribution is that its hazard rate is decreasing. Informally, this means that it has a fatter right tail than an exponential distribution (which has a constant hazard rate).

In a very different context (a theoretical analysis of auctions), Bulow and Klemperer (2002) show that \bar{y} will be negatively correlated with x , as long as the distribution of ϵ has a decreasing hazard rate. Proposition 1 lists this result, and Figure 1 illustrates it. (All proofs are listed in the Appendix.)

Proposition 1 (Bulow and Klemperer): Suppose $h(z)$ is strictly decreasing for all $z \geq z_0$. Then $\frac{d\bar{y}}{dx} < 0 \forall c \geq z_0 + x$.

Example 3: The lognormal distribution. As a further illustration of the Bulow-Klemperer result, suppose that ϵ is distributed according to the log-normal distribution. Sweet (1990)

has shown (see also Johnson, Kotz, and Balakrishnan 1994, 219-20) that the hazard rate of ϵ is maximized at a certain point, t_M .³ Thus, for all $z > t_M$, the hazard rate $h(z)$ is decreasing. This means that the hypothesis of Proposition 1 is satisfied when $c - x > t_M$. Hence, $d\bar{y}/dx$ is negative as long as $c - x > t_M$.

Of course, for many distributions the hazard rate is not decreasing; thus Proposition 1 does not apply to those cases. In fact, some distributions that are most commonly used in empirical applications—e.g., the normal, logistic, and Weibull—have increasing hazard rates. However, these distributions have upper tails that share a key property with the exponential distribution, which has a *constant* hazard rate. This property, which I call α , describes whether the derivative of the inverse hazard approaches zero in the limit.

Definition: A distribution satisfies *condition α* if

$$\lim_{z \rightarrow \infty} \frac{d}{dz} \frac{1 - F(z)}{f(z)} = 0.$$

Since the exponential distribution has a constant hazard rate, the derivative of its inverse hazard is zero. This is true *for all values of z* . Condition α is weaker. It only requires that this property hold as z approaches infinity. Recall from Example 1 that when the error term follows an exponential distribution, $\frac{d\bar{y}}{dx} = 0$, *for all values of c* . As the next proposition shows, when the distribution of error term satisfies *condition α* , $\frac{d\bar{y}}{dx}$ approaches zero as c becomes very large.

Proposition 2: Suppose condition α holds. Then $\lim_{c \rightarrow \infty} \frac{d\bar{y}}{dx} = 0$.

4. Four More Examples

To illustrate Proposition 2, I show that condition α is satisfied by the logistic, normal, Weibull, and extreme-value distributions. Therefore $\frac{d\bar{y}}{dx}$ has a limit of zero for these distributions.

³ Namely, at $t_M = \exp(\zeta + z_M\sigma)$, where ζ is the mean of $\log(\epsilon)$; σ is the standard deviation of $\log(\epsilon)$; z_M is the solution to $z_M + \sigma = \varphi(z_M)/[1 - \Phi(z_M)]$; and $\varphi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal.

Example 4: The Logistic Distribution. For the standard logistic distribution $F(z) = \Lambda(z) \equiv (1 + e^{-z})^{-1}$, and $f(z) = \Lambda(z)[1 - \Lambda(z)]$. Hence,

$$\frac{d}{dz} \frac{1 - F(z)}{f(z)} = \frac{d}{dz} \frac{1}{\Lambda(z)} = \frac{d}{dz} (1 + e^{-z}) = -e^{-z},$$

which has a limit of zero. Hence, the logistic distribution satisfies α .

Example 5: The Normal Distribution. When ϵ is distributed normal, it is easily shown that $\varphi'(z) = -z\varphi(z)$. Thus,

$$\frac{d}{dz} \frac{1 - \Phi(z)}{\varphi(z)} = -1 + \frac{[1 - \Phi(z)]}{\varphi(z)/z}.$$

L'Hopital's rule shows that the limit of second term on the right-hand side is $\frac{-1}{-1-1/z^2}$. Since this has a limit of one, the limit of the above expression is zero.

Example 6: The Weibull Distribution. The cdf for the Weibull distribution is $F(z) = 1 - e^{-z^a/a}$, $a > 0$, $z > 0$. Therefore, $f(z) = z^{a-1}e^{-z^a/a}$, and

$$\frac{1 - F(z)}{f(z)} = \frac{1 - (1 - e^{-z^a/a})}{z^{a-1}e^{-z^a/a}} = z^{1-a}$$

Thus, the derivative of this expression is $(1 - a)z^{-a}$, which has a limit of zero.

Example 7: The Extreme-Value Distribution. For the extreme-value distribution, $F(z) = e^{-e^{-z}}$, and $f(z) = e^{-z-e^{-z}}$. Hence,

$$\frac{1 - F(z)}{f(z)} = e^z e^{e^{-z}} - e^z.$$

The derivative of this expression is

$$e^z e^{e^{-z}} - e^z e^{e^{-z}} e^{-z} - e^z = \frac{e^{e^{-z}} - e^{e^{-z}-z} - 1}{e^{-z}}.$$

L'Hopital's rule implies that the limit of this expression is the limit of

$$\frac{-e^{e^{-z}} e^{-z} - e^{e^{-z}-z}(-e^{-z} - 1)}{-e^{-z}} = e^{e^{-z}} + e^{e^{-z}}(-e^{-z} - 1),$$

which is zero.

5. Main Result

We can now state the main result of the paper. This is that for any distribution that satisfies assumptions A1-A6, the limit of $\frac{d\bar{y}}{dx}$ is zero or negative. That is, under some very general conditions, the correlation between x and y will become zero or negative as the sample truncation becomes extreme.

The strategy for proving the result is to consider the limit of the derivative of the inverse hazard. I first show that the limit cannot be negative (otherwise the *level* of the inverse hazard would become negative, which is impossible).⁴ If the limit is zero, then condition α is satisfied, and Proposition 2 applies. If the limit is positive, then the tail of the distribution has a decreasing hazard, and Proposition 1 applies.

Theorem: $\lim_{c \rightarrow \infty} d\bar{y}/dx \leq 0$.

6. Empirical Applications

A potential criticism is that—despite the title of this paper—the main result does not necessarily imply that the correlation between two variables will *switch* signs. Instead, the correlation might only drop to zero. Worse, to obtain a switched-sign result (as in Proposition 1), I assumed that the error term had a distribution with a decreasing hazard, such as the Pareto or log-normal. Many researchers consider these types of distributions unorthodox and unlikely to occur in practice.

However, despite the common assumptions of empirical scholars, it is unreasonable to think that in practice the error terms for *all* observations will follow the same distribution. That is, while many of the errors may follow a normal distribution, a few might follow a Pareto, log-normal, or some other distribution with a decreasing hazard rate. And if even a tiny fraction follow the latter type of distribution, then this will cause a sign switch.⁵

⁴ An earlier of this paper adopted a much more complex, and less transparent, method for proving the main result. I am grateful to Simon Board for suggesting that I consider the derivative of the inverse hazard and to use the fact that its limit cannot be negative.

⁵ To see this, note that for any two variables their correlation is a weighted average of: (i) the correlation produced from observations that have errors with a decreasing hazard rate, and (ii) the correlation produced from observations that have other types of errors. By Proposition 1, the correlation from (i) is negative, and by the Theorem, the correlation from (ii) is non-positive. The weighted average of the two is negative if at least *some* of the errors have a decreasing hazard rate.

Another potential criticism is that the Theorem, strictly speaking, only applies to cases where the sample is zero percent of the population. However, there are many interesting empirical cases where the sample is *almost* zero percent of the population, and therefore it is reasonable to expect the theory of this paper to make predictions in those cases.

For instance, consider the championship heat of an event at the U.S. track and field championships. Technically, all U.S. citizens are eligible to compete in the event, but only eight typically qualify for the championship heat. Hence, the ratio of the sample to the population is approximately eight to 300 million.

In the remainder of the paper I demonstrate that the empirical implications of the main theoretical result seem to be widespread. To do this, I examine three very disparate areas of predictions—one area is normally analyzed by political scientists, another by biologists or exercise physiologists, and another by economists or education researchers. The predictions are sometimes surprising. Further, I provide some pieces of evidence—usually anecdotes or analyses by previous authors—that suggest that the predictions are accurate as well.

6.1. Race and Sex Advantages in Sports. In July, 2004, the online edition of *Track and Field News* discussed how few elite sprinters are white. In particular, it noted that between 1968 and 2000 only one white athlete qualified for the finals of an Olympic Trials 400 meters.⁶ This suggests that, in the population, white sprinters are disadvantaged relative to black sprinters—whether due to genetic factors, less desire or other cultural factors, or racism. However, shortly after the article appeared, two white runners, Jeremy Wariner and Andrew Rock, qualified for the 2004 Olympic Trials finals. A naive theory—that is, one not informed by the results of this paper—might predict that Wariner and Rock would run slower than the average black runner in the race. However, consistent with the theory of this paper, the average place of Wariner and Rock was faster than the average place of the black runners—Rock finished sixth and Wariner finished first. One month later, Wariner competed as the only white runner in the Olympic 400m final. He won that race, and one year later, he won the world-championship meet of 2005.

Cases such as these—where one race in a sport seems to have an advantage over other races, yet a member of a disadvantaged race is the star of near star of the sport—are not

⁶ The link to the article no longer exists. But a copy can be found at <http://www.polisci.ucla.edu/faculty/groseclose/pdfs/t.f.news.pdf>.

isolated. Perhaps the best-known counterpart to Wariner is Tiger Woods, a black athlete who dominates a typically white sport. Another is Lawrence Johnson, the only black pole vaulter to qualify for an Olympic final, who won the silver medal at the 1996 Games. Yet another example involves professional basketball, where the upper echelon is comprised disproportionately of black players, yet a white player, Steve Nash, has won the NBA's Most Valuable Player award for the last two seasons.

Another example involves auto racing. During her rookie season in 2005, Danica Patrick competed as the only woman in the Indianapolis 500. Of the 33 drivers who started the race, she finished fourth, and she was first among the rookie drivers. In the next year she finished eighth. Only one driver, Dan Weldon, who placed first in 2005 and fourth in 2006, had a higher average finish over the two years.

6.2. Wealthy Politicians and Electoral Success. Few would doubt that the personal wealth of politicians helps their electoral success. For example, the editorial page of the *New York Times* declared that Congress should reform a campaign finance system that “gives undue advantage to rich candidates ... whose spending subverts the democratic process.”⁷ Similarly, the *Los Angeles Times* opined that many good candidates are intimidated from running for office “because they can’t hope to match what rich candidates are able to spend.”⁸

However, suppose a researcher, analyzing the effect of wealth on electoral success, confined his or her analysis to a sample that only included incumbents. Such a sample would be comprised only of the very best campaigners (since incumbents have already demonstrated that they can win). As a consequence, the above theoretical results predict that such a researcher should find a nil effect of wealth or possibly a negative effect.

In an earlier paper Jeffrey Milyo and I (1999) analyzed this very problem with a sample of incumbents from the U.S. House of Representatives.⁹ We ran several econometric specifications, and none showed a significant positive effect of wealth. Meanwhile, one showed a weakly-significant *negative* effect. That specification examined the probability

⁷ “The Power of Rich Candidates,” *New York Times*, February 29, 1996, p. A20.

⁸ “The High Cost of Political Ambition,” *Los Angeles Times*, April 10, 1996, p. B8.

⁹ When we wrote the paper, we were aware of the attenuation bias that sample selection can cause. However, while U.S. law requires incumbent politicians to declare their financial assets, there is no similar requirement for challengers (nor, of course, for potential challengers). Consequently, wealth data was only available for incumbents.

that a “quality” challenger—one with previous electoral experience—would run against an incumbent. We found that millionaire incumbents experienced a 28% chance of facing a quality challenger, while the average incumbent only faced a 19% chance. This finding was statistically significant at the $p = .10$ level.

In that paper, we naively suggested that—although sample selection might mitigate the apparent effects of wealth—it should not reverse the effects.¹⁰ Of course, however, the theoretical results of this paper contradict that view. Further, if one is willing to assume that—in the population of all potential candidates—wealth really does have a positive effect on electoral success, then the Milyo-Groseclose results strongly support the predictions of this paper.

6.3. SAT scores and Income. Few would doubt that people with high SAT scores tend to earn higher incomes than those with low scores. William Bowen, the former president of Princeton University, and Derek Bok, the former president of Harvard University, examined this question in their famous book, *The Shape of the River* (1998). However, rather than examining a sample that draws from the entire population, they confined their analysis to a sample of matriculants from thirty of the most prestigious universities in the U.S.¹¹

Consistent with the predictions of this paper, Bowen and Bok found a *negative* relationship between SAT scores and income. The authors divided their sample into five categories, based on five ranges for the sum of a student’s verbal and math SAT score. These ranges were: below 1000, 1000-1099, 1100-1199, 1200-1299, and above 1299. Bowen and Bok regressed income on SAT scores and a number of control variables. The last column of their Table D.5.2 reported the results when the specification included all control variables, and the sample only included males. The specification predicted that the highest SAT group, i.e. those with scores of 1300 or above, would earn a yearly income of \$5,855 *less* than the lowest SAT group. (The specification respectively predicted that the second, third, and fourth highest SAT groups would earn \$1,132, \$421, and \$4560 less than the lowest SAT

¹⁰ In fact, that paper was the inspiration for this paper—to see if non-normal error terms could explain that empirical result.

¹¹ The schools in their sample ranged from the extremely selective (such as Princeton, Yale, and Stanford, which admit approximately 10% of all applicants) to the very selective (such as University of Michigan, Penn State, Miami [Ohio] University, and University of North Carolina at Chapel Hill, which admit approximately half of all applicants).

group.)¹²

Of course, rather than suggesting that there truly is a negative relationship between SAT scores and income, the theory of this paper suggests that the Bowen-Bok results are spurious. I predict that, if instead, Bowen and Bok had performed their analysis upon the entire population, then such an analysis would show a positive relationship between income and SAT scores.

One caveat to the above argument is that, although Bowen and Bok truncated their sample, the truncation was not very extreme. For instance, even the most selective schools in their sample admitted approximately ten percent of their applicants. This is not nearly as selective as the example with U.S. House members, where the sample included only the top one-in-one-hundred-thousand or so campaigners in the population, or the example with Olympic Trials 400m runners, where the sample included only the top one-in-ten-million or so runners in the population.

However, note that Proposition 1, the Bulow-Klemperer result, does not require the sample truncation to be extreme. It only requires that at some point in the distribution of the error term, the hazard rate decreases.

Next, note that Bowen and Bok's dependent variable is income. Some compelling evidence suggests that the distribution of income has a decreasing hazard rate. For instance, Lydall (1968) has studied income distribution over an extensive number of countries.¹³ He concludes that for nearly all the countries he studies, a lognormal distribution fits the data best when incomes are in the middle of the distribution—approximately between the 10th and 80th percentiles. However, for the very upper incomes, he suggests that a Pareto distribution fits the data better. Recall that the lognormal distribution has a decreasing hazard rate in its upper tail, and the Pareto distribution has a decreasing hazard rate for the

¹² The last column of Table D.5.3 reports the results of a similar specification for the income of females. The results predict that the highest SAT group (respectively, the second-, third-, and fourth- highest SAT groups) will earn \$701 (respectively, \$268, \$2,137, and \$-565) *less* than the lowest SAT group. In Table 5.1 Bok and Bowen provide an analysis that does not include control variables. Here again, these results show a negative relationship between SAT scores and income, or at best no relationship. For instance, among white male matriculants to the most selective schools (which Bok and Bowen label “SEL-1” schools), those with the lowest SAT scores tended to earn the highest incomes. (However, those with the highest SAT scores had the second-highest mean incomes.) Similar results occur with white female matriculants to the most selective schools.

¹³ See also Atkinson's (1983) review of these results.

entire distribution. Other evidence that the distribution of income has a decreasing hazard comes from Vilfred Pareto himself. Based on data from Prussia, Saxony, Peru, and the cities of Florence, Perugia, Basle, and Augsburg, Pareto concluded that a “law” described income distributions.¹⁴ The Pareto distribution is named after this law.¹⁵

That incomes are distributed according to such distributions means that we should expect switched-sign results even in cases where the sample is only moderately truncated. The Bowen-Bok result appears to be such a case.

6.4. Other Predictions. Many additional predictions follow from the main theoretical result. For example: (i) Although a college degree usually helps a manager rise on the corporate ladder, if a researcher gathers a sample of only the very top executives, then this paper predicts that if some of those executives do not have a college degree, then they likely will be more successful than the other executives. (ii) Before its annual draft, the National Football League invites top college players to perform a series of physical tests, such as the 40-yard dash, bench press, etc. This paper predicts that those who perform relatively poorly at the test will tend to be drafted earlier than those who perform relatively well. (iii) Top economics departments tend to hire assistant professors who have strong mathematical training. However, suppose that a hire at a top department has poor training. This paper predicts that he or she is more likely than his or her peers to earn tenure. (iv) Suppose a woman tends to choose boyfriends who are tall (or well-mannered or good-dressers, etc.). But suppose her current boyfriend is short (or ill-mannered or a bad-dresser, etc.). This paper predicts that she is more likely to marry him than her typical boyfriend.

7. Conclusion

This paper establishes a theoretical result, which at first glance might seem unintuitive: Two variables can be positively correlated in a population; however in an extreme sample,

¹⁴ See Atkinson (1983, p. 102).

¹⁵ Another caveat is that Proposition 1 assumes that the distribution of ϵ (not necessarily y) has a decreasing hazard. That is, to obtain the high-SAT-low-income prediction it must be true that, while holding all independent variables constant, income follows a distribution which has a decreasing hazard. Of course, Lydall and Pareto do not necessarily establish this for the independent variables that Bowen and Bok consider. However, they do establish that, within each country that they examine, incomes have a decreasing hazard. That is, the Lydall and Pareto analyses control for the country of residence, which we can interpret as an independent variable. Thus, Lydall and Pareto’s results indeed establish that ϵ , not just y , follows a distribution with a decreasing hazard.

they will be negatively correlated or have zero correlation. I apply this result to three disparate empirical areas: (i) race and sex advantages in sports, (ii) the effect of personal wealth on elections, and (iii) the relation between SAT scores and income. I am confident that these are only a small sample of the total areas to which the theory can be applied.

8. Appendix

Below, in the proof of Proposition 1, I follow the proof of Bulow and Klemperer (2002, p. 16, Lemma 1), only I substitute my own notation. Plus, I make one alteration to their proof. This is to prove a step that they skip. Bulow and Klemperer treat $\lim_{z \rightarrow \infty} z[1 - F(z)]$ as zero, but they do not prove this nor state assumptions that imply it. (Further, it is not necessarily true. E.g., let $F(z) = 1 - z^{-a}$, where $a \in (0, 1]$.) To prove the skipped step I prove the following lemma. It shows that a sufficient condition for the claim is that the conditional mean of ϵ is finite.

Lemma 1: $\lim_{z \rightarrow \infty} z[1 - F(z)] = 0$.

Proof: First note that $\forall z \geq 0$,

$$z[1 - F(z)] = z \int_z^\infty f(\epsilon) d\epsilon = \int_z^\infty z f(\epsilon) d\epsilon < \int_z^\infty \epsilon f(\epsilon) d\epsilon \leq \int_0^\infty \epsilon f(\epsilon) d\epsilon = \bar{\mu}(0)[1 - F(0)].$$

(A4) implies that the limit of $z[1 - F(z)]$ exists, and the above steps imply that is less than $\bar{\mu}(0)[1 - F(0)]$, which implies that it is finite. To show that it is zero, suppose to the contrary that instead it equals some c in $(0, \infty)$. Thus, it follows that

$$\lim_{z \rightarrow \infty} \frac{1 - F(z)}{1/z} = c,$$

and L'Hopital's rule implies

$$\lim_{z \rightarrow \infty} \frac{f(z)}{1/z^2} = c,$$

This implies that there exists a $z_0 > 0$ and $\delta \in (0, c)$ such that for all $z \geq z_0$, $z^2 f(z) > \delta$. This implies that $\forall z \geq z_0$, $z f(z) > \delta/z$, which implies

$$\bar{\mu}(z_0) = \frac{1}{[1 - F(z_0)]} \int_{z_0}^\infty z f(z) dz > \frac{1}{[1 - F(z_0)]} \int_{z_0}^\infty \frac{\delta}{z} dz = \infty.$$

This, in turn, contradicts (A6). It follows that $\lim_{z \rightarrow \infty} z[1 - F(z)] = 0$. ■

Proposition 1 (Bulow and Klemperer): Suppose the hazard rate of ϵ , $h(\cdot)$, is strictly decreasing for all $z \geq z_0$. Then $\frac{d\bar{y}}{dx} < 0 \forall c \geq z_0 + x$.

Proof: We want to show that the derivative of $\bar{\mu}(c - x)$ with respect to x is less than -1 . (See (1).) To do this, we first show that we can simplify $\int_{c-x}^\infty z f(z) dz$, the numerator of (2). Once we integrate this by parts (letting $u = z$ and $v = -[1 - F(z)]$), we get

$$\int_{c-x}^\infty z f(z) dz = \int_{c-x}^\infty [1 - F(z)] dz + (c - x)[1 - F(c - x)] - \lim_{z \rightarrow \infty} z[1 - F(z)].$$

(Bulow and Klemperer omit the term $(\lim_{z \rightarrow \infty} z[1 - F(z)])$.) First, by Lemma 1, we can eliminate $\lim_{z \rightarrow \infty} z[1 - F(z)]$ in the above expression. Next, note that the derivative of $\ln[1 - F(z)]$ is $-f(z)/[1 - F(z)] = -h(z)$. Hence, $1 - F(z) = \exp[-\int_{\underline{\epsilon}}^z h(\epsilon)d\epsilon]$, and the above equation can be rewritten as

$$\int_{c-x}^{\infty} e^{-\int_{\underline{\epsilon}}^z h(\epsilon)d\epsilon} dz + (c-x)[1 - F(c-x)].$$

Note that for any function $H(z)$, $\int_{c-x}^{\infty} H(z)dz = \int_{\underline{\epsilon}}^{\infty} H(z+c-x-\underline{\epsilon})dz$. Hence, the above equation can be rewritten as

$$\int_{\underline{\epsilon}}^{\infty} e^{-\int_{\underline{\epsilon}}^{z+c-x-\underline{\epsilon}} h(\epsilon)d\epsilon} dz + (c-x)[1 - F(c-x)].$$

Note that $\int_{\underline{\epsilon}}^{z+c-x-\underline{\epsilon}} h(\epsilon)d\epsilon = \int_{\underline{\epsilon}}^{c-x} h(\epsilon)d\epsilon + \int_{c-x}^{z+c-x-\underline{\epsilon}} h(\epsilon)d\epsilon$, which equals $-\ln[1 - F(c-x)] + \int_{c-x}^{z+c-x-\underline{\epsilon}} h(\epsilon)d\epsilon$, which equals $-\ln[1 - F(c-x)] + \int_{\underline{\epsilon}}^z h(\epsilon+c-x-\underline{\epsilon})d\epsilon$. Hence, the above equation equals

$$[1 - F(c-x)] \int_{\underline{\epsilon}}^{\infty} e^{-\int_{\underline{\epsilon}}^z h(\epsilon+c-x-\underline{\epsilon})d\epsilon} dz + (c-x)[1 - F(c-x)].$$

Recall that this equals $\int_{c-x}^{\infty} zf(z)$. Substituting this into (2) and simplifying gives

$$\bar{\mu}(c-x) = (c-x) + \left(\int_{\underline{\epsilon}}^{\infty} e^{-\int_{\underline{\epsilon}}^z h(\epsilon+c-x-\underline{\epsilon})d\epsilon} dz \right).$$

Consider the second term in parentheses. Note that the argument of $h()$ is always greater than or equal to $c-x$. Thus, according to the hypothesis of the proposition, for all $\epsilon \in [\underline{\epsilon}, \infty]$, $h()$ is decreasing in its argument, which means that $h(\epsilon+c-x-\underline{\epsilon})$ is increasing in x , which means that the entire second term is decreasing in x . Finally, since the derivative of the first term is -1 ,

$$\frac{d}{dx} \bar{\mu}(c-x) < -1.$$

Substituting this into (1) implies $\frac{d\bar{y}}{dx} < 0$. ■

Before proving Proposition 2, it is convenient first to prove the following four lemmas.

Lemma 2: $\lim_{z \rightarrow \infty} [\bar{\mu}(z) - z][1 - F(z)] = 0$.

Proof: Using the definition of $\bar{\mu}(z)$, we can write

$$[\bar{\mu}(z) - z][1 - F(z)] = \int_z^{\infty} \epsilon f(\epsilon)d\epsilon - z[1 - F(z)].$$

It is obvious that the first term has a limit of zero, and by Lemma 1, the second term also has a limit of zero. Hence, $\lim_{z \rightarrow \infty} [\bar{\mu}(z) - z][1 - F(z)] = 0$. ■

Lemma 3: $\frac{d}{dz} \underline{\mu}(z) = \frac{f(z)}{F(z)} [z - \underline{\mu}(z)]$.

Proof: Note that $\underline{\mu}(z) = \frac{1}{F(z)} \int_{\underline{\epsilon}}^z \epsilon f(\epsilon) d\epsilon$. Hence,

$$\begin{aligned} \frac{d\underline{\mu}}{dz} &= \frac{-f(z)}{[F(z)]^2} \int_{\underline{\epsilon}}^z \epsilon f(\epsilon) d\epsilon + \frac{1}{F(z)} z f(z) \\ &= \frac{-f(z)}{F(z)} \underline{\mu}(z) + \frac{f(z)}{F(z)} z. \end{aligned}$$

■

Lemma 4: If condition α holds, then $\lim_{z \rightarrow \infty} [1 - F(z)]^2 / f(z) = 0$.

Proof: First, note that, by α , there exists $z_0 > 0$ and $\delta > 0$ such that for all $z \geq z_0$,

$$\frac{d}{dz} \frac{1 - F(z)}{f(z)} \leq \delta.$$

Hence, for all $z \geq z_0$

$$\frac{1 - F(z)}{f(z)} \leq \frac{1 - F(z_0)}{f(z_0)} + \delta(z - z_0).$$

By multiplying both sides of the above by $1 - F(z)$, it follows that that for all $z \geq z_0$, $[1 - F(z)]^2 / f(z)$ is less than or equal to

$$[1 - F(z)] \left[\frac{1 - F(z_0)}{f(z_0)} - \delta z_0 \right] + \delta z [1 - F(z)].$$

The first term of the above expression has a limit of zero, since $1 - F(z)$ has a limit of zero. The second term has a limit of zero, by Lemma 1. It follows that $\lim_{z \rightarrow \infty} [1 - F(z)]^2 / f(z) = 0$. ■

Lemma 5: If condition α holds, then $\lim_{z \rightarrow \infty} [\bar{\mu}(z) - z]f(z) / [1 - F(z)] = 1$.

Proof: After multiplying numerator and denominator by $[1 - F(z)] / f(z)$, it follows that $[\bar{\mu}(z) - z]f(z) / [1 - F(z)]$ equals

$$\frac{[\bar{\mu}(z) - z][1 - F(z)]}{[1 - F(z)]^2 / f(z)}. \quad (3)$$

By the definitions of $\mu(z)$ and $\bar{\mu}(z)$, $\mu = F(z)\mu(z) + [1 - F(z)]\bar{\mu}(z)$. Hence, $\bar{\mu}(z) = \mu/[1 - F(z)] - \underline{\mu}(z)F(z)/[1 - F(z)]$. Substituting this into (3) gives

$$\frac{\mu - \underline{\mu}(z)F(z) - z[1 - F(z)]}{[1 - F(z)]^2/f(z)}. \quad (4)$$

By Lemmas 2 and 4, the numerator and denominator of (3) have limits of zero. Hence, the same is true of (4), and therefore we can use L'Hopital's rule to derive the limit of (4). The derivative of the numerator of (4) is

$$-\frac{d\mu}{dz}F(z) - \underline{\mu}(z)f(z) - [1 - F(z)] + zf(z),$$

which, by Lemma 3, equals

$$-f(z)[z - \underline{\mu}(z)] - \underline{\mu}(z)f(z) - [1 - F(z)] + zf(z),$$

which equals $-[1 - F(z)]$. The derivative of the denominator of (4) is

$$\frac{2[1 - F(z)]}{f(z)}(-f(z)) + [1 - F(z)]^2 \left(\frac{-1}{[f(z)]^2} \right) f'(z).$$

Hence, it follows that

$$\begin{aligned} \lim_{z \rightarrow \infty} \frac{\bar{\mu}(z) - z}{[1 - F(z)]/f(z)} &= \lim_{z \rightarrow \infty} \frac{-[1 - F(z)]}{-2[1 - F(z)] - \left[\frac{[1 - F(z)]^2}{[f(z)]^2} f'(z) \right]} \\ &= \lim_{z \rightarrow \infty} \frac{-1}{-2 - \left[\frac{1 - F(z)}{[f(z)]^2} f'(z) \right]} \end{aligned} \quad (5)$$

By α ,

$$\lim_{z \rightarrow \infty} \frac{d}{dz} \frac{1 - F(z)}{f(z)} = \lim_{z \rightarrow \infty} \frac{-f(z)}{f(z)} - \frac{1 - F(z)}{[f(z)]^2} f'(z) = 0.$$

This implies that the limit of the expression in large brackets in (5) has a limit of -1. This, in turn, implies that $\lim_{z \rightarrow \infty} [\bar{\mu}(z) - z]f(z)/[1 - F(z)] = 1$. ■

Proposition 2: If condition α holds, then $\lim_{c \rightarrow \infty} \frac{d\bar{y}}{dx} = 0$.

Proof: When we differentiate (2) by x we obtain

$$\begin{aligned} \frac{d\bar{\mu}(c-x)}{dx} &= \frac{[1 - F(c-x)](c-x)f(c-x) - f(c-x) \int_{c-x}^{\infty} \epsilon f(\epsilon) d\epsilon}{[1 - F(c-x)]^2} \\ &= \frac{f(c-x)}{1 - F(c-x)} [(c-x) - \bar{\mu}(c-x)] \end{aligned}$$

The above equation and Lemma 5 imply

$$\lim_{c \rightarrow \infty} \frac{d}{dx} \bar{\mu}(c - x) = -1.$$

Substituting this into (1) implies $\lim_{c \rightarrow \infty} \frac{d\bar{y}}{dx} = 0$. ■

Theorem: $\lim_{c \rightarrow \infty} d\bar{y}/dx \leq 0$.

Proof: Consider the derivative of the inverse hazard rate of ϵ ; i.e., $\frac{d}{dz}[1-F(z)]/f(z)$. First, note that its limit cannot be negative. If it were, then the inverse hazard rate itself (i.e. not just its derivative) would become negative at some point, which is impossible. Now suppose that the derivative of the inverse hazard rate has a limit of zero. If this is true, then the distribution of ϵ satisfies condition α , and by Proposition 2, $\lim_{c \rightarrow \infty} d\bar{y}/dx = 0$. Finally, suppose that the derivative of the inverse hazard has a positive limit. If this is true, then the hazard rate is decreasing in its tail, and Proposition 1 applies. Hence, $\lim_{c \rightarrow \infty} d\bar{y}/dx \leq 0$. ■

References

- Atkinson, A. B. 1983. *The Economics of Inequality*. Oxford: Clarendon Press.
- Bowen, William G., and Derek Bok. 1998. *The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions*. Princeton, N.J.: Princeton University Press.
- Bulow, Jeremy and Paul Klemperer. 2002. "Prices and The Winner's Curse", *RAND Journal of Economics*, 33, No. 1, (Spring 2002): pp. 1-21.
- Heckman, James. 1979. "Sample Selection Bias as a Specification Error," *Econometrica* 47: 153-161.
- Johnson, Norman L., Samuel Kotz, and N. Balakrishnan. 1994. *Continuous Univariate Distributions, Volume 1, Second Edition*. New York: John Wiley & Sons, Inc.
- Lydall, H. F. 1976. *The Structure of Earnings*. Oxford: Oxford University Press.
- Milyo, Jeffrey, and Tim Groseclose. 1999. "The Electoral Effects of Incumbent Wealth," *Journal of Law and Economics*, 42 (October): 699-722.
- Sweet, A. L. 1990. "On the Hazard Rate of the Lognormal Distribution," *IEEE Transactions on Reliability*, 39, 325-8.