Supplement to:

Bucca, Mauricio. 2024. "Colorism Revisited: The Effects of Skin Color on Educational and Labor Market Outcomes in the United States" Sociological Science 11: 517-552.

1 Online supplement - Colorism revisited: the effects of skin color on educational and labor market outcomes in the United States

Mauricio Bucca
Pontificia Universidad Católica de Chile

Sensitivity analysis of simulated linear structural causal model

I simulate a linear structural causal model (LSCM) (Pearl, 2011) and deploy three empirical strategies to identify the parameter of interest, θ, representing the effect of an individual's skin color (C_1) on the outcome Y_1. For these simulations, I set the true effect of skin color to zero ($\theta = 0$) and manipulate the influences of significant unobserved variables —- parental wealth (W_0) and ability (A_0, A_1) —- assuming no other confounders affect the outcomes. The objective is to evaluate the bias in $\hat{\theta}$ introduced by each identification strategy, given varying levels of omitted variable significance. The steps for the sensitivity analysis are as follows:

1. Step 1: Following the causal framework depicted in Figure **??**, I generate n=20^3 samples using the specified distributions for $\beta = 0$, $\delta = 0$, and $\theta = 0$.

$$
\begin{aligned}
&\text{Parental ability: } \mathbf{A}_0 \sim \text{Normal}(\mu = 0, \sigma = 1) \\
&\text{Parental skin color: } \mathbf{SC}_0 \sim \text{Normal}(\mu = 0, \sigma = 1) \\
&\text{Parental income: } \mathbf{Y}_0 \sim \text{Normal}(\mu = \delta \cdot \mathbf{A}_0 + 0.2 \cdot \mathbf{SC}_0, \sigma = 1) \\
&\text{Parental wealth: } \mathbf{W}_0 \sim \text{Normal}(\mu = 0.2 \cdot \mathbf{SC}_0, \sigma = 1) \\
&\text{Children's ability: } \mathbf{A}_1 \sim \text{Normal}(\mu = 0.5 \cdot \mathbf{A}_0, \sigma = 1) \\
&\text{Children's skin color: } \mathbf{SC}_1 \sim \text{Normal}(\mu = 0.7 \cdot \mathbf{SC}_0, \sigma = 1) \\
&\text{Other Children's attributes: } \mathbf{X}_1 \sim \text{Normal}(\mu = 0, \sigma = 1) \\
&\text{Children's income: } \mathbf{Y}_1 \sim \text{Normal}(\mu = \theta \cdot \mathbf{SC}_1 + \delta \cdot \mathbf{A}_1 + 0.2 \cdot \mathbf{X}_1 + 0.5 \cdot \mathbf{Y}_0 + \beta \cdot \mathbf{W}_0, \sigma = 1)
\end{aligned}
\tag{1}
$$

2. Step 2: I replicate Step 1 for 100 unique values each of $\beta \in (0, 1]$ and $\delta \in (0, 1]$. In every iteration, I estimate the models below using OLS and record $\hat{\theta}$:

$$
\begin{aligned}
&\text{(A)} \quad \hat{Y}_1 = \hat{\alpha}_A + \hat{\theta}_A \cdot C_1 + \hat{\beta} \cdot Y_0 \\
&\text{(B)} \quad \hat{Y}_1 = \hat{\alpha}_B + \hat{\theta}_B \cdot C_1 \\
&\text{(C)} \quad \hat{Y}_1 = \hat{\alpha}_C + \hat{\theta}_C \cdot C_1 + \hat{\omega} \cdot C_0
\end{aligned}
\tag{2}
$$

This method yields $101 \times 101 \times 3 = 30{,}603$ skin color effect estimates. With the true effect set to zero, any estimate deviation represents bias: $\text{bias}(\hat{\theta}) = \hat{\theta} - \theta$.

Figure A1 presents the sensitivity analysis outcomes, indicating that adjusting for parental skin color (strategy C) offers an unbiased estimate of skin color's impact, independent of unobserved variable significance. However, given data limitations, this approach is impractical
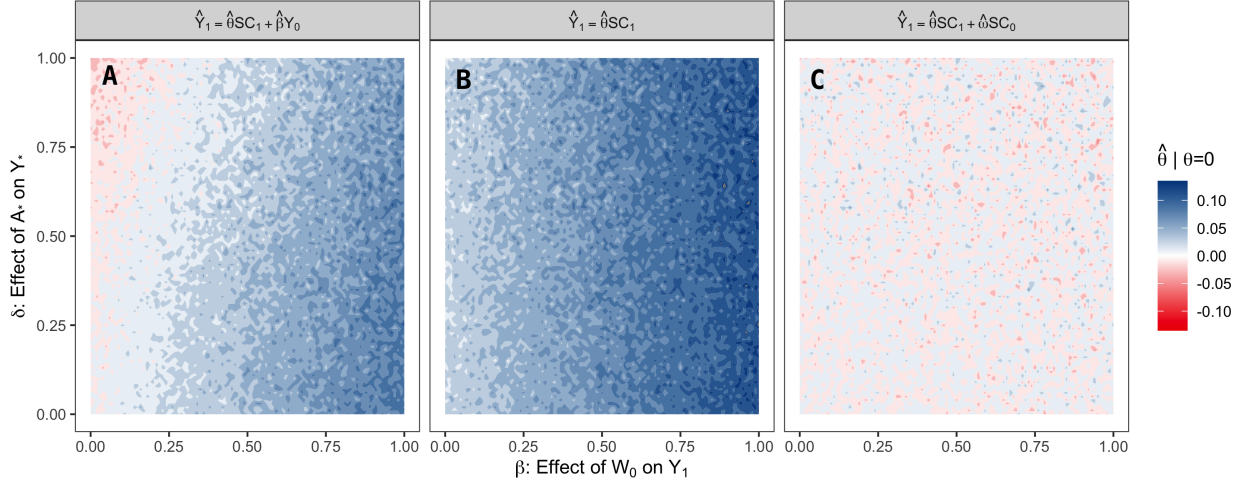
Figure A1: Bias in skin color effect estimates. Each panel shows a contour plot displaying the estimated skin color effect resulting from a particular identification strategy. In these plots whiter coloring indicates less bias, darker blue indicates larger upward bias and darker red indicates larger downward bias. The x-axis corresponds to the effect of unobserved parental wealth (W_0) on the outcome variable, and the y-axis is the effect of parents and children's unobserved ability (A_0, A_1) on the outcome. To facilitate the interpretation note that if the effect of unobserved parental wealth is in the range $\beta \in [0, 1]$ the correlation with the outcome variable is in the range $\rho_{W_0,Y_1} \in [0, 0.7]$. Likewise, if the effect of unobserved ability is in the range $\delta \in [0, 1]$ the correlation between parental ability and the outcome is in the range $\rho_{A_0,Y_1} \in [0, 0.5]$.

Tables

Table A1: OLS models for P(College Degree)

	White		Black		Hispanic	
	Baseline	SB-adjusted	Baseline	SB-adjusted	Baseline	SB-adjusted
AddHealth						
Skin Color	0.09***	0.10***	0.03	0.01	0.07***	0.07**
	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Parents Schooling		0.08***		0.07***		0.04***
		(0.004)		(0.01)		(0.01)
Family income 1994		0.10***		0.04*		0.08*
		(0.02)		(0.02)		(0.04)
Public assistance (any parent)		−0.002		−0.04		0.02
		(0.03)		(0.04)		(0.08)
Intact family		0.07***		−0.004		0.02
		(0.02)		(0.04)		(0.06)
Intercept	−5.24**	−4.98**	−0.01	−2.97	1.26	0.07
	(2.52)	(2.39)	(4.54)	(4.85)	(5.30)	(4.88)
NLSY97						
Skin Color	0.13***	0.07***	0.02	0.01	0.01	−0.01
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
Parents Schooling		0.04***		0.01*		0.02***
		(0.005)		(0.01)		(0.005)
Family income 1994		0.01***		0.03**		0.03***
		(0.003)		(0.01)		(0.01)
Public assistance		−0.19***		−0.13***		−0.04
		(0.03)		(0.05)		(0.03)
Intact family		0.19***		0.10***		0.06
		(0.02)		(0.03)		(0.04)
Intercept	−1.01	1.46	−20.10**	−20.62**	−7.17	−12.08
	(5.51)	(5.09)	(8.04)	(7.97)	(7.98)	(7.75)
Age and gender	yes	yes	yes	yes	yes	yes
Non SB-confounders	yes	yes	yes	yes	yes	yes
Interviewer's race	yes	yes	yes	yes	yes	yes
Observations AddHealth	4,788	3,567	1,532	963	985	647
Observations NLSY97	1,713	1,505	533	434	615	500

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

4

Table A2: OLS models for log Earnings

	White		Black		Hispanic	
	Baseline	SB-adjusted	Baseline	SB-adjusted	Baseline	SB-adjusted
AddHealth						
Skin Color	−0.02	−0.03	0.09**	0.04	0.19**	0.06
	(0.13)	(0.15)	(0.04)	(0.06)	(0.08)	(0.09)
Parents Schooling		0.07***		0.12***		0.03
		(0.02)		(0.02)		(0.02)
Family income 1994		0.23***		0.10		0.26***
		(0.04)		(0.08)		(0.09)
Public assistance (any parent)		−0.14*		−0.16		−0.04
		(0.08)		(0.15)		(0.24)
Intact family		−0.03		0.09		−0.13
		(0.06)		(0.12)		(0.16)
Intercept	3.07	9.42	15.53	20.32	12.87	10.76
	(6.03)	(6.46)	(13.48)	(17.82)	(9.72)	(13.08)
NLSY97						
Skin Color	0.16**	0.10	0.001	−0.02	0.04	0.004
	(0.06)	(0.08)	(0.08)	(0.05)	(0.08)	(0.08)
Parents Schooling		0.01		0.03		0.01
		(0.01)		(0.02)		(0.01)
Family income 1994		0.03***		0.03		0.07***
		(0.005)		(0.02)		(0.02)
Public assistance		−0.29***		−0.28***		−0.12
		(0.05)		(0.10)		(0.08)
Intact family		0.13**		0.28*		−0.01
		(0.05)		(0.14)		(0.10)
Intercept	3.63	4.65	16.34	13.17	−13.58	−8.70
	(11.30)	(12.11)	(27.05)	(23.31)	(18.28)	(18.60)
Age and gender	yes	yes	yes	yes	yes	yes
Non SB-confounders	yes	yes	yes	yes	yes	yes
Interviewer's race	yes	yes	yes	yes	yes	yes
Observations AddHealth	4,736	3,527	1,510	949	972	635
Observations NLSY97	1,667	1,466	480	389	584	473

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

5

Table A3: OLS models for log Household Income

	White		Black		Hispanic	
	Baseline	SB-adjusted	Baseline	SB-adjusted	Baseline	SB-adjusted
AddHealth						
Skin Color	0.05	−0.05	0.15***	0.08	0.28***	0.37***
	(0.15)	(0.16)	(0.05)	(0.09)	(0.10)	(0.13)
Parents Schooling		0.04**		0.16***		0.02
		(0.02)		(0.03)		(0.04)
Family income 1994		0.25***		−0.07		0.45***
		(0.06)		(0.10)		(0.14)
Public assistance		−0.36**		−0.64***		0.88***
		(0.15)		(0.20)		(0.28)
Intact family		0.04		0.28		−0.20
		(0.09)		(0.19)		(0.24)
Intercept	4.52	10.02	−11.39	7.25	35.44*	53.85**
	(9.15)	(9.21)	(15.71)	(19.86)	(18.81)	(22.98)
NLSY97						
Skin Color	0.08	0.03	0.09	0.04	0.20**	0.22**
	(0.05)	(0.06)	(0.06)	(0.06)	(0.09)	(0.09)
Parents Schooling		0.01**		0.03**		0.02
		(0.01)		(0.01)		(0.01)
Family income 1994		0.02***		0.04		0.05***
		(0.004)		(0.02)		(0.01)
Public assistance (any parent)		−0.23***		−0.43***		−0.13*
		(0.03)		(0.11)		(0.07)
Intact family		0.21***		0.24**		0.27***
		(0.04)		(0.11)		(0.08)
Intercept	16.47**	18.62**	31.54	31.62	−14.92	−23.66
	(8.00)	(8.78)	(22.90)	(22.03)	(15.50)	(15.01)
Age and gender	yes	yes	yes	yes	yes	yes
Non SB-confounders	yes	yes	yes	yes	yes	yes
Interviewer's race	yes	yes	yes	yes	yes	yes
Observations AddHealth	4,736	3,527	1,510	949	972	635
Observations NLSY97	1,667	1,466	480	389	584	473

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

6

Table A4: Fixed Effects OLS models for P(College Degree)

	AddHealth			NLSY97		
	White	Black	Hispanic	White	Black	Hispanic
Skin Color	0.03	−0.003	−0.01	0.02	−0.02	−0.002
	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)	(0.03)
Intercept	−4.24	9.10	−2.82	2.74	−9.13**	−6.94
	(3.72)	(7.51)	(2.79)	(5.68)	(4.47)	(8.14)
Age & gender	yes	yes	yes	yes	yes	yes
Observations	691	162	88	496	126	260

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

Table A5: Fixed Effects OLS models for log Earnings

	AddHealth			NLSY97		
	White	Black	Hispanic	White	Black	Hispanic
Skin Color	−0.09***	−0.06	0.03	0.01	−0.02	−0.05
	(0.02)	(0.05)	(0.04)	(0.03)	(0.04)	(0.04)
Intercept	7.41	8.64	6.96	1.35	−78.93***	−15.31
	(13.48)	(18.37)	(13.12)	(12.62)	(21.47)	(17.02)
Age & gender	yes	yes	yes	yes	yes	yes
Observations	683	161	88	478	111	259

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

Table A6: Fixed Effects OLS models for log Household Income

	AddHealth Data			NLSY97 Data		
	White	Black	Hispanic	White	Black	Hispanic
Skin Color	−0.04	0.16	−0.11	−0.01	0.09	0.05
	(0.05)	(0.10)	(0.15)	(0.02)	(0.08)	(0.04)
Intercept	18.27	−13.27	−7.45	−0.08	−10.41	−17.14
	(17.30)	(35.22)	(41.46)	(9.57)	(40.67)	(13.63)
Age & gender	yes	yes	yes	yes	yes	yes
Observations	685	161	87	523	130	280

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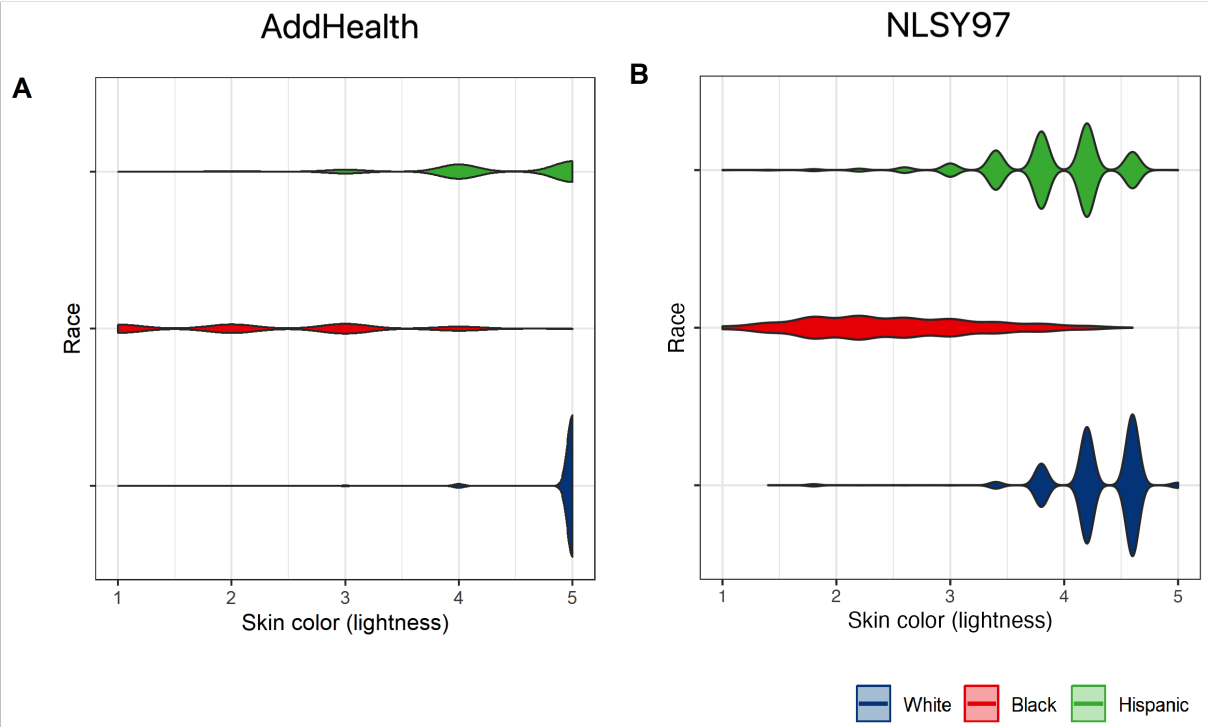
Figures

Figure A2: Skin color distribution and skin color associations by ethnoracial groups. Panel A shows the distribution of skin color for Whites, Blacks and Hispanics. Panels B, C and D show the average college attainment, log earnings and log household income of individuals of each ethnoracial groups at different skin color levels.
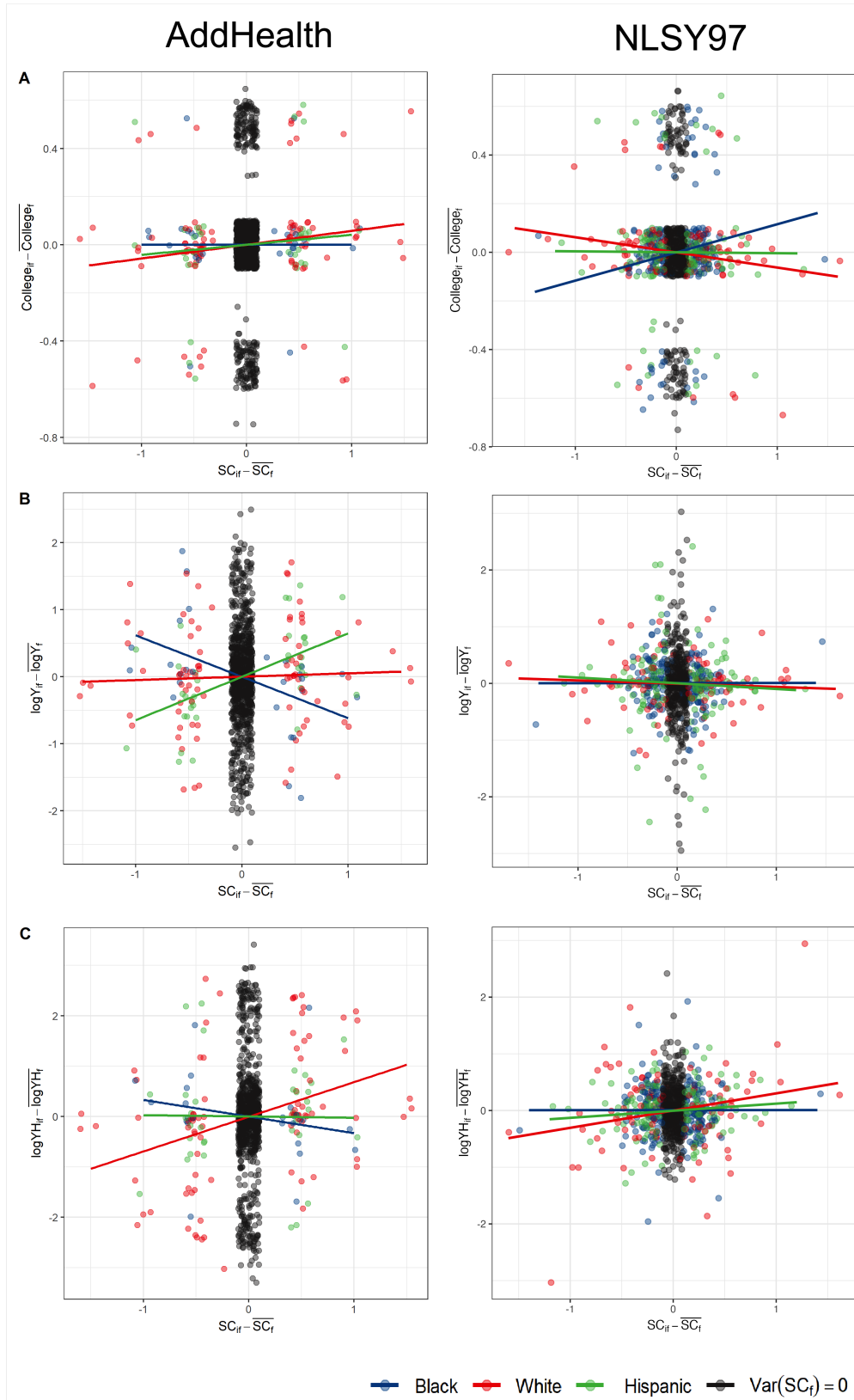
Figure A3: Within-family variation in skin color and socioeconomic outcomes. The x-axis corresponds to the deviation of an individual's skin color with respect to her family's average skin color. The y-axis corresponds to the deviation of an individual's outcome with respect to her family's average in that outcome. Black, White and Hispanic respondents are plotted in blue, red and green, respectively. Black dots correspond to individuals in color-homogeneous families (i.e., excluded from the identification sample in FE models). Dots are jittered to facilitate visualization.
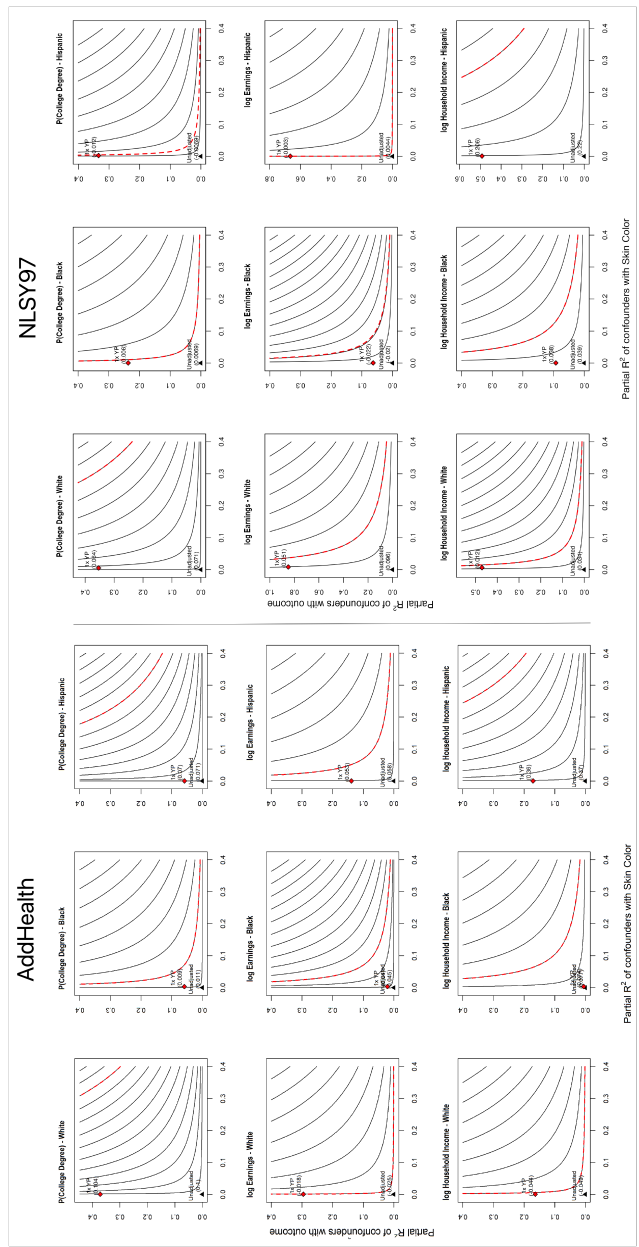
Figure A4: Sensitivity contour plot of skin color effect's point estimate. The x-axis corresponds to the hypothetical share of residual variance of the treatment explained by confounders. Likewise, the y-axis indicates the hypothetical share of residual variance of the outcome explained by confounders. The contour lines show the estimate of skin color effect that one would obtain in a regression model including unobserved confounders with such hypothetical strengths. In particular, the red dashed contour corresponds to the association between confounders, outcome and treatment combinations that would pull the effect estimate to exactly zero. The black triangle corresponds to the observed effect estimate, while the red rhombus is the effect estimate one would obtain in the presence of a confounder(s) as strong as the observed variable YP (parental income).

References

Pearl, Judea. 2011. *Causality: Models, reasoning, and inference, second edition*. Publication Title: Causality: Models, Reasoning, and Inference, Second Edition.