Supplementary Materials on "Equalization or Selection? Reassessing the 'Meritocratic Power' of a College Degree in Intergenerational Income Mobility"

A. The Logic of Residual Balancing

As noted in the main text, our goal is to construct a set of weights such that in the reweighted sample, the association between parental income rank X and other attributes $\boldsymbol{Z}$ does not depend on education. Specifically, we hope that the conditional density of $\boldsymbol{Z}$ given X is the same between college graduates and non-graduates, i.e., $f(z|x, c = 1) = f(z|x, c = 0) = f(z|x)$. In principle, we could estimate both $f(z|x, c)$ and $f(z|x)$ from data and construct the following weights:

$$w_i = \frac{f(z_i|x_i)}{f(z_i|x_i, c_i)}. \tag{1}$$

It can be shown that in the reweighted data, the conditional density of z given x in any educational group c will resemble that in the original population, i.e., $f_w(z|x, c) = f(z|x)$, and controlled mobility can be estimated via a weighted regression of adult income rank (Y) on parental income rank (X) among college graduates. However, since z is usually multidimensional, estimation of $f(z|x, c)$ and $f(z|x)$ is practically difficult. Fortunately, equation (1) can be rewritten as (by Bayes' rule)

$$w_i = \frac{f(c_i|x_i)}{f(c_i|z_i, x_i)}. \tag{2}$$

Since C is a binary variable, the conditional densities $f(c_i|x_i)$ and $f(c_i|\boldsymbol{z}_i, x_i)$ reduce to conditional probabilities, which can be estimated through two logistic regressions. This is the standard inverse probability weighting (IPW) approach to adjusting for time-varying confounding in causal inference (Robins et al. 2000).

However, it is widely recognized that IPW is highly sensitive to model misspecification, relatively inefficient, and susceptible to large finite sample biases (e.g., Lefebvre et al. 2008; Lunceford and Davidian 2004). The method of residual balancing, by contrast, is not only more efficient but also more robust to model misspecification than IPW (Zhou and Wodtke 2018). It can be seen as a generalization of the entropy balancing method (Hainmueller 2012) to causal inference in longitudinal settings. The basic idea is to prioritize finite sample balance rather than asymptotic balance because (a) the latter can never be achieved if the model for $f(c_i|\boldsymbol{z}_i, x_i)$ is misspecified, and (b) even if the model for $f(c_i|\boldsymbol{z}_i, x_i)$ is correctly specified, exactly balanced samples always produce more efficient estimates than probabilistically balanced samples (Imbens and Rubin 2015).

As mentioned earlier, our goal is to construct a set of weights such that in the reweighted sample, $f_w(\boldsymbol{z}|x, c)$ is as close to $f_w(\boldsymbol{z}|x)$ as possible. But since $\boldsymbol{z}$ is high-dimensional, balancing on the entire conditional distribution is practically difficult. Thus we focus on the conditional expectation $\mathbb{E}_w[\boldsymbol{Z}|X, C]$ instead of the full conditional distribution. Specifically, we hope that in the reweighted sample, $\mathbb{E}_w[Z_j|X, C]$ is as close to $\mathbb{E}_w[Z_j|X]$ as possible for each covariate Z_j. Define $\delta(Z_j) = Z_j - \mathbb{E}_w[Z_j|X]$. We can write $\mathbb{E}_w[Z_j|X, C]$ as

$$\begin{aligned} \mathbb{E}_w[Z_j|X, C] &= \mathbb{E}_w[(\mathbb{E}_w[Z_j|X] + \delta(Z_j))|X, C] \\ &= \mathbb{E}_w[Z_j|X] + \mathbb{E}_w[\delta(Z_j)|X, C]. \end{aligned}$$

Thus $\mathbb{E}_w[Z_j|X, C] = \mathbb{E}_w[Z_j|X]$ if and only if $\mathbb{E}_w[\delta(Z_j)|X, C] = 0$. The latter in turn implies $\mathbb{E}_w[\delta(Z_j)h(X, C)] = 0$ for any function $h(X, C)$. Setting $h(X, C)$ as 1 and C respectively leads to the following balancing conditions:

$$\begin{aligned} \mathbb{E}_w[\delta(Z_j)] &= 0 \\ \mathbb{E}_w[\delta(Z_j)C] &= 0 \end{aligned}$$

To this end, we first estimate $\mathbb{E}[Z_j|X]$ by fitting a generalized linear model (GLM) of Z_j on X or its nonlinear transformations (e.g., spline terms). Then $\delta(Z_j)$ can be estimated using the response residuals $Z_j^{\perp} = Z_j - g(\hat{\boldsymbol{\beta}}_j^T \boldsymbol{r}(X))$, where $\boldsymbol{r}(X) = [r_1(X), \ldots r_K(X)]$ is a vector of regressors and $g(\cdot)$ is the inverse link function of the GLM. Hence the sample analogs of the above balancing conditions become

$$\sum_i w_i z_{ij}^{\perp} = 0 \tag{3}$$

$$\sum_i w_i z_{ij}^{\perp} c_i = 0. \tag{4}$$

Moreover, to ensure that the reweighting procedure does not alter the marginal dependence of Z_j on X, we require that the maximum likelihood estimates $\hat{\boldsymbol{\beta}}_j$ of the GLM be invariant before and after reweighting. This can be achieved by imposing the score conditions for the reweighted sample:

$$\sum_i w_i z_{ij}^{\perp} r_k(x_i) = 0 \quad \text{for each } k\text{=}1, \ldots K. \tag{5}$$

Note that equations (3-5) have to hold for each covariate Z_j. With these constraints, we find a set of weights w_i that are as close as possible to the original sampling weights q_i by the Kullback–Leibler divergence metric:

$$\min_{w_i} \sum_i w_i \log(w_i / q_i).$$

This is a constrained optimization problem that can be solved via the method of Lagrange multipliers (see Hainmueller 2012 for details).

A reassuring property of the residual balancing weights is that if we fit a linear/logit regression of Z_j on both $\boldsymbol{r}(X)$ and college completion C, the coefficient on C will be exactly zero and the coefficients on all components of $\boldsymbol{r}(X)$ will be the same as those in the original sample (i.e., $\hat{\boldsymbol{\beta}}_j$). Thus, if the linear/logit model is a correct specification of $\mathbb{E}_w[Z_j|X, C]$, we achieve our goal of conditional mean independence, i.e., $\mathbb{E}_w[Z_j|X, C] = \mathbb{E}_w[Z_j|X]$. Meanwhile, because it minimizes the discrepancy between the new weights w_i and the original weights q_i, the residual balancing algorithm is unlikely to produce extremely large weights. As a result, subsequent analyses based on the reweighted sample are relatively efficient (compared with inverse-probability-weighted samples).

B. Alternative Methods for Adjusting for Selection

Compared with traditional methods for adjusting for selection, the residual balancing approach prioritizes finite sample balance rather than asymptotic balance. As a result, it is generally more efficient and less biased in finite samples. To check the robustness of my main finding, I have also examined controlled mobility using two alternative methods: inverse probability weighting (IPW) and propensity score matching.

Table B1: Estimates of Intergenerational Rank-Rank Slope by College Completion Where Inverse Probability Weighting Is Used to Adjust for Selection.

	Full Sample		Men		Women	
	Conditional Mobility	Controlled Mobility	Conditional Mobility	Controlled Mobility	Conditional Mobility	Controlled Mobility
Intercept	0.297***	0.306***	0.315***	0.324***	0.279***	0.287***
	(0.009)	(0.010)	(0.012)	(0.013)	(0.012)	(0.013)
Parental Income Rank	0.312***	0.322***	0.306***	0.317***	0.315***	0.329***
	(0.017)	(0.020)	(0.023)	(0.027)	(0.024)	(0.028)
College Degree	0.323***	0.090	0.342***	0.239***	0.311***	0.022
	(0.024)	(0.089)	(0.034)	(0.046)	(0.033)	(0.107)
Parental Income Rank * College Degree	-0.141***	0.088	-0.157**	-0.069	-0.131**	0.148
	(0.036)	(0.116)	(0.050)	(0.063)	(0.050)	(0.146)
Sample Size	4,673		2,370		2,303	

Note: †p<.1, *p<.05, **p<.01, ***p<.001 (two-tailed tests). Numbers in parentheses are heteroskedasticity-consistent robust standard errors.

To implement IPW, we need to estimate both the numerator and the denominator of equation (2). To this end, I fit two logit models, one with only parental income rank as the predictor and one with both parental income rank and the nine pre-college covariates (gender, race, Hispanic status, mother's years of schooling, father's presence, number of siblings, urban residence, educational expectation, and the AFQT percentile score) as predictors. The fitted values of these models are used to construct the inverse probability weights (equation 2). These weights are then multiplied by the NLSY custom weights to account for the stratified multistage survey design of NLSY79. Figure B1 shows that in

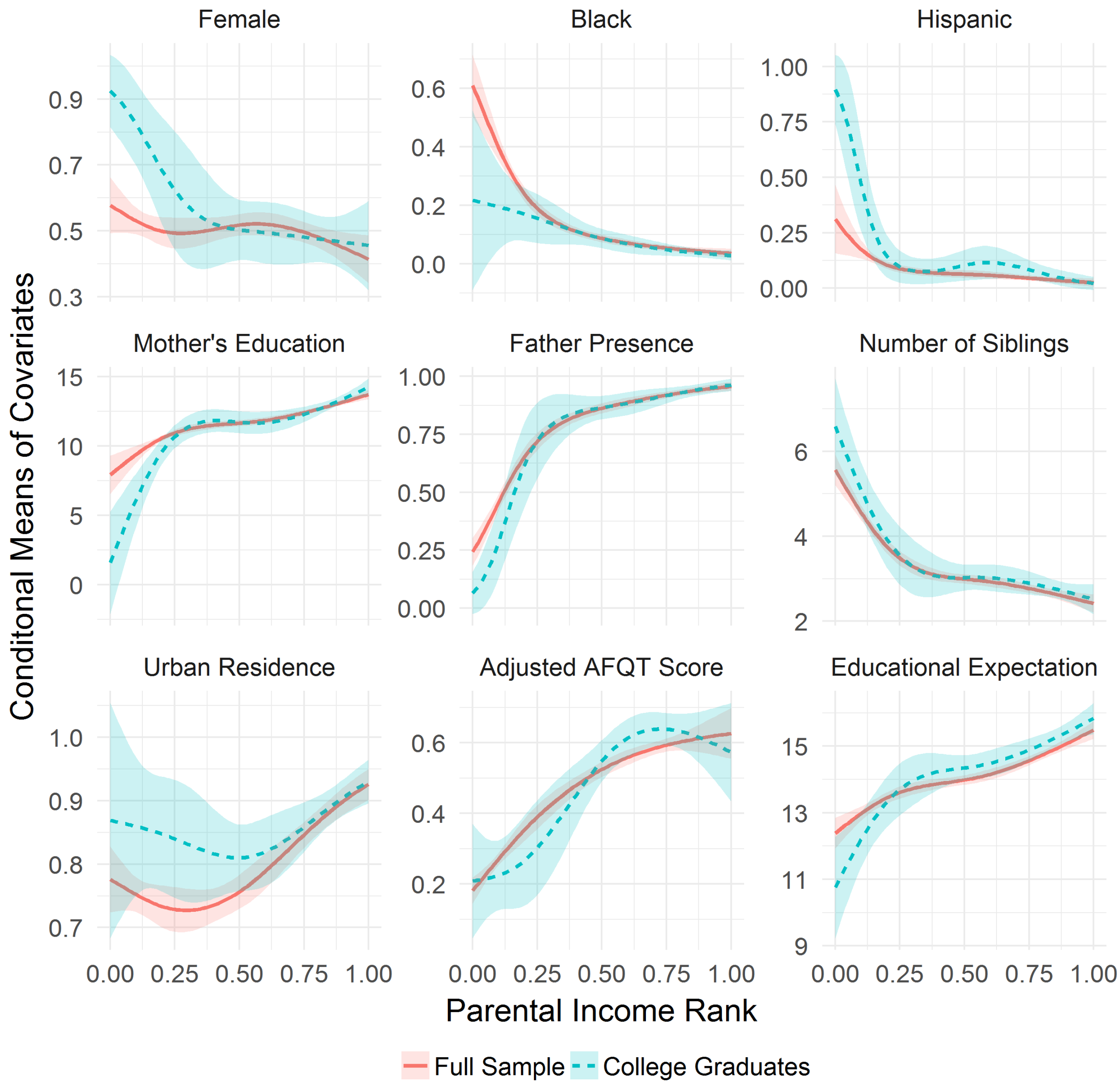

Figure B1: Fitted Conditional Means of Covariates Given Parental Income Rank among the Full Sample and College Graduates, Inverse Probability Reweighted Sample
Note: All conditional means are fitted as a natural cubic spline of parental income rank with three degrees of freedom and adjusted by inverse probability weights. Ribbons represent 95% asymptotic confidence intervals.

the inverse probability weighted sample, the conditional means of the covariates given parental income rank are reasonably balanced between college graduates and the full sample. Table B1 reports the estimated intergenerational rank-rank slopes in both the original sample and the inverse probability reweighted sample. We can see that the estimates of controlled mobility do not differ significantly between college graduates and non-graduates.

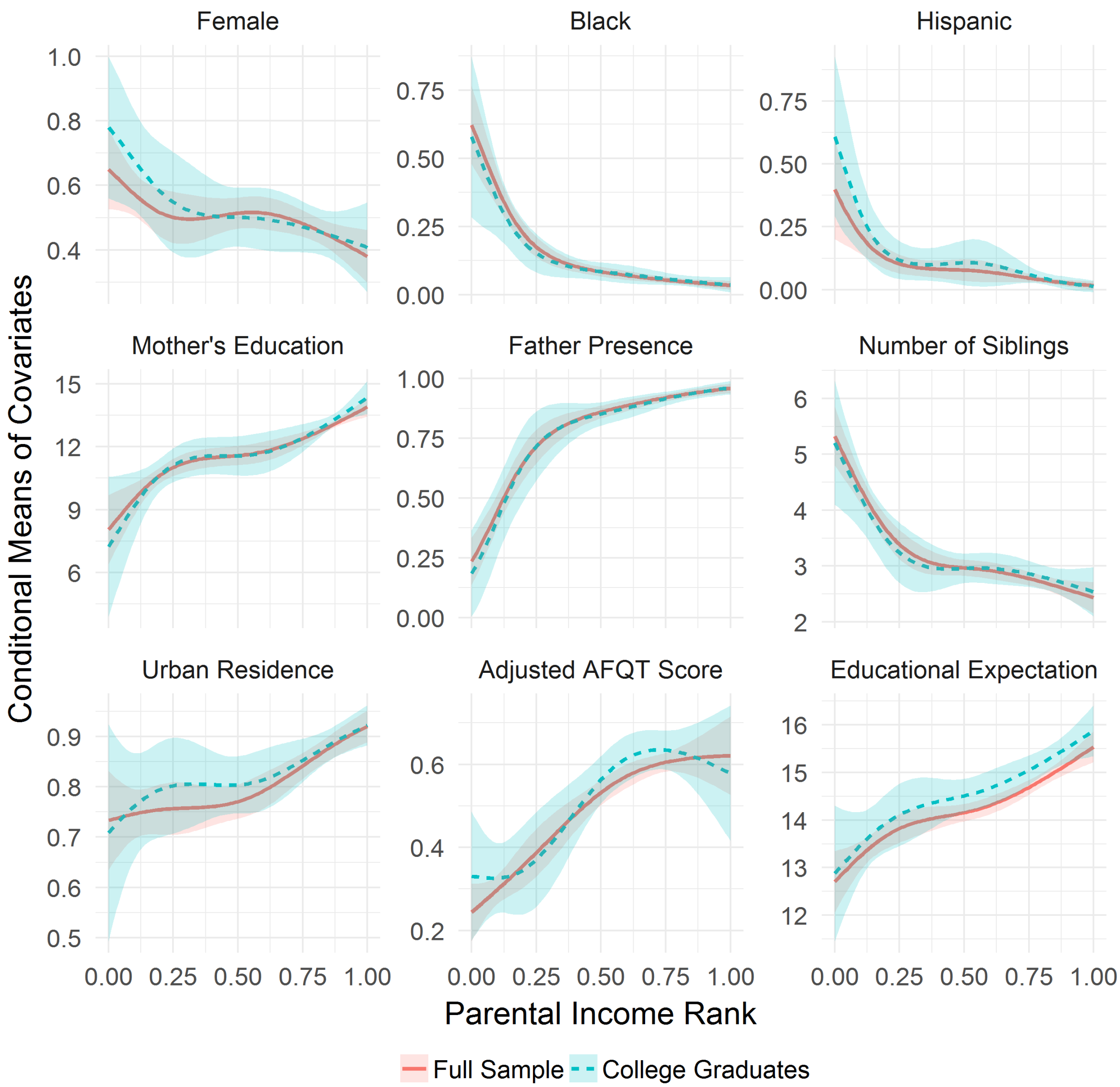

Figure B2: Fitted Conditional Means of Covariates Given Parental Income Rank among the Full Sample and College Graduates, Propensity Score Matched Sample
Note: All conditional means are fitted as a natural cubic spline of parental income rank with three degrees of freedom and adjusted by propensity score matching weights. Ribbons represent 95% asymptotic confidence intervals.

To implement propensity score matching, I first estimate the propensity scores by fitting a logit model of college graduation with both parental income rank and the nine pre-

Table B2: Estimates of Intergenerational Rank-Rank Slope by College Completion Where Propensity Score Matching Is Used to Adjust for Selection.

	Full Sample		Men		Women	
	Conditional Mobility	Controlled Mobility	Conditional Mobility	Controlled Mobility	Conditional Mobility	Controlled Mobility
Intercept	0.297***	0.307***	0.315***	0.324***	0.279***	0.294***
	(0.009)	(0.011)	(0.012)	(0.014)	(0.012)	(0.015)
Parental Income Rank	0.312***	0.313***	0.306***	0.308***	0.315***	0.296***
	(0.017)	(0.026)	(0.023)	(0.028)	(0.024)	(0.035)
College Degree	0.323***	0.171**	0.342***	0.197**	0.311***	0.185*
	(0.024)	(0.057)	(0.034)	(0.076)	(0.033)	(0.077)
Parental Income Rank * College Degree	-0.141***	0.011	-0.157**	-0.007	-0.131**	-0.022
	(0.036)	(0.079)	(0.05)	(0.107)	(0.050)	(0.113)
Sample Size	4,673		2,370		2,303	

Note: †p<.1, *p<.05, **p<.01, ***p<.001 (two-tailed tests). Numbers in parentheses are heteroskedasticity-consistent robust standard errors.

college covariates as predictors (i.e., the same model used for estimating the denominator of equation 2). The estimated propensity scores are then used to construct a matched sample for college graduates. Specifically, I use one-to-one matching with replacement while allowing for ties. That is, for each college graduate, a non-graduate with the closest propensity score is matched. But in cases where multiple non-graduates are sufficiently close to the college graduate (with the absolute difference in propensity score smaller than 10^{-5}), they are all included in the matched dataset with evenly distributed fractional weights. Similarly, we construct a matched sample for non-college graduates. I then combine the two matched samples, collapse multiple records of the same individual onto a single record by adding up her weights, and update each individual's weights as the product of her matching weight and NLSY custom weight. In this final matched sample, the joint distribution of parental income rank and the nine pre-college covariates is *expected* to be the same between college graduates and the general population (assuming correct specification of the propensity score model). Thus the conditional distribution of the nine covariates given parental income rank is also *expected* to be the same between college graduates and the general population. Figure B2 shows that in this matched sample, the conditional means of the covariates given parental income rank are reasonably balanced between college graduates and the full sample. Table B2 reports the estimated intergenerational rank-rank slopes in both the original sample and the propensity score matched sample. We can see that the results are substantively the same as those based on residual balancing (Table 2).

C. Results on Intergenerational Income Elasticity (IGE)

In the main text, we examined income mobility using the intergenerational rank-rank slope. Alternatively, we can use the intergenerational income elasticity (IGE), i.e., the slope parameter from a regression of log adult income on log parental income. To estimate the IGE, I exclude a few respondents with zero or negative incomes. Tables C1 reports the IGE-based results of conditional and controlled mobility. Echoing Table 2, the estimated interaction effect between parental income and college completion is large and negative in the original sample but drops substantially in magnitude in the reweighted sample.

Table C1: Estimates of Intergenerational Income Elasticity by College Completion

	Full Sample		Men		Women	
	Conditional Mobility	Controlled Mobility	Conditional Mobility	Controlled Mobility	Conditional Mobility	Controlled Mobility
Intercept	6.438***	6.322***	6.895***	6.718***	6.025***	5.944***
	(0.220)	(0.225)	(0.318)	(0.319)	(0.298)	(0.310)
Log Parental Income	0.396***	0.412***	0.354***	0.376***	0.433***	0.447***
	(0.022)	(0.022)	(0.032)	(0.032)	(0.030)	(0.031)
College Degree	1.987***	0.068	1.843*	0.015	2.248***	0.396†
	(0.497)	(1.113)	(0.776)	(1.099)	(0.630)	(1.443)
Log Parental Income * College Degree	-0.135**	0.024	-0.118	0.038	-0.163**	-0.012
	(0.048)	(0.105)	(0.075)	(0.105)	(0.062)	(0.137)
Sample Size	4,628		2,344		2,284	

Note: †p<.1, *p<.05, **p<.01, ***p<.001 (two-tailed tests). Numbers in parentheses are heteroskedasticity-consistent robust standard errors. Observations with zero or negative incomes are excluded.

D. Balance Plots for Three Educational Groups

To assess the role of college selectivity, we have coded education as a trichotomous variable: non-college graduates, college graduates who attended only nonselective colleges, and college graduates who attended selective colleges. Figures D1 and D2 show the estimated conditional means of the nine pre-college covariates given parental income rank before and after reweighting, respectively. We can see that the conditional means of the covariates are much more balanced across educational groups in the reweighted sample than in the original sample.

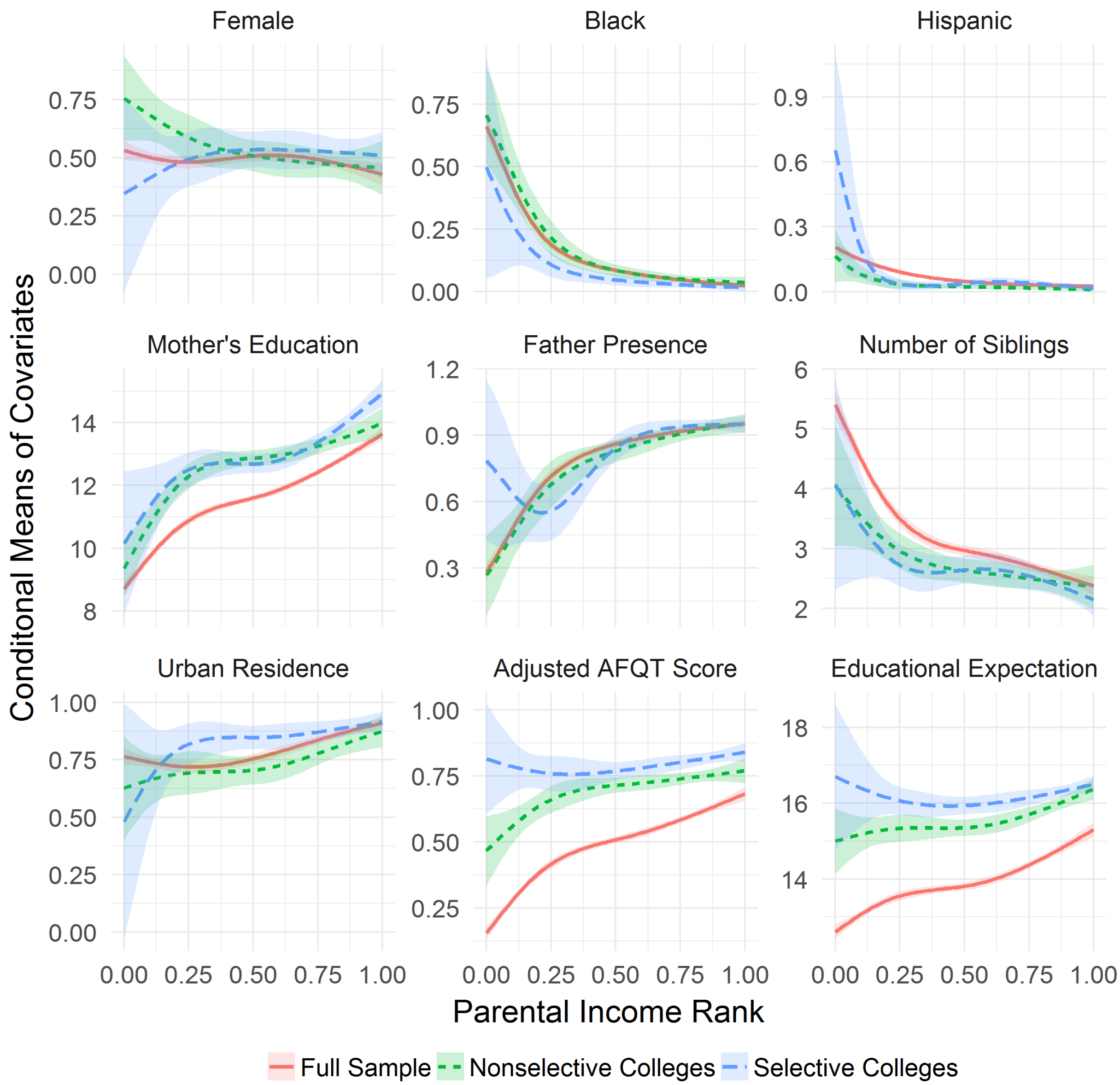

Figure D1: Fitted Conditional Means of Covariates Given Parental Income Rank among the Full Sample, College Graduates from Nonselective Colleges, and College Graduates from Selective Colleges, Original Sample.

Note: All conditional means are fitted as a natural cubic spline of parental income rank with three degrees of freedom and adjusted by NLSY custom weights. Ribbons represent 95% asymptotic confidence intervals.

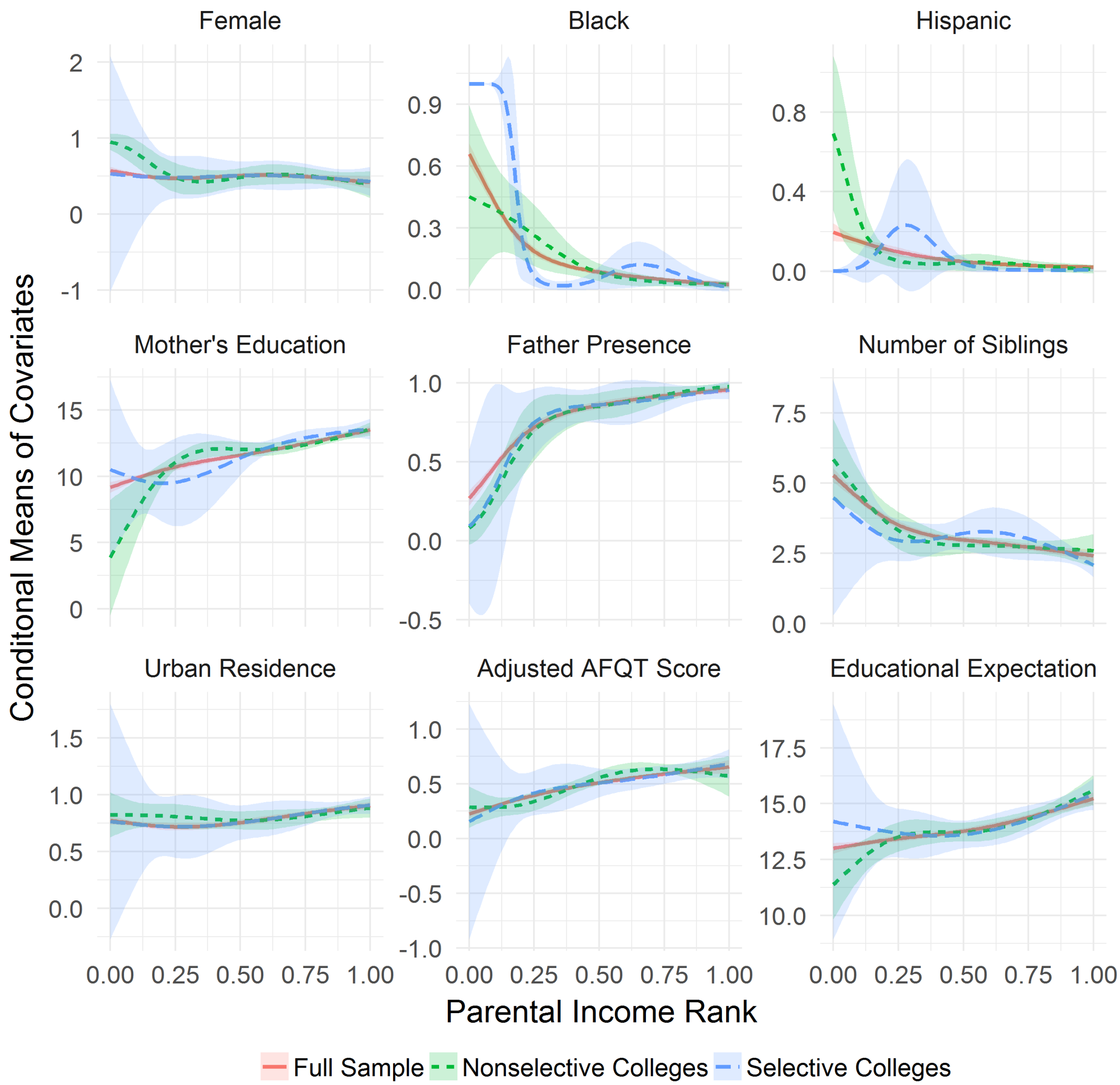

Figure D2: Fitted Conditional Means of Covariates Given Parental Income Rank among the Full Sample, College Graduates from Nonselective Colleges, and College Graduates from Selective Colleges, Reweighted Sample.

Note: All conditional means are fitted as a natural cubic spline of parental income rank with three degrees of freedom and adjusted by residual balancing weights. Ribbons represent 95% asymptotic confidence intervals.

E. Results under Alternative Sample Restriction Criteria

In the main analysis, I restricted my sample to those respondents who were at most 18 years old in 1979. To maintain a relatively large sample size, I retained all respondents who had at least one valid observation of parental family income and one valid observation of adult family income. The results, however, are robust to different sample restriction criteria. Table E1 reports the rank-rank slope estimates of conditional and controlled mobility for the full NLSY79 sample (without age restriction). Table E2 reports the rank-rank slope estimates of conditional and controlled mobility for those respondents who had at least two valid observations for both parental family income and adult family income. Both sets of results indicate that compared with conditional mobility, estimates of controlled mobility are much less distinguishable, both substantively and statistically, between college graduates and non-graduates.

Table E1: Estimates of Intergenerational Rank-Rank Slope by College Completion for the Full NLSY79 Sample.

	Full Sample		Men		Women	
	Conditional Mobility	Controlled Mobility	Conditional Mobility	Controlled Mobility	Conditional Mobility	Controlled Mobility
Intercept	0.288***	0.302***	0.301***	0.312***	0.276***	0.292***
	(0.007)	(0.008)	(0.010)	(0.011)	(0.010)	(0.011)
Parental Income Rank	0.335***	0.343***	0.335***	0.354***	0.327***	0.330***
	(0.014)	(0.016)	(0.020)	(0.021)	(0.021)	(0.023)
College Degree	0.340***	0.186**	0.368***	0.214***	0.319***	0.158†
	(0.020)	(0.064)	(0.030)	(0.063)	(0.027)	(0.082)
Parental Income Rank * College Degree	-0.175***	-0.070	-0.206***	-0.077	-0.146***	-0.047
	(0.031)	(0.084)	(0.045)	(0.082)	(0.043)	(0.108)
Sample Size	6,532		3,214		3,138	

Note: †p<.1, *p<.05, **p<.01, ***p<.001 (two-tailed tests). Numbers in parentheses are heteroskedasticity-consistent robust standard errors.

Table E2: Estimates of Intergenerational Rank-Rank Slope by College Completion, Excluding Respondents with Fewer Than Two Observations in Either Parental Family Income or Adult Family Income.

	Full Sample		Men		Women	
	Conditional Mobility	Controlled Mobility	Conditional Mobility	Controlled Mobility	Conditional Mobility	Controlled Mobility
Intercept	0.295***	0.302***	0.307***	0.314***	0.282***	0.287***
	(0.010)	(0.010)	(0.013)	(0.014)	(0.014)	(0.015)
Parental Income Rank	0.317***	0.331***	0.318***	0.334***	0.314***	0.332***
	(0.019)	(0.021)	(0.027)	(0.030)	(0.028)	(0.030)
College Degree	0.324***	0.184*	0.352***	0.211**	0.304***	0.179†
	(0.027)	(0.078)	(0.039)	(0.068)	(0.038)	(0.106)
Parental Income Rank * College Degree	-0.145***	-0.059	-0.171**	-0.039	-0.127*	-0.088
	(0.041)	(0.102)	(0.058)	(0.091)	(0.058)	(0.136)
Sample Size	3,620		1,855		1,765	

Note: †p<.1, *p<.05, **p<.01, ***p<.001 (two-tailed tests). Numbers in parentheses are heteroskedasticity-consistent robust standard errors.

F. Results under Alternative Definitions of Income

In the main analysis, I defined adult income as the sum of husbands', wives', and other co-residential family members' annual disposable incomes from a variety of sources, including wages and salary, farm and business income, and several government programs such as unemployment compensation. Table F1 reports the rank-rank slope estimates of conditional and controlled mobility where income is restricted to include only husbands' and wives' pre-tax earnings, i.e., wages and salary plus farm and business income.[1] Table F2 reports the rank-rank slope estimates of conditional and controlled mobility where income is restricted to include only the respondent's pre-tax earnings. We can see that our main results are highly consistent under alternative definitions of income.

Table F1: Estimates of Intergenerational Rank-Rank Slope by College Completion where Adult Income Is Defined as Total Family Earnings.

	Full Sample		Men		Women	
	Conditional Mobility	Controlled Mobility	Conditional Mobility	Controlled Mobility	Conditional Mobility	Controlled Mobility
Intercept	0.303***	0.309***	0.310***	0.315***	0.296***	0.301***
	(0.008)	(0.009)	(0.011)	(0.012)	(0.012)	(0.013)
Parental Income Rank	0.301***	0.324***	0.298***	0.324***	0.302***	0.327***
	(0.017)	(0.019)	(0.023)	(0.026)	(0.025)	(0.027)
College Degree	0.326***	0.149*	0.367***	0.242**	0.295***	0.105
	(0.025)	(0.067)	(0.036)	(0.084)	(0.034)	(0.083)
Parental Income Rank * College Degree	-0.145***	-0.013	-0.191***	-0.115	-0.109*	0.033
	(0.038)	(0.090)	(0.054)	(0.123)	(0.053)	(0.111)
Sample Size	4,673		2,370		2,303	

Note: †p<.1, *p<.05, **p<.01, ***p<.001 (two-tailed tests). Numbers in parentheses are heteroskedasticity-consistent robust standard errors.

[1] Technically speaking, farm and business income consists of both labor income and capital income, although these two components cannot be separated from the NLSY data.

Table F2: Estimates of Intergenerational Rank-Rank Slope by College Completion where Adult Income Is Defined as Total Individual Earnings.

	Full Sample		Men		Women	
	Conditional Mobility	Controlled Mobility	Conditional Mobility	Controlled Mobility	Conditional Mobility	Controlled Mobility
Intercept	0.346*** (0.008)	0.354*** (0.009)	0.405*** (0.012)	0.412*** (0.012)	0.293*** (0.010)	0.300*** (0.011)
Parental Income Rank	0.233*** (0.018)	0.242*** (0.020)	0.296*** (0.022)	0.315*** (0.025)	0.142*** (0.023)	0.153*** (0.026)
College Degree	0.282*** (0.029)	0.161* (0.082)	0.351*** (0.032)	0.247*** (0.058)	0.241*** (0.039)	0.134 (0.099)
Parental Income Rank * College Degree	-0.149** (0.046)	-0.011 (0.103)	-0.210*** (0.048)	-0.122 (0.083)	-0.103 (0.063)	0.022 (0.124)
Sample Size	4,673		2,370		2,303	

Note: †p<.1, *p<.05, **p<.01, ***p<.001 (two-tailed tests). Numbers in parentheses are heteroskedasticity-consistent robust standard errors.

G. Results under Alternative Age Cutoffs for Defining College Graduates

In the main analysis, I coded college graduates as those who had received a bachelor's degree by age 30. Tables G1 and G2 report the rank-rank slope estimates of conditional and controlled mobility where the age cutoff for defining college graduates is 25 and 35, respectively. We can see that our main results are fairly robust under alternative age cutoffs for defining college graduates.

Table G1: Estimates of Intergenerational Rank-Rank Slope by College Completion where the Age Cutoff for College Graduation is 25.

	Full Sample		Men		Women	
	Conditional Mobility	Controlled Mobility	Conditional Mobility	Controlled Mobility	Conditional Mobility	Controlled Mobility
Intercept	0.301***	0.306***	0.316***	0.322***	0.287***	0.289***
	(0.008)	(0.009)	(0.012)	(0.012)	(0.012)	(0.013)
Parental Income Rank	0.333***	0.356***	0.328***	0.348***	0.335***	0.363***
	(0.017)	(0.017)	(0.023)	(0.024)	(0.024)	(0.025)
College Degree	0.336***	0.198*	0.373***	0.229***	0.302***	0.202†
	(0.027)	(0.080)	(0.036)	(0.062)	(0.038)	(0.110)
Parental Income Rank * College Degree	-0.175***	-0.086	-0.200***	-0.080	-0.151**	-0.126
	(0.039)	(0.100)	(0.054)	(0.082)	(0.057)	(0.137)
Sample Size	4,673		2,370		2,303	

Note: †p<.1, *p<.05, **p<.01, ***p<.001 (two-tailed tests). Numbers in parentheses are heteroskedasticity-consistent robust standard errors.

Table G2: Estimates of Intergenerational Rank-Rank Slope by College Completion where the Age Cutoff for College Graduation is 35.

	Full Sample		Men		Women	
	Conditional Mobility	Controlled Mobility	Conditional Mobility	Controlled Mobility	Conditional Mobility	Controlled Mobility
Intercept	0.292***	0.303***	0.310***	0.320***	0.275***	0.285***
	(0.008)	(0.009)	(0.012)	(0.013)	(0.012)	(0.013)
Parental Income Rank	0.312***	0.326***	0.306***	0.324***	0.314***	0.33***
	(0.017)	(0.019)	(0.023)	(0.027)	(0.024)	(0.027)
College Degree	0.337***	0.163*	0.353***	0.216***	0.327***	0.151†
	(0.023)	(0.067)	(0.032)	(0.052)	(0.032)	(0.085)
Parental Income Rank * College Degree	-0.154***	-0.014	-0.166***	-0.047	-0.146**	-0.018
	(0.035)	(0.086)	(0.048)	(0.069)	(0.049)	(0.110)
Sample Size	4,673		2,370		2,303	

Note: †p<.1, *p<.05, **p<.01, ***p<.001 (two-tailed tests). Numbers in parentheses are heteroskedasticity-consistent robust standard errors.

H. The Effects of College Expansion on Intergenerational Mobility: A Simulation Study

In this study, we have found that the "college premium" in intergenerational income mobility is largely driven by selection processes rather than an equalizing effect of a bachelor's degree. We have thus argued that simply expanding the pool of college graduates is unlikely to be effective in promoting intergenerational mobility. Quite the contrary, given the strong influence of parental income on educational attainment and the fact that college completion is still far from universal among high-income children in the US (Bloome et al. 2018), a mechanical (non-preferential) expansion of higher education might attract disproportionately more students from economically advantaged backgrounds. Thus, in the short run, intergenerational income mobility might even increase as a result of college expansion. Below, I conduct a simulation study to investigate how intergenerational income mobility would change as a result of gradual expansions of college education.

The key idea of this simulation is to change only the proportion of college graduates while holding all other aspects of the status attainment process constant. To set up the simulation parameters in a way that mimics empirical reality, I first fit two generalized additive models (GAM), one modeling college graduation as a function of both parental income rank and the nine pre-college covariates, and the other modeling log adult family income as a function of parental income rank, education (a binary indicator for college graduates), and the nine pre-college covariates. In the meanwhile, I create a hypothetical population by replicating the NLSY79 sample 100 times, keeping only information on parental income rank and the nine pre-college covariates. To examine the effect of a "pure" college expansion, I construct a continuum of scenarios by gradually changing the intercept of the college graduation model (the first GAM) from very negative to very positive values, while keeping all other parameters of the estimated GAMs unchanged. Then, in each of these scenarios, I generate model-based draws of college graduation status and adult family income for each individual of the hypothetical population, and then calculate the intergenerational rank-rank slopes for college graduates, non-graduates, and the

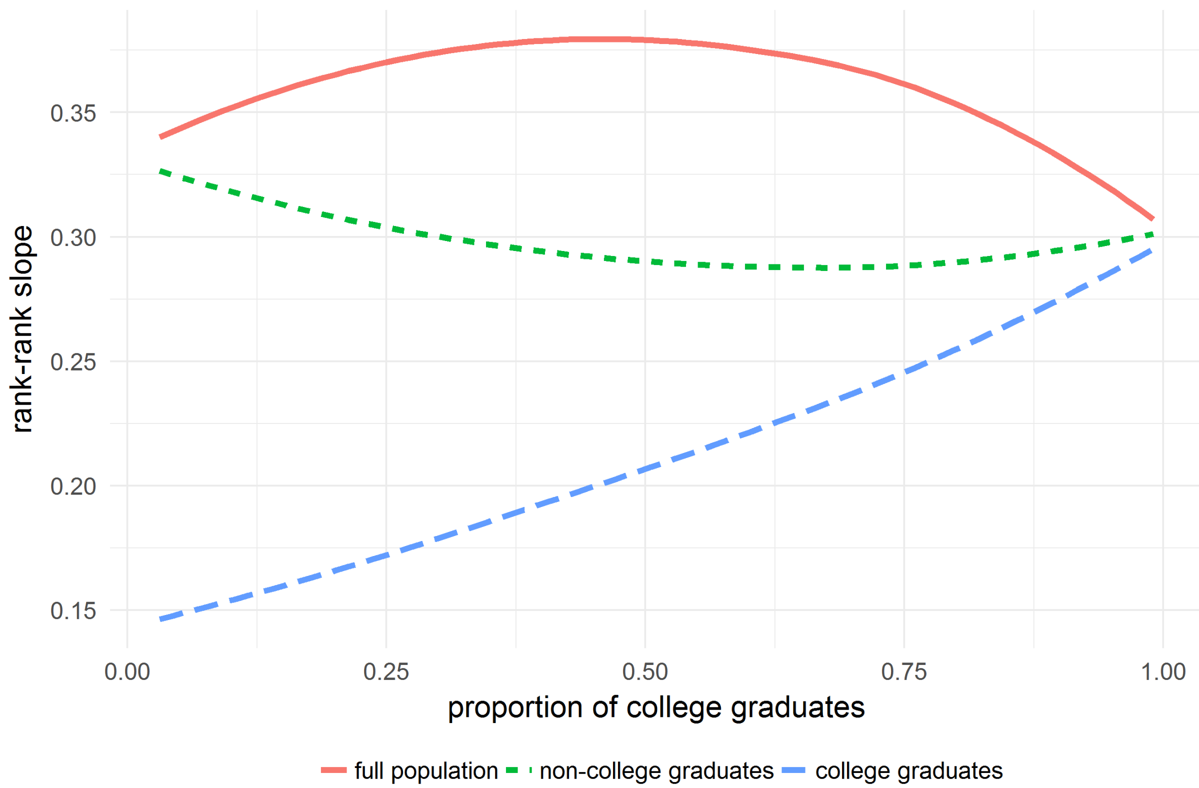

Figure H1: Intergenerational Rank-Rank Slopes under Varying Proportions of College Graduates in Simulated Data.

entire population.

The results are summarized in Figure H1. We can see that when the proportion of college graduates in the population is relatively low, the intergenerational rank-rank slope is much lower among college graduates (long dashed line) than among non-graduates (short dashed line), indicating a "college premium" in intergenerational income mobility. Yet, as the proportion of college graduates increases, the college premium diminishes. More important, the rank-rank slope in the general population (solid line) exhibits a non-monotone trend. As the pool of college graduates expands, the overall rank-rank slope increases at first, peaks when about half of the population are college-educated, and declines thereafter. Thus, considering that the proportion of college graduates in recent US cohorts is still below 50% (Bauman 2016), an incremental expansion of postsecondary education (per se) is unlikely to have a considerable effect on intergenerational income mobility.

References

Bauman, Kurt. 2016. "College Completion by Cohort, Age and Gender, 1967 to 2015." Technical report, U.S. Census Bureau. URL: `https://www.census.gov/content/dam/Census/library/working-papers/2016/demo/SEHSD-WP2016-04.pdf`.

Bloome, Deirdre, Shauna Dyer, and Xiang Zhou. 2018. "Educational Inequality, Educational Expansion, and Intergenerational Income Persistence in the United States." *American Sociological Review* 83:1215–1253.

Hainmueller, Jens. 2012. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis* 20:25–46.

Imbens, Guido W and Donald B Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Lefebvre, Genevieve, Joseph AC Delaney, and Robert W Platt. 2008. "Impact of Misspecification of the Treatment Model on Estimates from a Marginal Structural Model." *Statistics in Medicine* 27:3629–3642.

Lunceford, Jared K and Marie Davidian. 2004. "Stratification and Weighting via the Propensity score in Estimation of Causal Treatment Effects: A Comparative Study." *Statistics in Medicine* 23:2937–2960.

Robins, James M, Miguel Angel Hernan, and Babette Brumback. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." *Epidemiology* 11:550–560.

Zhou, Xiang and Geoffrey T Wodtke. 2018. "Residual Balancing Weights for Marginal Structural Models: with Application to Analyses of Time-varying Treatments and Causal Mediation." *arXiv preprint arXiv:1807.10869* .