Mass Incarceration

New Jim Crow, Class War, or Both?

NATHANIEL LEWIS

People's Policy Project

Cover

Bosse, Abraham. *Visiting the Prisoners (Visiter les prisonniers)*.
ca. 1629–1666. The Metropolitan Museum of Art. Harris Brisbane Dick Fund, 1926. Public domain. Accessed via web on 7 January 2017.
www.metmuseum.org/art/collection/search/390811

About

People's Policy Project is a think tank founded in 2017. The primary mission of 3P is to publish ideas and analysis that assist in the development of an economic system that serves the many, not the few.

PeoplesPolicyProject.org

Author

Nathaniel Lewis is an independent researcher and a graduate of New College of Florida.

Colophon

Titles are set in Harriet Display and Bodoni URW.
Body is set in Harriet Text. Data values are set in Franklin Gothic.

Design by Jon White.

Using data from the National Longitudinal Study of Adolescent to Adult Health, I analyze racial and class disparities in incarceration. My analysis shows that class status has a large and statistically significant effect on **(1)** whether or not men aged 24–32 years have ever been to jail or prison; **(2)** whether or not men are jailed after being arrested; **(3)** whether or not men have spent more than a month in jail or prison; and **(4)** whether or not men have spent more than a year in jail or prison. After controlling for class, I do not find race to be a statistically significant factor for the first three outcome categories, but I do find that race has a significant impact on whether or not a man has spent more than a year in prison or jail.

Introduction

Americans—black Americans in particular—are incarcerated at jaw-dropping rates. The US is home to less than 5% of the world's population, but over 20% of the world's prisoners (Walmsley 2016). Blacks make up about 13% of the US population, but 40% of the prison population (Sakala 2014). Countless studies have demonstrated this racial disparity within a highly carceral state, including disparities in arrest rates (Langan 1995), plea bargain offers (Berdejó 2017), and sentencing outcomes (Kansal 2005).

There are two common explanations for these disparities in left-of-center thought. The first explanation is that mass incarceration exists as a racist system for managing black people as black people following the end of formal Jim Crow laws and the successes of the civil rights movement. Michelle Alexander offers this view in her widely acclaimed book The New Jim Crow (Alexander 2012). The second explanation is that mass incarceration exists as a capitalist system for managing poor people, following the rollback of the liberal social welfare state and other neoliberal reforms that left the working class to fend for itself. This perspective is presented by Cedric Johnson in his essay "The Panthers Can't Save us Now" (Johnson 2017).

Both of these views posit that black Americans are disproportionately poor because of a long history of racism in America, but one holds that black people are disproportionately incarcerated primarily because they are black, while the other holds that it is primarily because they are poor.

The discovery of racial disparities in the criminal justice system cannot, by itself, tell us which of these views is closer to the truth since both views predict and explain racial disparities. Therefore, we need to look at incarceration through some measure of class in conjunction with race. Below, I briefly review studies that touch on this issue and then proceed to my own study of this question.

1

Literature Review

Rabury and Kopf (2015) use the **Bureau of Justice Statistics** (BJS) **Survey of Inmates in State Correctional Facilities and the Census Bureau's American Community Survey** (ACS) to produce descriptive statistics on the pre-incarceration income of prisoners compared to the income of the non-incarcerated population. Their study found that the median annual pre-incarceration income of black male prisoners aged 27–44 years was $17,625, while the median income for non-incarcerated black males in the same age group was $31,245. Among white male prisoners and non-prisoners in that age group, the annual income was observed to be $21,975 and $47,505, respectively. Overall, 57% of all male prisoners made less than $22,500 before they were incarcerated, while 57% of non-incarcerated males made over $37,500.

Although a useful contribution, Rabury and Kopf's study suffers from some important limitations. The study presents a detailed comparison of income for incarcerated and non- incarcerated men and women, but this is not cross-tabulated by race. Thus, we cannot say, for example, that poor blacks are equally as likely to be locked up as poor whites, or whether rich blacks are equally as likely to be incarcerated as rich whites. Further, these statistics are potentially confounded **(1)** by re-incarcerated prisoners, who might have been earning less than they would have had they never been imprisoned, and **(2)** by prisoners who had been imprisoned from a young age.

Western and Pettit (2010) use the **National Longitudinal Survey of Youth** (NLSY) to compare incarceration rates across race and education levels. They conclude that education has a stronger relationship with incarceration than race, but that both education and race are important factors. Their findings are summarized in the table below.

TABLE 1

Risk of Incarceration by Age 30-34 for Men Born Between 1975-1979

	Less Than HS	High School	College
White	28%	6.2%	1.2%
Black	68%	21.4%	6.6%
Latino	19.6%	9.2%	3.4%

Table 1 presents the cumulative likelihood of having ever been incarcerated by age 30–34 for racial and educational groups as provided in Western and Pettit (2010).

Western and Pettit's study sheds important light on educational disparities in incarceration, but it does not analyze class. Education can serve as a rough proxy for class or socioeconomic status when other population characteristics are unknown, but there are problems with using educational attainment as the sole measure of class. One problem is that at any given educational level blacks have much lower incomes and wealth than their white counterparts.

The below table shows the median wealth of families headed by someone between the age of 25 and 34 in 2016 as reported by the **Federal Reserve's Survey of Consumer Finances** (SCF).

TABLE 2

Median Wealth of Families with Head Ages 25–34 in 2016

	Less Than HS	High School	Bachelor's or Higher
White	$4,290	$24,300	$52,205
Black	$3,640	$2,800	$1,300
Latino	$9,500	$17,200	$21,735

Table 2 presents the median wealth of families headed by individuals between the ages of 25 and 34 for racial and educational groups as provided by the Survey of Consumer Finances.

Below are the median earnings of men between the age of 25 and 34 in 2016 as reported by the Census Bureau in the Annual **Social And Economic** (ASEC) **Supplement of the Current Population Survey** (CPS).

TABLE 3

Median Earnings of Men Ages 25–34 in 2016

	Less Than HS	High School	Bachelor's or Higher
White	$25,492	$35,022	$58,012
Black	$20,292	$29,090	$55,502
Latino	$22,541	$26,935	$44,785

Table 3 presents the median earnings of men between the ages of 25 and 34 for racial and educational groups as provided by the Current Population Survey.

Since income and wealth are important aspects of class, a measure of class that relies solely on educational attainment is inadequate.

Zaw, Hamilton, and Darity (2016) use the NLSY to compare black and white rates of incarceration at different wealth levels. They find that both race and wealth impacted incarceration rates. Specifically, blacks at each wealth decile have a higher chance of ending up in jail or prison than whites within the same decile, while blacks within the lower deciles have a higher chance of imprisonment than blacks at higher deciles. Remarkably, they also observe that blacks of all but the highest decile were incarcerated at higher rates than even the lowest decile whites.

Zaw et al. use household wealth of people in their early 20s, which is theoretically an apples-to- apples comparison among respondents. However, as journalist Ryan Cooper argues (2016), this approach likely obscures the possibility that higher-class people in their 20s might be temporarily in debt due to student debt. Thus, someone who would otherwise belong to the higher classes might temporarily make a youthful appearance in the lower wealth deciles. Indeed, as seen in Table 2 above, wealth among young black families is actually lower for higher educational groups than lower educational groups.

In order to avoid the confounding effect that incarceration may have on present wealth, Zaw et al. eliminate from their sample anyone who had been incarcerated before the NLSY started measuring respondents' wealth. By doing so, a relatively large subsample of people who had been jailed are excluded. The study ends up finding 87 whites and 231 blacks who had been incarcerated by 2012, but this excludes 103 whites and 85 blacks who had been incarcerated before 1985. Further, as the researchers report, the wealth level of the excluded sample was considerably lower than that of the retained sample. Were this subsample included, the effects of wealth would have likely been amplified, with the effects of race diminished.

Due to the constraints of the data, Zaw et al.'s method for determining who had been incarcerated mostly counts respondents who were in jail or prison at the time of the yearly survey. While this approach does count all respondents who served terms over one year, it likely undercounts those who, for example, might have been in jail once for three months or in jail for a few weeks every few years. This does not have an obvious effect—it could be that the survey underreports black, white, rich, and poor exactly proportionally—but it is worth noting as a limitation, since it is not clear exactly what their measure of incarceration means in real-world terms.

Finally, while descriptive statistics can be highly informative, if we are working with unequally small subsamples, it can be misleading to generalize findings to the entire population, or to posit a real effect rather than random chance. In the case of Western and Pickett (2010), describing large differences in incarceration rates by education level and race, we can probably dispense with statistical tests and let the descriptive statistics do the work, as the effect is apparently so large. We would, no doubt, find both education level and race to be statistically significant factors in incarceration rates, with large effects. In the case of Zaw et al., we end up with 38 blacks in the ninth wealth decile versus 208 whites in that same decile versus 310 blacks in the second decile. Thus it only takes 3 weighted respondents to describe the ninth wealth decile as having a 9.65% incarceration rate for blacks. In this case, intuition suggests that inferential statistical testing would be prudent so that we might determine the individual impact of each variable and the generalizability of these impacts to the entire US population.

2

Methods

Overview

For my analysis, I use the publicly available datasets for the **National Longitudinal Study of Adolescent to Adult Health** (Add Health). This study follows a nationally representative sample of people who were in grades 7 through 12 during the first wave of 1994–95. Wave IV was conducted in 2008 when participants were 24–32 years old. The public dataset for respondents who completed both waves has a sample size of 5,114. The restricted dataset has a much larger sample size and is restricted in order to make it difficult to deduce any of the respondents' identities from the data. Although this type of restriction is common for public datasets, it does introduce the possibility of selection bias which may impact any given study. The restricted dataset would therefore be optimal for use in the present study, but is difficult to obtain, and the public dataset is apparently well-constructed, with care put to ensure as accurate an analysis as possible (e.g., the inclusion of special post-stratified weight and cluster variables specific to the public dataset.)

While the Add Health study focuses on aspects of health, Wave IV asks a series of questions on incarceration. Respondents are asked **(1)** whether they have been jailed after an arrest, **(2)** whether they have been to jail or prison since turning 18, and **(3)** how many months and years total they have spent in jail or prison. Thirty of the respondents were in prison at the time of the survey, eliminating the need to ask if they'd ever been arrested or in jail or prison. There is also a series of follow-up questions concerning types of crimes and charges, which is outside of the scope of the present study, but which may be valuable for anyone looking to use Add Health data for similar purposes.

One drawback of the dataset is that it is limited to an age group of 24–32 year-olds, meaning we are potentially missing any disparities that open up after the age of 32. However, according to a BJS report on the "Lifetime Likelihood of Going to State or Federal Prison" (Bonczar and Beck 1997), this age range captures the majority of people who will go to state or federal prison. Thus it is reasonable to look at this age range.

Since we are interested in how race and class relates to incarceration disparities, we need a race variable and a class variable. For the race variable, I simply take self-reported race from Wave I of the study. Self-reported sex is also taken from Wave I, as sex is highly relevant to incarceration statistics. Though I include all races and both males and females in my subsequent regressions, my primary focus is on the differences between non-Hispanic white males and black males, since males represent the vast majority of the incarcerated population and since the disparity between blacks and whites is the matter of interest.

As partially demonstrated in the literature review section, finding a suitable variable to use as a proxy for class is far from straightforward, particularly when applying it to young adults of different races, and even more particularly when applying it to an outcome like incarceration, which itself can influence class status. There are two basic approaches. The first is to find a single aspect of class to use as a proxy. Most commonly educational attainment, wealth level, income level, or occupational status is used as a stand-in for class. However, any of these given variables faces a set of confounders making them incapable of presenting a robust picture of class on their own. Following from this limitation, the second approach is to create a composite variable that combines information from the various markers of class into a single index. Several different variables are weighted and summed, and the result is a more complete picture of class status than any of the single variables might give on its own.

There are different approaches to creating a composite index variable, each with its advantages and drawbacks. For robustness, this study employs three approaches: the mathematically sophisticated principal component approach; the clearly communicated intuitive weighting approach; and an agnostic approach of equally weighting each of the variables. These approaches are discussed in more detail below, following a discussion of the variables that were chosen as components of class status.

Component variables of the class composite index variable

1. Educational attainment. Educational attainment is often useful as a proxy for class. For example, Western and Pettit (2010), discussed above, examine incarceration rates by race and educational attainment and find stark disparities along both dimensions. Educational attainment is also a socially understood marker of class, as seen in the political realm where the "working class" is often defined simply as those without a college degree. Moreover, in the long run, educational attainment is highly correlated with other markers of class, such as income, wealth, and occupational status.

There are two primary drawbacks to using educational attainment by itself as the only proxy for class in the present study. First, as discussed already, blacks with the same educational attainment as whites have lower incomes than their counterparts, and drastically lower levels of wealth. Therefore, if an analysis demonstrates dramatic differences in incarceration rates between blacks and whites with the same educational attainment, the results may be influenced by latent variables such as income and wealth. Secondly, as is the case for other markers of class, educational attainment is potentially determined in part by whether or not somebody has been incarcerated.

Wave IV of the Add Health survey asks a series of questions on educational attainment, which I reduced to five levels indicating highest educational attainment for each respondent: **(1)** did not graduate high school, **(2)** graduated high school, **(3)** some college, **(4)** graduated college, and **(5)** at least some post-college.

---

2. Current household assets. Wealth is potentially a strong proxy for class, since many factors go into its possession, and it can be directly measured. However, it can also be a problematic metric when used on its own, particularly when measured in young adults, for several reasons.

First, young adults have not had many years to potentially accumulate wealth. Even young adults who end up inheriting a lot of wealth may not have received their inheritance yet. Second, as discussed in the literature review section, young adults who might, by any reasonable standard, be considered higher class might have low levels of current wealth due to student loans. Looking only at assets, rather than debt, mitigates this confounder. Third, wealth is also, like educational attainment, potentially confounded by whether or not somebody has been incarcerated (Zaw et al. 2016).

Wave IV of the Add Health survey asks respondents what their total household assets sum up to, excluding home equity. The question is asked categorically so that respondents can choose one of nine dollar ranges. Since these levels were unequally distributed among the sample, I reduced the nine levels to six levels, roughly approximating sextiles.

---

3. Homeownership status. Although I would have liked to include home equity in the measure of household assets, the survey asks participants to exclude home equity from their responses. Instead of recording each respondent's home equity, the survey asks participants whether they are homeowners or not. Because home ownership is a major aspect of wealth accumulation and thus an important class marker, I include it separately in the class composite index variable.

4. Current household income. Like wealth, income is potentially a strong candidate as a stand- alone proxy for class, though again, it is better to look at individuals aged at the peak of their earning power, rather than young adults. Somebody who has spent a lot of years in college might just be coming out and starting at the bottom of his respective career ladder. Meanwhile, somebody who began working directly out of high school may have already advanced near the top of his career ladder by the age of 32. Additionally, as with wealth and educational attainment, income is potentially confounded by whether or not somebody has been incarcerated (Geller et al. 2006).

Wave IV asks respondents their total household income from all sources. The question is again asked categorically, this time with 12 levels. I took the middle dollar amount from each range (i.e. $30,000 to $39,000 was coded as $35,000) and divided this amount by the square root of the household size in order to produce an equivalized household income for each person (following Kochnar and Cohn 2011.) The highest level was asked as "more than $150,000," which I coded initially as $200,000. With the adjusted household income figure calculated, I sorted respondents into quintiles.

---

5. Household income at time of adolescence. Parental income in adolescence is strongly related to an individual's future class position as an adult (Chetty 2014). Parental income in adolescence also has many virtues as an indicator of class. Unlike every other variable in our index, because parental income in adolescence precedes any adult behavior, it cannot be confounded by whether or not somebody has been incarcerated since the age of 18. However, as a stand-alone proxy for class, this variable is inadequate, since the correlation between parental income and eventual offspring income, while very strong, is not perfect (Chetty 2014; The Pew Charitable Trusts 2012).

It should be noted that, for this variable in particular, a large amount of data (23%) was missing, due to many parents not participating in the survey. To correct for this and all other missing data, I employed **multivariate imputation by chained equation** (MICE). For a general and basic introduction to this method, see Azur et al (2011) and van Buuren and Groothuis-Oudshoor (2011).

The parental survey from Wave I asks parents to estimate household income by thousands of dollars. As with current household income, discussed above, I adjusted this figure by household size and sorted the results into quintiles.

Composite indices

I test three different composite indices in the regression models below.

1. PCA Composite. Principal component analysis (PCA) is a validated method for creating socioeconomic status indices (Vyas and Kumaranayake 2006). For an introduction to PCA, see Shlens (2003). Briefly, a set of variables is mathematically sorted by different linear combinations into a number of principal components, each of which is uncorrelated with the other so that any given component explains a given amount of variation within the set of variables.

PCA is a mathematically elegant solution to constructing composite variables, but it is important to understand what, in practical terms, the results of such an analysis indicate. Crudely put, a given principal component is meant to represent a separate dimension of information, and the loading (essentially weighting) of a variable within that principal component is meant to represent how much unique information that variable contributes in relation to the others. PCA cannot directly tell the interpreter what kind of information a principal component represents, and results are very sensitive to the way that variables are coded going into the analysis.

2. Intuitive composite. Because PCA can be somewhat of a black box, it is sometimes useful to construct an index based on intuitive judgment. This way, the reasoning behind the weights can be clearly stated, and the reader of the analysis can determine for herself whether they make sense (Cowan et al., 2012). Following this reasoning, I also construct a composite variable, using the same component variables as above, based on my understanding of the virtues and limitations of each component variable with respect to the investigation at hand.

Because this study attempts to isolate the effect of class status on risk of incarceration, rather than the effect of incarceration on class status, I weight household income at adolescence as heavily as all other variables combined, since it is the only variable used that is immune to the confounding effect of incarceration on class status. In addition to this feature, household income at adolescence is highly correlated with adulthood markers of class: children from high-income parents are likely to end up with high levels of educational attainment, as well as high levels of adult household income, adult wealth, and so on. Indeed, if I could pick only one variable to run the analysis with, it would be this one. However, due to the possibility of social mobility, in this instance, this variable is best supplemented by current markers of class for young adults.

I saw no obvious reason to weight any of the remaining variables more than the others, with the exception of home ownership. There, the information presented is very general, i.e., whether or not somebody owns a home. Thus I considered this to be part of a broader assets category, the primary component of which is household assets excluding home equity. Ideally, the data would present household assets including home equity value, but since it does not, I decided that home ownership should carry around one-quarter of the weight of the broader assets category.

To summarize, the intuitive weighting that I settled on has household income at adolescence constituting one-half of the composite variable, with current income, current assets (including homeownership), and current educational attainment each constituting one-sixth of the index, as can be seen below in **Table 4** below.

3. Equal weight composite. Finally, for the purposes of comparison and robustness testing, a composite that weights each of the five component variables equally was created.

Table 4 shows the weights applied to each variable to construct the three composite indices tested in my regression model.

TABLE 4

Variable	Levels	PCA weight	Intuitive weight	Equal weight
Hh income at adolescence	0.2, 0.4, 0.6, 0.8, 1	0.418	0.5	0.2
Current hh income	0.2, 0.4, 0.6, 0.8, 1	0.123	0.167	0.2
Current hh assets less house	0.167, 0.333, 0.5, 0.667, 0.833,1	0.097	0.127	0.2
Home ownership status	0,1	0.075	0.04	0.2
Educational attainment	0.2, 0.4, 0.6, 0.8, 1	0.288	0.167	0.2

Table 4 compares the weights applied to the component variables for each composite class index variable.

Statistical analysis

Logistic regressions were run for each of the four outcomes: **(1)** whether or not someone had been incarcerated at all since the age of 18, **(2)** whether they had been incarcerated for more than a month in total, **(3)** whether they had been incarcerated for more than a year total, and **(4)** for those who had ever been arrested, whether or not they had been jailed after arrest. Each regression was run with seven different models: one for each of the four component variables and one for each of the three composite variables. All models contained race and sex variables. The Results section primarily details the results of the model containing the PCA composite since it is the most rigorous model. An abbreviated table comparing the seven models can be found in **Appendix B**.

For each analysis, the post-stratified weights supplied with the Add Health public dataset were applied in order to generalize the results, and the provided cluster variable was applied in order to produce accurate standard errors, using the 'survey' package in R (Lumley 2017). Because there were 10 imputed datasets, the analysis was run on each and the results pooled according to Rubin's formulas (Rubin 1987) in order to produce averaged point estimates and standard errors adjusted to the uncertainty of the imputations, using the R package 'mitools' (Lumley 2015).

Results

Table 5 presents the results of the regressions for each outcome using the model containing the PCA-generated class variable. The results are stated in terms of log odds. The Intercept represents black males aged 24-32 from the lowest class quintile. If we wish to know estimated odds instead of log odds, we can exponentiate, and from there, divide odds by one plus odds to get probability. For example, if we want to know the probability of white females from the fourth class quintile having been jailed, we start with the Intercept, add the coefficients for Female, Non-Hispanic White, and 4th class quintile, to get log odds of: -0.084 + -1.424 + -0.199 + -1.258 = -2.965. Then we exponentiate that to get odds of: exp(-2.965) = 0.052. Finally, we can get probability: 0.052/(1+0.052) = 0.049, or a 4.9% chance.

Table 5

Ever Jailed	Log odds	Standard error	Wald X^2	$P(>X^2)$
Intercept *(Black, male, 1st class quintile)*	-0.084	0.141	0.35	0.55
Female	-1.424	0.118	146.6	1.00E-16
Non-Hispanic White	-0.199	0.124	2.6	0.11
2nd class quintile	-0.478	0.138	12	0.00055
3rd class quintile	-0.992	0.155	40.9	1.60E-10
4th class quintile	-1.258	0.176	51.2	8.30E-13
5th class quintile	-2.038	0.209	94.8	1.00E-16
All class quintiles			122.2	1.00E-16

Table 5 continued on next page →

Jailed After Arrest				
	Log odds	Standard error	Wald X^2	$P(>X^2)$
Intercept *(Black, male, 1st class quintile)*	0.995	0.178	49.8	1.70E-12
Female	-0.537	0.156	20.9	4.90E-06
Non-Hispanic White	-0.072	0.154	0.34	0.56
2nd class quintile	-0.223	0.221	2.6	0.11
3rd class quintile	-0.841	0.206	29.4	5.580E-08
4th class quintile	-0.929	0.238	28	1.20E-07
5th class quintile	-1.629	0.248	60.6	6.90E-15
All class quintiles			84.2	1.00E-16

Jailed More Than A Month				
	Log odds	Standard error	Wald X^2	$P(>X^2)$
Intercept *(Black, male, 1st class quintile)*	-1.198	0.236	25.8	3.80E-07
Female	-2.005	0.204	96.9	1.00E-16
Non-Hispanic White	-0.281	0.221	1.6	0.2
2nd class quintile	-0.467	0.216	4.7	0.031
3rd class quintile	-0.922	0.258	12.8	0.00035
4th class quintile	-1.235	0.260	22.5	2.10E-06
5th class quintile	-2.325	0.380	37.4	9.80E-10
All class quintiles			54.4	4.40E-11

Jailed More Than A Year				
	Log odds	Standard error	Wald X^2	$P(>X^2)$
Intercept *(Black, male, 1st class quintile)*	-1.230	0.214	76.1	1.00E-16
Female	-2.461	0.317	437.4	1.00E-16
Non-Hispanic White	-0.813	0.231	42.9	5.80E-11
2nd class quintile	-0.992	0.289	51.4	7.40E-13
3rd class quintile	-1.596	0.373	105.9	1.00E-16
4th class quintile	-2.073	0.403	139.1	1.00E-16
5th class quintile	-3.449	0.834	217.7	1.00E-16
All class quintiles			331.3	1.00E-16

↑ *Table 5. Regression table for the binary outcomes of having been arrested, having been arrested but not jailed, having been jailed for more than a month, and having been jailed for more than a year, for people aged 24–32 years, by sex, race, and class; data source is Add Health public datasets for Wave I and Wave IV.*

As can be seen from the Wald statistics presented in **Table 5**, class as a whole is highly statistically significant ($p < .001$) for each outcome. For each outcome the difference between the bottom class quintile and the third class quintile is statistically significant ($p < .001$), with a large effect size that grows larger as the gap between the classes grows larger. For example, in this model, white men at the bottom quintile have a 43% probability of ever having been jailed, while white men at the top quintile have a 8.9% probability. The statistically significant large effects hold for all models, whether a composite index or a single variable is used as a proxy for class, though the effect is more varied across the component variable models. (A comparison of all models can be found in **Appendix B**.)

For all composite variable models, race is not statistically significant by a $p < .05$ threshold for three of the four outcomes. For these three outcomes, the effect of being white instead of black is much less than of the effect of being in the second class quintile instead of the first class quintile. For some of these same outcomes, race becomes statistically significant in the models that use a single variable, instead of a composite variable, as a proxy for class. For these single variable models, the effect of being in the second class level instead of the first remains stronger, in most cases, than the effect of being white instead of black; in all cases, the effect of being in the third, rather than the first, class level is stronger than the effect of race.

In all models, race does become highly significant ($p < .001$) for the most extreme outcome, i.e., whether or not someone has spent a year or more in jail or prison by the age of 24–32 years, though for all but one of the seven models the effect of being in the middle rather than bottom class level was stronger than the effect of being white rather than black.

For both a more nuanced and a clearer understanding of these results, we can use the coefficients from the regressions along with the covariance matrix to pull out and graph predicted probabilities for black males and white males at each class quintile within an 83.4% confidence interval. The 83.4% confidence interval is convenient for graphically demonstrating statistical significance because overlapping intervals indicate no significance at the $p < 0.05$ level, while non-overlapping intervals indicate significance (Knol et al. 2011). Figure 1 shows these graphs for each outcome, using the PCA composite model. (Similar graphs for all models can be found in **Appendix B.**)

Figure 1

Black White

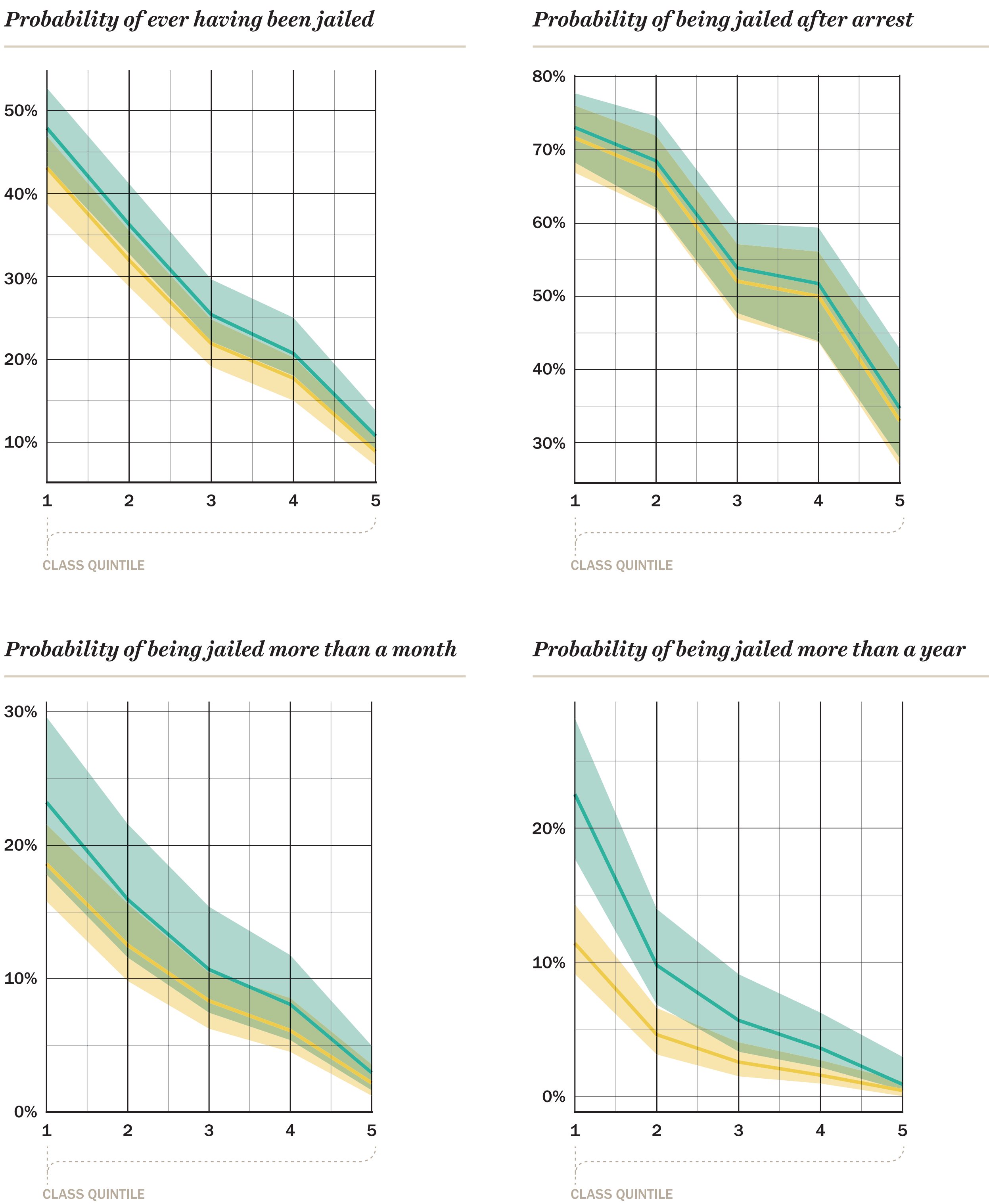


Figure 1 shows the probabilities of having been jailed, having been jailed after being arrested, having been jailed for more than a month, and having been jailed for more than a year, for males aged 24–32, by race and class, with an 83.4% confidence interval.

The graphs in Figure 1 present a clear view of the effects of race and class on various incarceration measures, from which we can make any number of comparisons.

For example, in the bottom left-hand graph, the probability of being jailed more than a month is higher for whites in the bottom class than it is for blacks in the second class; however, since the confidence intervals overlap (if we extend them horizontally), this difference is not statistically significant. We see in all cases the dramatic effects of class; however, the effect of race is perhaps not as dramatic as we might have thought, except when it comes to the lower classes and the probability of being jailed for more than a year.

The relatively small effect of race does not mean that there are relatively small racial disparities in incarceration rates. As **Figure 2** demonstrates, class is not evenly distributed across races.

FIGURE 2

Percent of Black and White Men in Each Class Quintile

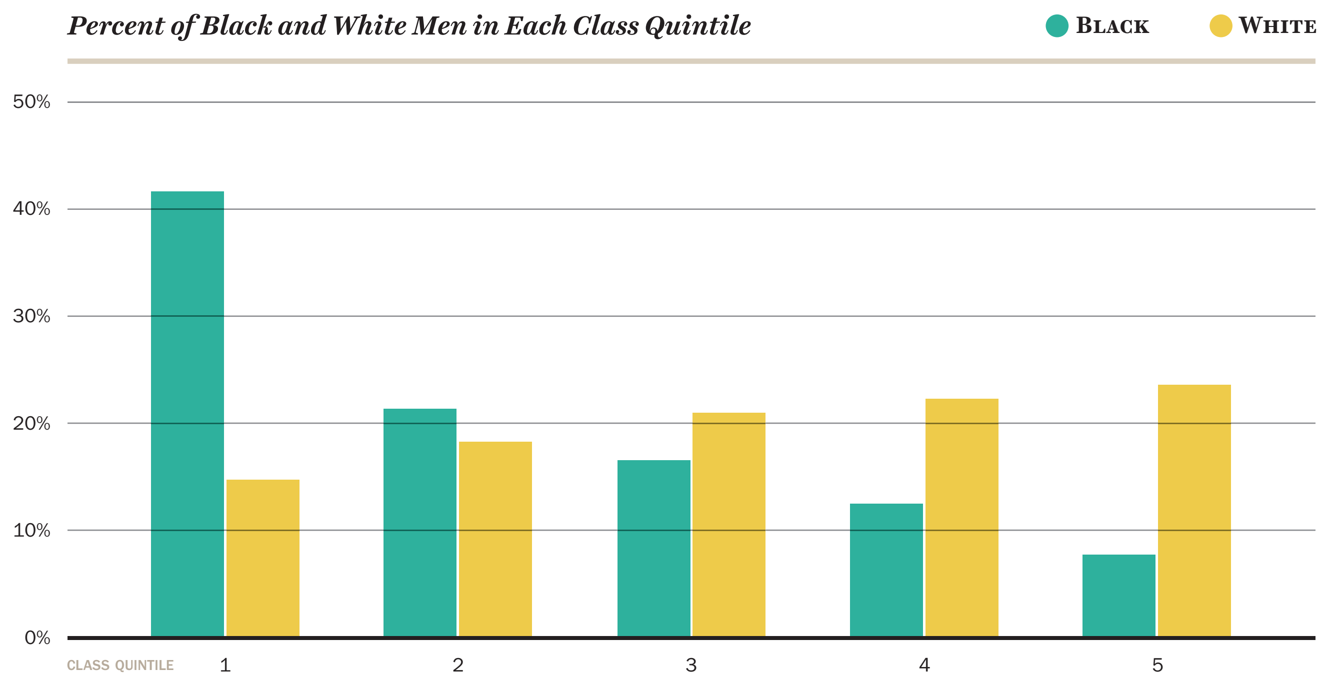


Figure 2 shows what percent of whites and blacks, aged 24–32 years, of both sexes are in each class quintile.

If class were evenly distributed across race, all of the bars in Figure 2 would be at 20%. But they are not. Blacks are highly concentrated in the lowest class and whites are disproportionately represented at the highest class. This indicates that even if within-class incarceration rates were exactly the same for blacks and whites, blacks would still be incarcerated at higher rates, since lower classes are incarcerated at higher rates, and blacks are concentrated in the lower classes.

To better illustrate this, I first calculate the overall estimated probability of sampled white males of all classes facing each outcome and the overall estimated probability of all sampled blacks facing the same outcome. I then compare this to a counterfactual situation, where class is evenly distributed among races (i.e., all the bars in Figure 2 are set at 20%), using predicted probabilities of incarceration at each point of race and class. The difference between the actual predicted probabilities and the counterfactual predicted probabilities represents the proportion of racial disparity explained by the different class composition of black and white men. **Table 6** presents the results.

Table 6

	Black	*White*	*Difference*	*Proportion accounted for by class disparities*
Ever jailed				
Sampled	35.4%	22.8%	12.6pp	72.2%
Counterfactual	28.2%	24.7%	3.5pp	
Jailed after arrest				
Sampled	63.2%	52.7%	10.5pp	84.8%
Counterfactual	56.3%	54.7%	1.6pp	
Jailed for more than a month				
Sampled	16.1%	8.7%	7.4pp	64.9%
Counterfactual	12.1%	9.5%	2.6pp	
Jailed for more than a year				
Sampled	13.0%	3.5%	9.5pp	53.7%
Counterfactual	8.5%	4.1%	4.4pp	

Table 6. Sampled and counterfactual estimated probabilities for each outcome for all class quintiles, and the proportion of racial disparities accounted for by class disparities

The results indicate that class explains at least half of racial disparities for each outcome, assuming my point estimates are either correct or biased in the same direction.

3

Conclusion

This study takes a careful account of class and how it relates to race and incarceration rates. Previous studies interested in racial disparities across various outcomes all too often fail to control for class at all, or else pick a single variable as a proxy for class, which comes with a set of confounders. The constructed class variables used here attempt to balance out the confounders lurking in any one proxy variable. The result, robust across different methods of composite construction, is that class appears to be a larger factor than usually reported when studying racial disparities. It may indeed come as a surprise to many that race is not a statistically significant factor for many incarceration outcomes, once class is adequately controlled for.

To an extent, this study provides weight to the assertion that mass incarceration is primarily about the systematic management of the lower classes, regardless of race. It would be reasonable to conclude then that if policymakers wished to eliminate the phenomenon of mass incarceration, and the negative effects it has on black Americans, they should look to reducing class disparities in universal ways. For example, single-payer health care, a federal job guarantee, a universal basic income, a livable minimum wage, universal childcare, universal education. These are all policies that would likely reduce class disparities and provide the material means to lift a large swath of people out of the scope of the criminal justice system.

On the other hand, this study demonstrates a large racial gap, even controlling for class, when it comes to the most devastating outcome: long appearances in jail and prison. The current popular effort to draw attention to racial disparities as racial disparities certainly seems to still hold validity in light of this study. Nevertheless, while a focus on reducing class disparities in a material fashion clearly will not be enough to completely solve the problem of racial bias, it seems evident that this approach would do a lot of good for poor blacks and poor whites alike with respect to the cruel machinery of mass incarceration.

4

Appendix A

PRINCIPAL COMPONENT ANALYSIS DETAILS

As discussed in the main body, the primary model analyzed in this paper employs a composite index constructed through PCA. Analysis was run on the set of component variables, with their values set to the average across the 10 imputations. Each of the variables was converted from factors to numeric variables (see Vyas and Kumaranayake (2006)), and run through the 'prcomp' function in the standard R package 'stats.' **Table A1** presents the results.

TABLE A1

	PC1	*PC2*	*PC3*	*PC4*	*PC5*	*Standardized weighted sum*
Proportion of variance	0.449	0.197	0.155	0.118	0.082	1
HH income at adolescence	0.444	0.434	0.244	0.741	0.073	0.418
HH income current	0.553	-0.019	-0.260	-0.158	-0.775	0.123
Educational attainment	0.462	0.369	0.357	-0.643	0.331	0.288
HH assets minus home equity	0.459	-0.321	-0.631	0.069	0.532	0.097
Home ownership status	0.272	-0.756	0.589	0.087	-0.004	0.075

Table A1 shows the PCA on selected variables. The top row indicates how much variance each principal component accounts for. The remaining rows present the PCA coefficients, or loadings, for each variable within a given PCA. The final column presents the final weight assigned to each variable.

There are different approaches to using PCA to construct an index. One is to simply take the loadings from the first PC and use them directly as weights. I take an approach similar to Krishnan (2010), which is to additively combine the PCs, weighted by the proportion of variance each one explains, and then standardize the results. So, for example, the final weight for educational attainment is the PC1 loading for that variable (0.433) multiplied by the proportion of variance that PC1 explains (0.431), plus the same formula for PC2 (0.364 * 0.197), plus the same for PCs 3, 4, and 5. The sum of all of this is then standardized so that the final weights of all variables add up to 1.

PCs can be difficult to interpret. The reason that I include them all is that it seems as though these variables together, and the different linear combinations that constitute the various PCs, must describe different dimensions of class (indeed that's why they were all chosen in the first place, to account for the shortcomings of each other), and because each PC makes up a nontrivial amount of total variance.

So, for example, current household assets minus home equity and current household income are negatively correlated with PCs 2 and 3, while household income at adolescence and educational attainment are positively correlated. I interpret this as meaning that these are the dimensions along which some of the confounders discussed in the main body are addressed, e.g., blacks and whites may have the same education level, but different income levels, and people with high educational attainment in their late 20s may have relatively low levels of current income due to spending so much time in school.

To put it another way, PC1 seems to tell us the degree to which all of the variables agree: people with high education levels generally have high income as did their parents. The remaining PCs tell us about the exceptions specific to the dataset, which the main body of this paper allots considerable space to discussing.

5

Appendix B

MODEL COMPARISON

In this appendix, I provide a comparison of the seven models tested across two different incarceration outcomes: whether or not somebody has been incarcerated since the age of 18 years, and whether they have been incarcerated for more than a year total.

Table B1 is a regression table, while **Figures B1** and **B2** are a series of graphs comparing the same outcomes across models in the fashion of Figure 2 in the main body. Readers can refer to the **Methods** section for a discussion of which variables are included in which model.

In general, the models with composite variables agree closely with each other, particularly on the race coefficients, and the general relative importance of class over race when it comes to incarceration. The single variable models are more varied, and in general attribute greater effects to race than the composite variables do.

Table B1 on next page →

Table B1

	PCA composite	Intuitive composite	Equal weight composite	Current income	Current assets	Education level	Parental income
Ever jailed							
Intercept (Black male, first level)	-0.08 (0.14)	-0.10 (0.14)	-0.02 (0.14)	-.03 (0.15)	-0.08 (0.14)	0.41 * (0.16)	-0.18 (0.15)
Female	-1.42 † (0.12)	-1.45 † (0.12)	-1.46 † (0.12)	-1.51 † (0.12)	-1.47 † (0.12)	-1.29 † (0.12)	-1.40 † (0.11)
Non-Hispanic White	-0.20 (0.12)	0.22 (0.13)	-0.20 (0.13)	0.37 * (0.13)	-0.49 † (0.13)	-0.42 † (0.11)	-0.35 * (0.13)
Level 2	-0.48 † (0.14)	-0.40 * (0.13)	-0.67 † (0.14)	-0.73 † (0.13)	-0.40 * (0.17)	-0.70 † (0.16)	-0.41 * (0.15)
Level 3	-0.99 † (0.16)	-0.92 † (0.16)	-0.92 † (0.14)	-0.79 † (0.14)	-0.69 † (0.14)	-1.16 † (0.13)	-0.83 † (0.16)
Level 4	-1.26 † (0.18)	-1.18 † (0.18)	-1.29 † (0.15)	-0.95 † (0.18)	-0.99 † (0.17)	-2.38 † (0.24)	-1.01 † (0.15)
Level 5	-2.04 † (0.21)	-1.73 † (0.20)	-2.15 † (0.26)	-1.46 † (0.21)	-0.53 † (0.16)	-2.66 † (0.24)	-0.98 † (0.17)
Level 6					-1.01 † (0.16)		
Jailed more than a year							
Intercept (Black male, first level)	-1.23 † 0.21	-1.20 † (0.24)	-1.12 † (0.23)	-0.98 † (0.25)	-1.19 † (0.27)	-0.95 † (0.23)	-1.33 † (0.24)
Female	-2.46 † 0.32	-2.51 † (0.32)	-2.54 † (0.33)	-2.69 † (0.33)	-2.56 † (0.32)	-2.26 † (0.32)	-2.46 † (0.31)
Non-Hispanic White	-0.81 † 0.23	-0.80 † (0.23)	-0.83 † (0.25)	-0.94 † (0.25)	-1.26 † (0.25)	-1.16 † (0.23)	-0.97 † (0.24)
Level 2	-0.99 † 0.29	-0.91 † (0.30)	-1.45 † (0.34)	-1.71 † (0.36)	-1.06 † (0.37)	-0.61 † (0.24)	-0.55 † (0.31)
Level 3	-1.60 † 0.37	-1.55 † (0.37)	-1.47 † (0.31)	-1.81 † (0.35)	-1.00 † (0.36)	-1.30 † (0.25)	-2.27 † (0.49)
Level 4	-2.07 † 0.40	-2.03 † (0.42)	-1.85 † (0.37)	-1.85 † (0.42)	-1.16 † (0.39)	-4.99 † (0.78)	-1.47 † (0.39)
Level 5	-3.45 † 0.83	-2.74 † (0.51)	-3.38 † (0.75)	-2.61 † (0.53)	-0.97 † (-0.35)	-2.97 † (0.86)	-1.44 † (0.36)
Level 6					-1.03 † (0.33)		

Table B1 is a regression table comparing seven different models across two different incarceration outcomes. Log odd coefficients are listed above standard errors in parentheses. "" denotes significance at a p <.05 threshold while "†" denotes significance at a p <.005 threshold.*

Figure B1

Probability of ever having been jailed

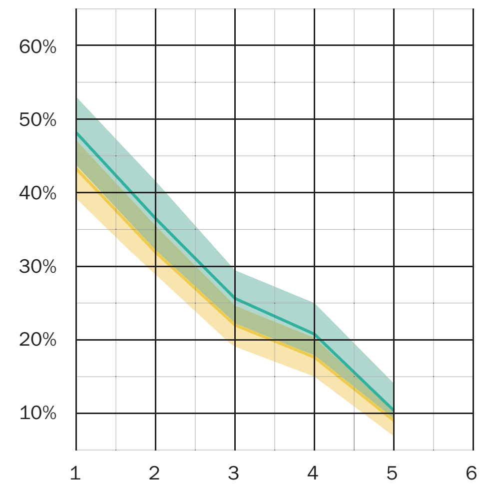
Figure B1 shows a comparison of the predicted probabilities of males aged 24–32 years (by race and class status) having been incarcerated since the age of 18 across seven models with various proxy variables for class.

Each vertical scale represents a range from **5%** (bottom line) to **65%** (top line).

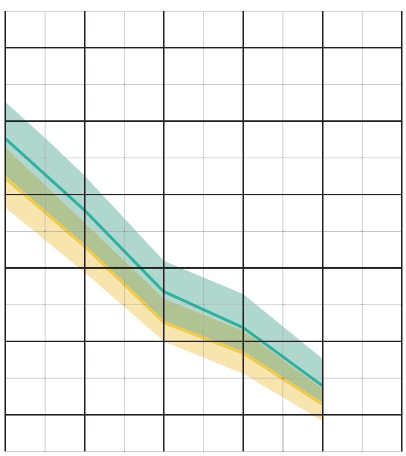
Each horizontal scale represents a range from **1** (leftmost line) to **6** (rightmost line).

White Black

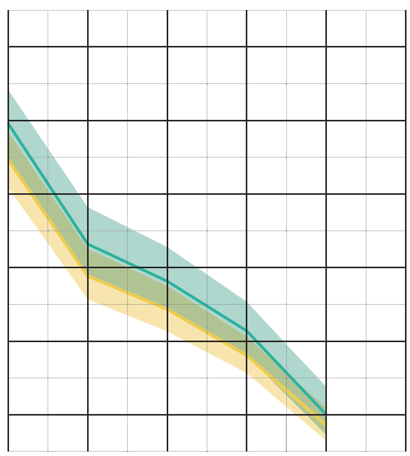
PCA composite



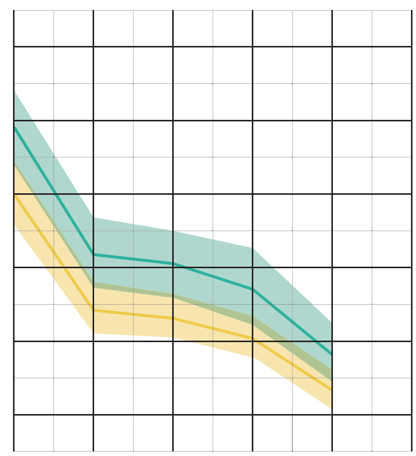
Intuitive composite



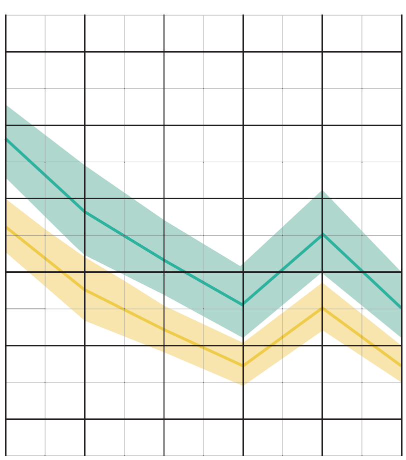
Equal weight composite



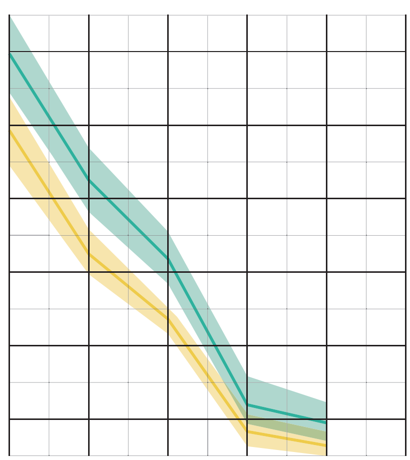
Current household income



Household assets (less home equity)



Education level



Household income at adolescence

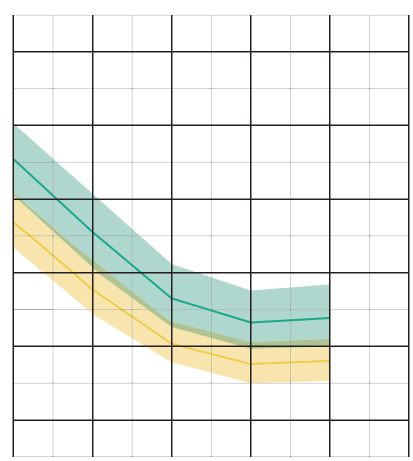


Figure B2

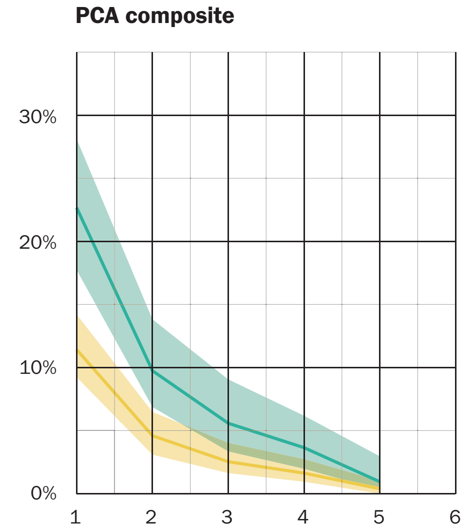
Probability of being jailed more than a year

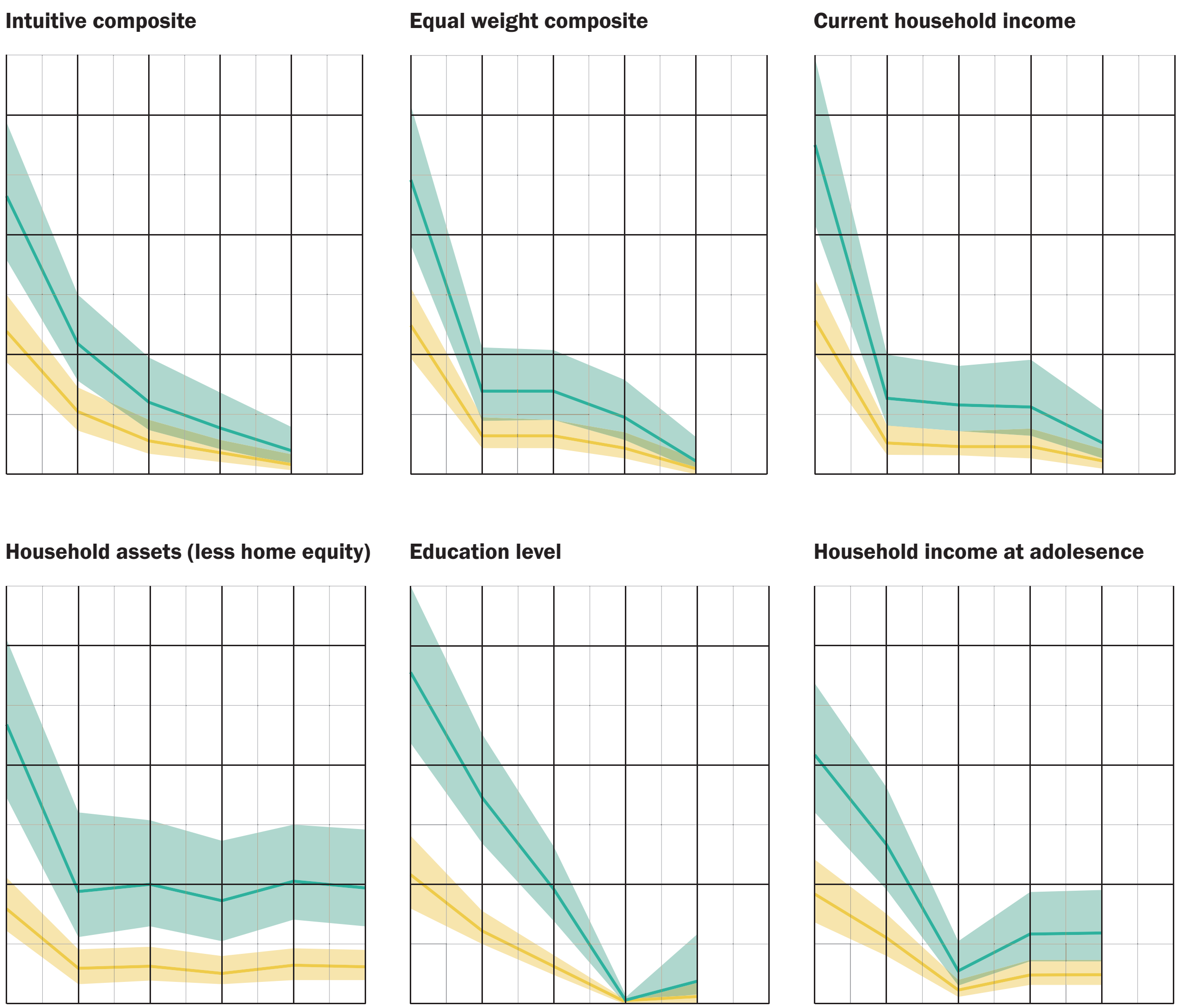
Figure B2 shows a comparison of the predicted probabilities of males aged 24–32 years (by race and class status) having been incarcerated for more than a year total since the age of 18 across seven models with various proxy variables for class.

Each vertical scale represents a range from **0%** (bottom line) to **35%** (top line).

Each horizontal scale represents a range from **1** (leftmost line) to **6** (rightmost line).

White Black





6

Acknowledgements

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

Many thanks as well to Shelley Batts, a medical writer and scientific consultant, for carefully editing an earlier version of this paper (don't blame her for the finished draft though), and providing sharp and invaluable insight. Thanks to Eric Chase-Sosnoff, a PhD candidate in political science at the University of Chicago, for lively discussions around a previous draft of the study. For his statistical guidance, I would like to thank Joshua Loftus, an assistant professor at NYU Stern.

Finally, I would like to thank Khai Zaw, statistical research associate at Duke University, who read over a draft of this paper and offered some very helpful feedback.

7

References

Alexander, M. (2012). The New Jim Crow: Mass Incarceration in the Age of Colorblindness. New York: The New York Press. http://newjimcrow.com/

Berdejó, C. (2017). Criminalizing Race: Racial Disparities in Plea Bargaining. Loyola Law School, Los Angeles Legal Studies Research Paper No. 2017-39. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3036726

Bonczar, T., & Beck, A. (1997). Lifetime Likelihood of Going to State or Federal Prison. Bureau of Justice Statistics. https://bjs.gov/content/pub/pdf/Llgsfp.pdf

Chetty, R., Hendren, N., Kline, P., Saez, E., Turner, N. (2014). Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility. http://www.equality-of-opportunity.org/assets/documents/mobility_trends.pdf

Cooper, R. (2016). This landmark prison study has 1 big problem. The Week. http://theweek.com/articles/645683/landmark-prison-study-1-big-problem

Cowan, C., Hauser, R., Kominski, R., Levin, H., Lucas, S., Morgan, S., Spencer, M., & Chapman, C. (2012). Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress: A Theoretical Foundation. Recommendations to the National Center for Education Statistics. https://nces.ed.gov/nationsreportcard/pdf/researchcenter/Socioeconomic_Factors.pdf

Federal Reserve. (2016). Survey of Consumer Finances. https://www.federalreserve.gov/econres/scfindex.htm

Geller, A., Garfinkel, I., & Western, B. (2006). The Effects of Incarceration on Employment and Wages: An Analysis of the Fragile Families Survey. Center for Research on Child Wellbeing. http://www.saferfoundation.org/files/documents/Princeton- Effect%20of%20Incarceration%20on%20Employment%20and%20Wages.pdf

Harris, K. (2009). The National Longitudinal Study of Adolescent to Adult Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002; Wave IV, 2007-2009. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill. http://www.cpc.unc.edu/projects/addhealth

Johnson, C. (2017). The Panthers Can't Save Us Now. The Catalyst, 1(1). https://catalyst-journal.com/vol1/no1/panthers-cant-save-us-cedric-johnson

Kansal, T. (2005). Racial Disparity in Sentencing: A Review of the Literature. The Sentencing Project. https://www.opensocietyfoundations.org/sites/default/files/disparity.pdf

Knol, M., Pestman, W., & Grobbee, D. (2011). The (mis)use of overlap of confidence intervals to assess effect modification. European Journal of Epidemiology, 26(4), 253-254. https://link.springer.com/article/10.1007/s10654-011-9563-8

Kochnar, R., & Cohn, D. (2011). Fighting Poverty in a Bad Economy, Americans Move in with Relatives, Appendix B. Pew Research Center. http://www.pewsocialtrends.org/2011/10/03/appendix-b-adjusting-household-income-for-household-size/

Krishnan, V. (2010). Constructing an Area-based Socioeconomic Index: A Principal Components Analysis Approach. Early Child Development Mapping Project Alberta. http://www.cup.ualberta.ca/wp-content/uploads/2013/04/SEICUPWebsite_10April13.pdf

Langan, P. (1995). The Racial Disparity in U.S. Drug Arrests. Bureau of Justice Statistics. https://www.bjs.gov/content/pub/pdf/rdusda.pdf

Lumley, T. (2015). Package 'mitools'. The Comprehensive R Archive Network. https://cran.r- project.org/web/packages/mitools/mitools.pdf

Lumley, T. (2017). Package 'survey'. The Comprehensive R Archive Network. https://cran.r- project.org/web/packages/survey/survey.pdf

Rabury, B., & Kopf, D. (2015). Prisons of Poverty: Uncovering the pre-incarceration incomes of the imprisoned. Prison Policy Initiative. https://www.prisonpolicy.org/reports/income.html

Rubin, D. (1987). Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons.

Sakala, L. (2014). Breaking Down Mass Incarceration in the 2010 Census: State-by-State Incarceration Rates by Race/Ethnicity. Prison Policy Initiative. https://www.prisonpolicy.org/reports/rates.html

Shlens, J. (2003). A Tutorial on Principal Component Analysis. https://www.cs.princeton.edu/picasso/mats/PCA-Tutorial-Intuition_jp.pdf

The Pew Charitable Trusts. (2012). Pursuing the American Dream: Economic Mobility Across Generations. Economic Mobility Project. http://www.pewtrusts.org/~/media/legacy/uploadedfiles/pcs_assets/2012/pursuingamericand reampdf.pdf

van Buuren, S. & Groothuis-Oudshoor, K. (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3). https://www.jstatsoft.org/article/view/v045i03

Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: how to use principal components analysis. Health Policy and Planning, 21(6), 459-468. https://academic.oup.com/heapol/article/21/6/459/612115

Walmsley, R. (2016). World Prison Population List, eleventh edition. Institute for Criminal Policy Research. http://www.prisonstudies.org/sites/default/files/resources/downloads/world_prison_population_list_11th_edition_0.pdf

Western, B., & Pettit, B. (2010). Incarceration & social inequality. Dædalus, Summer 2010. https://www.amacad.org/content/publications/pubContent.aspx?d=808

Zaw, K., Hamilton, D., & Darity, W. (2016). Race, Wealth and Incarceration: Results from the National Longitudinal Survey of Youth. Race and Social Problems, 8(1), 103-115. https://link.springer.com/article/10.1007/s12552-016-9164-y