IQ and Poverty: Testing the Bell Curve on a Novel Data Set

IQ and Poverty: Testing the Bell Curve on a Novel Data Set

*Dody Eid**

April 30, 2018

Abstract

In their controversial work *The Bell Curve*, Charles Murray and Richard Herrnstein hypothesize that intelligence plays a significant role in determining later life outcomes. Using data from the 1979 National Longitudinal Survey of Youth, Herrnstein and Murray specifically claim that Youth IQ is more important than Parental Socioeconomic Status in affecting future odds of living in poverty. We explore their hypothesis using the same statistical decisions but on a new data set: the 1997 National Longitudinal Survey of Youth. Our findings mirror those of Herrnstein and Murray; IQ has a larger effect than socioeconomic status on the likelihood of being in poverty, though the overall effect of both is diminished. We find that increasing Youth IQ by one standard deviation is expected to decrease the odds of ending up in poverty by 35 percentage points, compared to the 13 percentage point decrease from similarly increasing Parental Socioeconomic Status.

Introduction

What determines the likelihood of being in poverty? The factors that influence poverty are heavily debated (Corcoran 1995). Economic upbringing, education, geography, and generational income are examples of indicators one might analyze in predicting poverty status. Some literature suggests that a child's economic environment plays a heavy role in future outcomes, such as completed schooling (Greg J. Duncan and Smith 1998). In their seminal work *The Bell Curve*, Charles Murray and Richard Herrnstein (henceforth HM) examine a more controversial variable to explain poverty: intelligence (Herrnstein and Murray 1994).

*A special thank you to David Kane for all his support.

The Bell Curve is centered on one message: intelligence matters. HM go to great lengths to explain the existence of general intelligence and its influence on numerous outcomes in American society. They have been the subject of heavy criticism; critics scrutinize everything from their characterization of general intelligence to their statistical methods to their overall purpose (Bernie Devlin and Roeder 1994). While the more controversial sections of *The Bell Curve* venture into the relationship between race and intelligence, this paper is primarily interested in how intelligence affects the poverty status of an individual, and how that effect compares to the effect of childhood socioeconomic status. By running a logistic regression with Age, Youth IQ, and Youth SES as RHS variables and poverty status as the LHS variable, HM find the effect of IQ on likelihood of being in poverty to larger than that of SES.

The empirical basis for HM's claim rests on the 1979 National Longitudinal Survey of Youth (NLSY '79) (Will Beasley and Meredith 2018). This data set includes information on 12,685 individuals from childhood to adulthood. The NLSY '79 has infromation on education, parental characteristics, employment status over time, cognitive test scores, and income, among other things. HM restrict their analysis to white indivduals, and eliminate observations with missing values for their variables of interest. To measure socioeconomic status, they create an index which captures the education of the parents, household income, and occupational prestige. Their measurement of IQ stems from a cognitive test known as the Armed Forces Qualification Test (AFQT). Finally, their measurement of poverty is a simple dummy variable that indicates whether an individual's income was below the poverty line in 1989. HM conclude that increasing an individual's intelligence by one standard deviation has a more significant impact on decreasing poverty than increasing an individual's socioeconomic status by one standard deviation.

One strong criticism of HM's methodology is centered on their measurement of family SES; Korenman and Winship explain its incompleteness and ultimately find SES to be at least as important and in some cases more important than IQ (1995). Moreover, one might wonder if HM would see different results had they made other legitimate econometric decisions. Given these concerns, this paper tests HM's hypothesis on a different dataset: the 1997 National Longitudinal Survey of Youth. The NLSY '97 includes information on 8,984 individuals,

including most of the same variables from 1979 data set but with a few excpetions and altercations. The AFQT was given by the NLSY not as raw scores but rather as percentiles. For example, we would not be able to see Joe's raw score but we would be able to see that he scored in the 73rd percentile. Secondly, poverty was not given as a dummy variable. Instead, a poverty ratio is given, which is defined as the ratio between an individual's income and the poverty line. Lastly, occupational prestige and their corresponding Duncan values were not available, unlike the NLSY '97.

Without making any meaningful changes to HM's methodology, we reaffirm the hypothesis that IQ is more important than family SES in avoiding poverty, though both of these co-variates' effects are smaller than those found by HM. Running a logistic regression with IQ, SES, and Age in 1997 as independent variables and poverty status in 2007 as the dependent variable, we find the IQ effect to be approximately three times the size of the SES effect.

Analysis

Motivation

A typical concern surrounding econometric findings discussed by Gelman and Loken is that analytic decisions may be data dependent when they ideally ought to remain data indepenent (2013). Pre-registration is a way to avoid data dependency. It requires that you make analytic decisions before the collection of data and make little to no changes to those decisions as you apply them to the data. By using nearly the same methodology as HM on a new data set, we test their hypotheses without data dependent analytic decisions, essentially mimicking pre-registration.

We first summarize HM's theoretical model. HM are interested in the following logistic regression.

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 IQ_{sd} + \beta_2 SES_{sd} + \beta_3 Age + \mu$$

Where P is the probability of being in poverty, IQ_{sd} is the person's childhood IQ measured

in standard deviations, and SES_{sd} is parental socioeconomic status measured in standard deviations. HM conclude that:

$$\beta_1 < \beta_2 < 0$$

That is, IQ has a larger impact than parental SES on lowering the probability of being in poverty. The statistical path to this conclusion is laid out in the next section. All data analysis is done in R (R Core Team 2013).

Measurements

The Sample

The Bell Curve uses data from the 1979 National Longitudinal Survey of Youth. The NLSY '79 has data on 12,686 individuals, including information on household income, cognitive test scores, and poverty status. HM are interested in two relationships: the relationship between an individual's IQ score in 1979 and their poverty status in 1989, as well as the relationship between an indiviual's socioeconomic environment in 1979 and their poverty status in 1989. For the NLSY '79, most individuals are in their early to late teens in 1979. HM filter out non-whites and those who are out of the labor force due to schooling in 1989.

We supplement HM's analysis with data from the 1997 National Longitudinal Survey of Youth (NLSY '97). There are three important variables that need replicating from HM's 1979 analysis: Poverty, SES, and IQ. In the context of the NLSY '97, these variables are an individual's poverty status in 2007, an individual's socioeconomic environment in 1997, and an individual's IQ in 1997. We filter out data to only include whites. This leaves us with 5232 observations. We do not make adjustments based on employment status in 2007, but this sample restriction does not significantly change our results.

IQ

HM measure an individual's IQ using the AFQT (Armed Forces Qualification Test). This test is taken by youths in their late teens, and it is described by HM as "[doing] what a good IQ test is supposed to do-tap into a general factor rather than specific bits of learning or skill-as well as or better than its competitor." HM originally used AFQT scores from 1980,

but the test underwent a revision in 1989 and HM decided to run all their analyses with the revised scores. However, they did not restrict updated scores to only 18 to 23 year olds, but instead looked at all AFQT scores. The test is influenced by age, but HM believe the effect is close to enough to linear that including age as an independent variable in their regressions should correct for this. Moreover, the AFQT is skewed such that scores are bunched on the high end of the distribution. HM adjust for skew by examining an individual's centile in the distribution and assigning a standardized score. This adjustmnet ensures a distribution with tails on both sides of the mean, which HM believe are essential to capturing the value of having a higher IQ. However, HM note that changing from AFQT 1980 to 1989 and adjusting for skew do not change the results for most of their regressions.[1]

Our measurement of IQ from the NLSY 97 is very close to the measurement used by HM in 1979 but with a few caveats. The NLSY 97 has data on AFQT scores in 1997, but it is given as percentiles instead of raw figures. We give each individual a standardized AFQT score by subtracting the mean percentile and dividing by the standard deviation. While its distribution is roughly normal, there is a slight left skew, as HM found with their data. We make no adjustment to correct for skew, and we do not use sample weights. Like HM, we filter out observations who have NA for AFQT. This, along with our previous exclusion of nonwhites, leaves us with 3133 observations.

SES

An individual's SES index is determined by parents' income, years of schooling, and occupation. For the '79 data set, HM define family income as the average household income for 1978 and 1979, unless only one of those years is available, then they only use that year. They then log this value (a typical transformation for income data), and standardize it. Mother's and father's education is measured by highest grade completed and is also standardized. Lastly, HM measure occupational prestige using modified Duncan values.[2] An average of these standardized scores is the final SES value for the individual, and then SES itself is standardized by subtracting the mean and dividing by the standard deviation.

[1] For more discussion on the AFQT, see page 570-586 of *The Bell Curve*

[2] For more discussion of SES, see pages 573-575 of *The Bell Curve*

Our calculation of SES for the '97 data follows HM's methodology almost identically, with the exception of calculating occupational prestige. We have the following raw pieces of data from the NLSY '97: biological mom highest grade completed (HGC), biological dad HGC, residential mom HGC, residential dad HGC, household income in 1997, and household income in 1996. Our measurement of SES is an index calculated using standardized values for grades completed by the parents and household income, similar to the methods used by Herrnstein and Murray, with the exception that this SES index excludes Duncan values for occupational prestige. In *The Bell Curve*, household income was defined as an average between the years 1978 and 1979 (or one of those years if the other was not available). For the NLSY '97, we define household income as an average between the years 1996 and 1997 (or one of those two years if the other was not available). We then take a log transformation of household income (and add 1 to avoid problems with zeros in logs), and standardize. Parental education is calculated as the average of residential mom HGC and residential dad HGC (or one of those two values if the other is missing). The use of residential parents instead of biological parents has the purpose of isolating environmental rather than natural effects on a child. We then average the standardized income and standardized education value for each individual (or use one if the other is missing) to calculate SES. Finally, these SES values are standardized. On top of previous filters, we filter out those with NA for SES, leaving us with 3101 observations.

Poverty

HM measured poverty status as a binary variable that indicates whether an individual was below the poverty line 10 years after 1979 (1989). We employ the exact same definition, but for the '97 data, we are interested in measuring poverty status in 2007. The NLSY '97, unlike the NLSY '79, does not come with a dummy variable for poverty status. It instead has a "poverty ratio" for each individual, which is the ratio of an individual's income to the poverty line. This variable takes on values 0-1000 (any negative values we code to be NA). If one's value is between 0 and 99, one's ratio is less than 1, indicating income below the poverty line. We code such values as "1" and all higher values as "0" to replicate the dummy variable for poverty used by HM. Note that there are 3730 observations left after filtering

out NA values. Also, after coding the dummy variable, I find 481 people in poverty and 3249 people out of poverty. The percentage of people in poverty for the NLSY '97 is higher than that of the NLSY '79, perhaps as a result of significant financial trouble in 2007.

Results

We replicate Herrnstein and Murray's analysis on the NLSY '79 below and compare them with results from the NLSY '97. Table 1 shows percentages of people who end up in poverty for different cognitive and socioeconomic classes. Table 2 does the same but on different individuals from a different data set: the NLSY '97. Both tables are restricted to white individuals. Table 1 has a sample size of 6430, while Table 2 has a sample size of 3101.

Table 1 highlights HM's claims in *The Bell Curve*: that those in the lowest cognitive class have a higher percentage of people who end up in poverty than those in the lowest socioeconomic class. The effect also works the other way: the middle to high cognitive classes have less people who end up in poverty than the middle to high SES classes. Those in lowest cognitive class have 29.3% of individuals go into poverty, compared to the 19.7% figure for the lowest socioeconomic class. Table 2 shows similar effects but they are not as large. The difference between the lowest cognitive class and the lowest SES class is around 3 percentage points, compared to the 10 percentage point difference in Table 1. Overall, the benefits of high IQ compared to high SES, and the consequences of low IQ compared to low SES, are replicated in Table 1 and confirmed to a lesser degree in Table 2.

Table 1: Poverty by Cognitive and Socioeconomic Class - NLSY 1979

	'79 Parents' SES	'89 % in Poverty	'79 Cognitive Class	% in Poverty
1	Very High	2.9	Very Bright	2.1
2	High	2.8	Bright	3.4
3	Mid	7.4	Normal	6.3
4	Low	12.3	Dull	16.1
5	Very Low	19.7	Very Dull	29.3
6	Sample	7.0	Sample	7.0

This table indicates the percentage of people who end up in poverty for different cognitive and socioeconomic classes from the 1979 National Longitudinal Survey of Youth, used by Herrnstein and Murray. "'79 Parents' SES" indicates the socioeconomic classes in which a young adult in 1979 could have grown up. "'79 Cognitive Class" indicates the cognitive classes for the same group. The cognitive and socioeconomic classes are defined as such: "Very Low SES" is the bottom 20 Percent of families in the SES distribution, "Low SES" is the next 20 Percent of the distribution, and so on. The same percentile breaks are used for cognitive classes. The "Percent in Poverty" columns indicate what percentage of the people in the specific classes end up in poverty in 1989. This table replicates the findings of HM in The Bell Curve, who also use the 1979 National Longitudinal Survey of Youth. Namely, that high cognitive classes have a lower future poverty rate than high SES classes. The reverse effect is also displayed; low cognitive classes have a higher future poverty rate than low SES classes.

Table 2: Poverty by Cognitive and Socioeconomic Class - NLSY 1997

	'97 Parents' SES	'07 % in Poverty	'97 Cognitive Class	% in Poverty
1	Very High	11.3	Very Bright	10.6
2	High	9.4	Bright	6.9
3	Normal	8.5	Normal	9.5
4	Low	11.8	Dull	11.6
5	Very Low	19.8	Very Dull	22.1
6	Sample	12.2	Sample	12.2

This table indicates the percentage of people who end up in poverty for different cognitive and socioeconomic classes from the 1997 National Longitudinal Survey of Youth. "'97 Parents' SES" indicates the socioeconomic classes in which a young adult in 1997 could have grown up. "'97 Cognitive Class" indicates the cognitive classes for the same group. The cognitive and socioeconomic classes are defined as such: "Very Low SES" is the bottom 20 Percent of families in the SES distribution, "Low SES" is the next 20 Percent of the distribution, and so on. The same percentile breaks are used for cognitive classes. The "Percent in Poverty" columns indicate what percentage of the people in the specific classes end up in poverty in 2007. Overall, the table supports the idea that IQ is more important than SES in avoiding future poverty, though the effect is not as large as the one from the NLSY 1979.

In Figure 1, we divide people into 20 Youth IQ and Youth SES classes instead of five. An x-axis value of "5th Percentile" indicates the class of people between the 0th and 5th percentile. The next class is labled as "10th Percentile" and includes people between the 5th and 10th percentile. This continues until the 100th Percentile. Only the 5th, 25th, 50th, and 100th Percentile classes are marked for aesthetics. Each class has a corresponding percentage of people who end up in poverty ten years later. There are two data sets: the NLSY '79, represented in blue and purple, and the NLSY '97, represented in red and orange. Both are restricted to white individuals.

One can see the importance of IQ relative to SES for avoiding poverty. Low values in the IQ distribution generally come with larger percentages in poverty than low values in the SES distribution. The effect also works the other way: high values in the IQ distribution come with lower percentages in poverty than high values in the SES distribution. However, these effects are more pronounced for data from NLSY '79, as indicated by the larger gaps between the blue and purple curves.

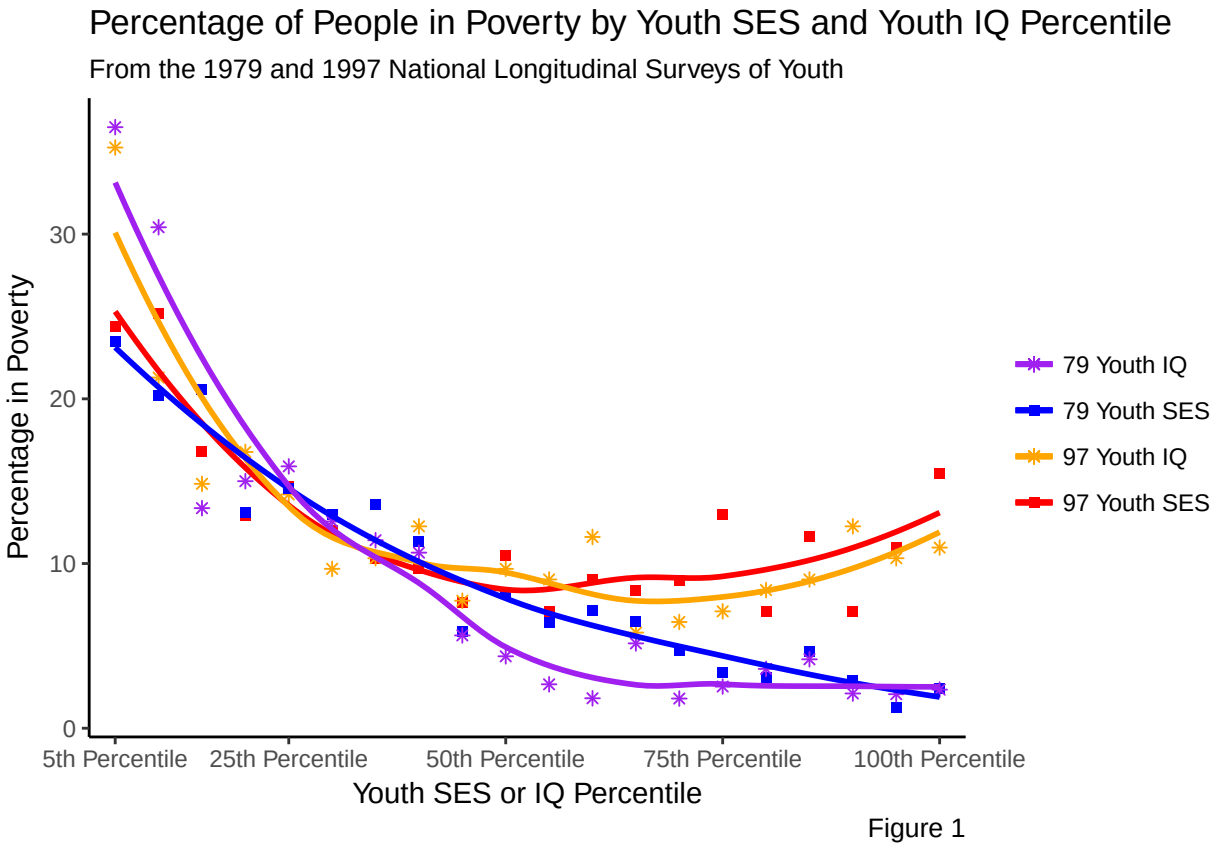

Figure 1: This graphic shows percentage of people who end up in poverty at different percentiles in the IQ or SES distribution. Youth IQ is measured using the Armed Forces Qualification Test (AFQT) for individuals in their teens. Youth SES is a composite of parental income and education. The x-axis is grouped into percentile chunks. The 5th percentile tick indicates people between the 0th and 5th percentile, the 25th percentile tick indicates people between the 20th and 25th percentile, and so on. Each of these groupings has a certain percentage of people who end up in poverty ten years after their youth. These percentages are indicated by the points. Note that having extremely low IQ is more detrimental than having extremely low SES and having high IQ is more beneficial than having high SES for both the 1979 and 1997 NLSY, though these discrepancies are more pronounced for the 1979 data set.

Given the uncertainty of each estimate, one may find Figure 1 to be uninformative. We supplement our findings with two logistic regressions. Table 3 is a replication of HM's analysis on the NLSY '79, and Table 4 is our regression for the NLSY '97.

Table 3: Relationship Between IQ/Parental Socioeconomic Status and Poverty - 1979

	Dependent variable:
	Poverty in 1989
Youth IQ	-0.938^{***}
	(0.075)
Age	-0.025
	(0.056)
Youth SES	-0.327^{***}
	(0.067)
(Intercept)	-2.593^{***}
	(0.066)
Observations	4,552
Log Likelihood	−1,194.531
Akaike Inf. Crit.	2,397.063

Note: *p<0.1; **p<0.05; ***p<0.01

This regression uses data from the National Longitudinal Survey of Youth 1979. The sample is restricted to whites and those who were out of the labor force in 1989 due to schooling. "Poverty in 1989" is a binary dependent variable which indicates whether an individual was below the poverty line in 1989. "Youth IQ" is the standardized AFQT (Armed Forces Qualification Test) score for an individual in their late teens, with sample weights and skew adjustments discussed by HM in Appendix 2 of The Bell Curve. "Youth SES" is the standardized socioeconomic status of an individual's parent in the year 1979. This index is calculated using standardized values for grades completed by the parents, household income, and Duncan values for occupational prestige. Observations with missing values for any of the LHS or RHS variables were removed from the sample.

Table 4: Relationship Between IQ/Parental Socioeconomic Status and Poverty - NLSY 97

	Dependent variable:
	Poverty in 2007
Youth IQ	-0.345^{***}
	(0.063)
Age	-0.177^{***}
	(0.056)
Youth SES	-0.133^{**}
	(0.062)
(Intercept)	-2.037^{***}
	(0.058)
Observations	3,101
Log Likelihood	−1,113.421
Akaike Inf. Crit.	2,234.841

Note: *p<0.1; **p<0.05; ***p<0.01

This regression uses data from the National Longitudinal Survey of Youth 1997. The sample is restricted to whites. "Poverty in 2007" is a binary dependent variable which indicates whether an individual was below the poverty line in 2007. "Youth IQ" is the standardized AFQT score of an individual in their teens, without the adjustments for skew used by HM. "Youth SES" is the standardized socioeconomic status of an individual's parent in the year 1997. This index is calculated using standardized values for grades completed by the parents and household income, similar to the methods used by Herrnstein and Murray, with the exception that this SES index excludes Duncan values for occupational prestige. Observations with missing values for any of the LHS or RHS variables were removed from consideration.

Table 3 shows that increasing an individual's IQ by one standard deviation is associated with a reduction in the odds of being in poverty by 93 (!) percentage points. The equivalent change in SES is associated with a 32 percentge point reduction. Table 4 is the same logistic regression but on the NLSY 1997. In this data set, increasing IQ by one standard deviation is associated with a 35 percentage point reduction in the odds of being in poverty, compared to the 13 percentage point reduction from increasing SES by one standard deviation. The importance of IQ relative to SES is confirmed, though both effects are smaller compared to the effects from the NLSY '79.

Conclusion

In *The Bell Curve*, Charles Murray and Richard Herrnstein hypothesize that an individual's IQ at a young age is more important than his family's socioeconomic environment in determining whether he ends up in poverty. Their analysis uses data from the 1979 National Longitudinal Survey of Youth. We apply their methodology on a new data set, the 1997 NLSY, and support their idea that a strong relationship exists between childhood IQ and future chances of being in poverty. Moreover, this relationship is stronger than the relationship between family SES and future chances of being in poverty.

We apply nearly the same methodology as HM in analyzing IQ, SES, and poverty. HM measure family SES as a composite of parental education, income and occupation. We do the same except we exclude occupation. We also use the AFQT as an indicator of IQ, without correcting for skew, and measure the poverty status of an individual 10 years after their youth. Each score is standardized so as to allow comaprisons between changes in IQ and SES.

We run two logistic regressions of Youth IQ, Youth SES, and Age on Poverty Status 10 years later: one that replicates HM's analysis for the NLSY '79, and a new one on the NLSY '97. For the NLSY '79, we find that increasing Youth IQ by one standard deviation yields a 94 percentage point decrease in the odds of ending up in future poverty, compared to the 33 percentage point decrease associated with increasing Youth SES by one standard deviation.

For the '97 data, the former number is 35 while the latter is 13. We therefore support HM's hypothesis that Youth IQ is more important than Youth SES in avoiding future poverty, but our effects are smaller overall.

The discrepancy in magnitudes between the NLSY '97 and '79 analyses could be attributed to a couple factors. First, by excluding occupational prestige, we may not be capturing the full extent of socioeconomic status on an individual. Second, we make no skew adjustment to IQ scores, whereas HM did, and this may very well diminish the effect of IQ on poverty, especially at the tails of the distribution. Third, since we measure poverty in 2007, the economic recession could have plausibly contributed to high poverty rates seen across cognitive and SES classes in the NLSY '97. Moreover, the recession could disproportionately affect those on the lower ends of both distributions. While we make no adjustment of when poverty is calculated to stay as close to HM's methodology as possible, future research interested in this question ought to examine poverty across multiple years.

Our results have implications for policymakers interested in lowering poverty rates. Targetting an individual's socioeconomic environment may not be as effective as putting dollars towards increasing their IQ. Many federal programs rest on the assumption that outcomes can be improved with training, funding, and support from the government. The extent to which IQ affects individual outcomes is at the center of debates on education, welfare, and other government programs (Winship and Korenman 1997). Like *The Bell Curve*, our research warrants further examination on how much of IQ is nature and how much is nurture. Herrnstein and Murray take the view that government and one's environment can do little to improve individual IQ. Some research suggests that the environment can have a significant causal effect on future life outcomes (Chetty and Katz 2016). Regardless, it is hard to dispute the role intelligence plays in an individual's life, and we hope to encourage further research on the complex relationship between IQ and poverty.

References

Bernie Devlin, Daniel P. Resnick, Stephen E. Fienberg, and Kathryn Roeder. 1994. *Intelligence, Genes, and Success: Scientists Respond to *the Bell Curve**. Springer Science.

Chetty, Nathaniel Hendren, Raj, and Lawrence Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Project." *American Economic Review* 4 (106). The National Bureau of Economic Research.

Corcoran, M. 1995. "Rags to Rags: Poverty and Mobility in the United States." *Annual Review of Sociology* 21. University of Michigan: 237–67. https://doi.org/10.1146/annurev.so.21.080195.001321.

Gelman, Andrew, and Eric Loken. 2013. "The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There Is No 'Fishing Expedition' or 'P-Hackin' and the Research Hypothesis Was Posited Ahead of Time."

Greg J. Duncan, Jeanne Brooks-Gunn, W. Jean Yeung, and Judith R. Smith. 1998. "How Much Does Childhood Poverty Affect the Life Chances of Children?" *American Sociological Review* 63 (3): 406–23. http://www.jstor.org/stable/2657556?seq=1#page_scan_tab_contents.

Herrnstein, Richard, and Charles Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. The Free Press.

Korenman, Sanders, and Christopher Winship. 1995. "A Reanalysis of the Bell Curve." *NBER Working Paper Series*, no. 5230. The National Bureau of Economic Research. http://www.nber.org/papers/w5230.

R Core Team. 2013. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Will Beasley, David Bard, Joe Rodgers, and Kelly Meredith. 2018. *National Longitudinal Survey of Youth*. NLSY. http://www.bls.gov/nls/.

Winship, Christopher, and Sanders Korenman. 1997. "Does Staying in School Make You

Smarter? The Effect of Education on IQ in the Bell Curve." Springer, New York, NY, 215–34. https://doi.org/10.1007/978-1-4612-0669-9_10.