Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

Accounting for Individual-Specific Heterogeneity in Intergenerational Income Mobility

---

CAMA Working Paper 18/2024
February 2024 (Revised in December 2024)

Yoosoon Chang
Indiana University
Centre for Applied Macroeconomic Analysis, ANU

Steven N. Durlauf
University of Chicago

Bo Hu
Indiana University

Joon Y. Park
Indiana University

Abstract

This paper proposes a fully nonparametric model to investigate the dynamics of intergenerational income mobility for discrete outcomes. In our model, an individual's income class probabilities depend on parental income in a manner that accommodates nonlinearities and interactions among various individual and parental characteristics, including race, education, and parental age at childbearing. Consequently, we offer a generalization of Markov chain mobility models. We employ kernel techniques from machine learning and further regularization for estimating this highly flexible model. Utilizing data from the Panel Study of Income Dynamics (PSID), we find that race and parental education play significant roles in determining the influence of parental income on children's economic prospects.

Keywords

intergenerational income mobility, ordered multinomial probability model, nonparametric estimation, heterogeneous treatment effects, reproducing kernel Hilbert space, effects of parental education

JEL Classification

C10, C50, D30, I24

Address for correspondence:

(E) cama.admin@anu.edu.au

ISSN 2206-0332

Accounting for Individual-Specific Heterogeneity in Intergenerational Income Mobility *

Yoosoon Chang† Steven N. Durlauf‡ Bo Hu§ Joon Y. Park¶

December 30, 2024

Abstract

This paper proposes a fully nonparametric model to investigate the dynamics of intergenerational income mobility for discrete outcomes. In our model, an individual's income class probabilities depend on parental income in a manner that accommodates nonlinearities and interactions among various individual and parental characteristics, including race, education, and parental age at childbearing. Consequently, we offer a generalization of Markov chain mobility models. We employ kernel techniques from machine learning and further regularization for estimating this highly flexible model. Utilizing data from the Panel Study of Income Dynamics (PSID), we find that race and parental education play significant roles in determining the influence of parental income on children's economic prospects.

JEL Classification: C10, C50, D30, I24

Key words and phrases: intergenerational income mobility, ordered multinomial probability model, nonparametric estimation, heterogeneous treatment effects, reproducing kernel Hilbert space, effects of parental education

---

*We would like to thank Kristina Butaeva, Ruli Xiao, Guo Yan, Liyuan Yang, and participants in the *Conference on Inequality and Mobility*, co-organized by the Stone Center for Research on Wealth Inequality and Mobility (SCRWIM) at the University of Chicago and the Institute of Economic Research at Seoul National University, as well as the *Conference on Intergenerational Mobility* jointly organized by SCRWIM at University of Chicago and the Stone Center on Inequality Dynamics at University of Michigan, for their comments. This paper is part of the research activities at the Centre for Applied Macroeconomics and Commodity Prices (CAMP) at the BI Norwegian Business School.

†Indiana University and CAMA, yoosoon@iu.edu

‡University of Chicago, sdurlauf@uchicago.edu

§Indiana University, hu21@iu.edu

¶Indiana University, joon@iu.edu

1 Introduction

This paper proposes an approach to studying intergenerational income mobility based on a generalization of Markov chain mobility models. We group households into different income categories and assume that the probabilities of an individual from different family backgrounds being in different income categories are determined by various characteristics of a person's parents. Our generalization allows transition probabilities between parent and child to depend nonparametrically on parental characteristics. This type of dependence is natural from the perspective of theories of intergenerational mobility, whether models in which parental income determines investment, models in which parental income determines the neighborhoods and/or schools of children, models in which discrimination creates persistent Black/White mobility differences, or models in which parental skills as determined by education or experience affect the productivity of education on children's human capital formation; see Durlauf et al. (2022) for the elaboration of the many paths by which Markov transition probabilities will differ across families. This framework enables us to study the joint distribution of parental-child income pairs and therefore income mobility dynamics at the aggregate level.

Previous research has extensively explored heterogeneity in transition processes, with race being a standard dimension of analysis; see Duncan (1968) and Hout (1984) for older classic studies and Bhattacharya and Mazumder (2011) and Bloome (2014) for more recent contributions. Our aim is to provide novel tools that capture this heterogeneity in richer ways than previous studies. To achieve this, we develop a fully nonparametric ordered multinomial probability model that can accommodate highly general nonlinear relationships between parental income status and the income class into which children move. In addition, we incorporate factors such as race, parental education, and parental age at childbearing to influence the conditional probability structure linking parental and offspring income statuses without relying on functional form assumptions linking offspring income and parental characteristics. There are other approaches to studying heterogeneous effects. For example, Brand and Xie (2010) consider group-specific treatment effects, Abadie et al. (2002) introduces quantile treatment effects, and one may estimate the joint distribution of the factor and outcome using copulas as in Chetty et al. (2017). These approaches usually study levels instead of class probabilities, and they assume specific functional forms for the interaction structure.

The flexibility of our model presents challenges in terms of estimation. It has long been understood that fully nonparametric estimation of a nonlinear model can be exceedingly difficult, especially when dealing with a large number of factors and/or a sizable sample size. Even in conventional probit or logit models with linear index functions, estimation can become computationally daunting when handling extensive datasets or numerous regressors. To overcome this issue, we employ kernel methods from the machine learning literature coupled with further regularization through principal component analysis (PCA). These tools have become increasingly prevalent for addressing high-dimensional problems.[1] By leveraging these methods, we introduce a new approach to estimate our fully nonparametric multinomial choice model, which is robust in environments with large samples and/or a large number of covariates, thereby circumventing the curse of dimensionality while maintaining computational efficiency.

We illustrate our methods by applying our multinomial probability model to the Panel Study of Income Dynamics (PSID) data to examine how gender, race, parental education, parental age at childbirth, and parental income status interact to influence offspring income status. Our analysis reveals significant racial disparities, with Black individuals more likely to fall into the low-income category and less likely to belong to the middle- and high-income categories, particularly among those raised in middle-income families. We also find that parental college education substantially reduces the likelihood of a child being in the low-income category and increases the chances of belonging to the middle- and high-income categories. This positive effect of parental education is particularly pronounced for individuals with middle-to-high-income parents and those born when their parents are in their late twenties to mid-thirties, maximizing the predictive probability of a child attaining high-income status. Collectively, race, parental education, and parental age at childbirth can influence the probabilities of low-income status for children by about 20 percent, given a certain parental income level. This provides compelling evidence of the ways in which heterogeneity in downward mobility can occur for middle-income families.

Relationships between family-level characteristics and future incomes have of course been extensively studied. Dahl and Lochner (2012) demonstrate nonlinearities in the intergenerational income transmission process. Maralani (2013) and

[1]From the methodological point of view, our paper is a multinomial extension of Yan (2023), which develops a framework to analyze fully nonparametric large dimensional binary choice models.

Bacic and Zheng (2024) demonstrate racial differences in intergenerational educational mobility. Lopoo and DeLeire (2014) and Bloome (2017) show how a family structure affects intergenerational income effects. Our overall findings are corroborative of past work. Our contribution is to demonstrate the robustness of previous claims when the analysis allows for a much more general structure than past studies. In particular, we are able to show that particular family-level factors retain their predictive power when simultaneously considered with other factors and are able to do this by allowing nonlinear relationships between each factor and offspring income category probabilities. Beyond this, our methods give additional substantive insights. In particular, we show that there are stark differences between the children of non-Black families with college-educated parents, and born when their parents are around 30 versus children who are Black, born to parents at the age of 18, and without college degrees. Thus one set of background conditions makes it hard for some children to achieve socioeconomic success in income and education while another makes it hard to fail.

The paper is organized as follows: We begin with a brief motivation for our work in Section 2. Section 3 introduces our ordered multinomial choice model for income class probabilities and discusses its estimation and inference. Section 4 describes the Panel Study of Income Dynamics data we use. Section 5 applies our methods to explore the effects of various factors on the relationship between parental and offspring's income statuses. Finally, Section 6 concludes the paper. Technical details and more robustness checks are provided in the four Appendices that follow.

2 Beyond Linearity in Intergenerational Mobility Analysis

The workhorse model to study intergenerational income mobility is

$$\log(y_c) = \alpha + \beta \log(y_p) + \varepsilon$$

where y_c and y_p are specific measures of the child's income and the parental income, respectively, and ε represents individual-specific heterogeneity. The parameter β is the intergenerational elasticity of income and has become the primary measure of the persistence of income across generations.

As such, this workhorse model has nothing to say about the evolution of intergenerational persistence, since β is a constant. Researchers have therefore augmented this model to include additional factors. A frequently used regression model takes the form of

$$\log(y_c) = \alpha + \beta \log(y_p) + \gamma' s + \varepsilon$$

where s is a vector of factors beyond parental income that are believed to shape a child's income.

Although the augmented model enables one to study the effects of factors beyond parental income on intergenerational mobility, it is still very restrictive since it does not allow interactions between different factors in determining the income level of the child. As such, it preserves an implicit dichotomy between the measure of intergenerational mobility, β, and other mechanisms. Social science theory does not justify this independence. For example, it could be the case that the effect of parental income on children's future income is affected by discrimination or by parental education. This has led to a literature that allows β to differ by categories such as race. By implication, products of variables are usually taken to capture the interactions of different determinants of offspring outcomes. However, this is not an entirely satisfactory solution, since it amounts to a second-order Taylor series approximation of the interactions of different variables, and there is no theoretical basis for thinking such an approximation will be particularly accurate. And of course, this observation applies to efforts to introduce nonlinearities in the effects of parental income based on polynomial generalizations of the linear model.

Our objective is to propose a framework that can accommodate rich interactions and nonlinearities. We propose a fully nonparametric model to link these factors to the probabilities of a child belonging to different relative income classes. Unlike the IGE model which focuses on levels of income, we consider probabilities that link the income classes of parents and children. We choose this outcome variable for several reasons. First, our model permits a natural integration of interactions by making income class probabilities functions of various factors. Second, income categories such as the middle class hold a distinct substantive interest from absolute income levels. Third, many of the publicly available income data contain left and/or right-censored observations and might contain zero/negative income figures. Estimating an IGE with censored data might lead to biased estimates, and taking logs with zero and/or negative values could be problematic, even with some of the usual transformation techniques such as adding one before taking logs (Chen and Roth, 2024). Our approach, by analyzing income classes instead of income levels, remains robust in the presence of such data issues.

In the next section, we propose a fully nonparametric multinomial choice model that can be used to study the link between various factors and an individual's probabilities of membership into different income classes.

3 Methodology

3.1 Multinomial Model for Income Class Probabilities

The evolution of income distributions over time is evident. A pertinent inquiry arises: what factors propel these changes, and how exactly do they impact income distribution dynamics? To address this, we propose a nonparametric ordered multinomial choice model.

To be specific, let $j = 1, \ldots, m$ denote the m income classes. In our study, we shall set $m = 3$, and let $j = 1, 2$ and 3 represent the low-, middle- and high-income classes, respectively. We use subscript $i = 1, 2, \ldots, n$ to index individuals in our sample and use $\pi_j(x)$ to denote the probability of belonging to class j for the individual with covariates x. We shall call these probabilities the income class probabilities hereafter. Evidently, $\sum_{j=1}^{m} \pi_j(x_i) = 1$ for all $i = 1, \ldots, n$, indicating that each individual's probabilities across all income classes sum up to one.

Individuals' characteristics (x_i) are related to their income class probabilities by the functions $\pi_j(\cdot)$ in the form of an ordered multinomial choice model

$$\pi_j(x) = \mathbb{P}\left\{\tau_{j-1} < y_i^* \leq \tau_j \middle| x_i = x\right\}$$

for $j = 1, \ldots, m$, with the convention $\tau_0 = -\infty$ and $\tau_m = \infty$, where y_i^* is a latent variable that represents the unobservable permanent income of the individual i, which depends on the individual covariates x_i, and $\tau_1, \ldots, \tau_{m-1}$ are constant income thresholds that determine the categories of permanent income. For convenience, from now on we shall simply call y_i^* the permanent income. We set the permanent income (y_i^*) to be determined by the covariates (x_i) through

$$y_i^* = g(x_i) + u_i, \tag{1}$$

where g is a nonparametric function to be estimated, and (u_i) is the random component that represents the heterogeneity in permanent income not captured by the covariates (x_i). We shall estimate the distribution of the random component nonparametrically.

Our framework offers a high level of flexibility and generality. We depart from the usual linear setting by allowing for a general nonlinear form of g, taking values in a sufficiently large function space. The function space employed in our analysis allows for a precise approximation of any continuous function over a compact subset of its domain. This departure from linearity is not solely about freedom in functional forms; rather, it empowers us to explore the heterogeneous impacts of factors on income distribution, and thus intergenerational mobility. In addition, it facilitates the exploration of intricate interactions among various factors that influence income distributions and intergenerational mobility, far beyond those allowed in conventional linear discrete choice models. The generalities of our approach will be explained further in Section 3.2. Moreover, unlike parametric models such as logit or probit, which assume a Gaussian or logistic distribution for the random component (u_i), our model does not confine the random component to any specific distribution family. This grants us greater adaptability in mirroring real income distributions, for example, allowing the presence of fat tails in the income distribution.

To identify the effects of various factors in our discrete choice model, we may either not impose any level restriction on the function g and set one of the parameters in $\tau = (\tau_1, \tau_2, \ldots, \tau_{m-1})$ at a fixed number, say $\tau_1 = 0$, or we restrict the level of g and allow all the parameters in τ to vary freely. In the paper, we set

$$\inf_{x \in D} g(x) = 0,$$

where D is the support of (x_i), which imposes a level restriction on the function g, so that all the parameters in $\tau = (\tau_1, \tau_2, \ldots, \tau_{m-1})$ are identified without any further restriction.

If the random component (u_i) has an invertible cumulative distribution function (CDF) F, it follows that

$$\tau_j - g(x) = F^{-1}\left(\sum_{k=1}^{j} \pi_k(x)\right)$$

for $j = 1, \ldots, m-1$. This implies that, for (π_j) given, parametric models such as logit and probit specifying F as the cumulative distribution function (CDF) of standard Gaussian and logistic distribution, respectively, impose unintended and uninterpretable restrictions on g whenever $m > 2$. This problem does not arise in our model, where we allow the distribution of the random component to be fully nonparametric.

We need to introduce an appropriate identification condition to separately identify the unknown function g in the systematic component and the distribution of the random component.[2] In this paper, however, they are not separately identified, since our analysis will be focused only on various choice probabilities. We leave for our future work the structural analysis based on the function g in the systematic component, which is identified by an appropriate identifying restriction.

3.2 Heterogeneous Effects of Factors

The study of income intergenerational mobility is based on the belief that the income status of parents is linked to the adult income status of their children. One intriguing quantitative question is: if one family has a higher income than another by a certain margin, how does this difference affect the likelihood of their offspring belonging to a specific income class in their adulthood?

While all multinomial choice models can offer insights into this question, their efficacy varies. To articulate this more formally, let $\pi_j(x)$ represent the probability of an offspring's income falling within class j, where x denotes the logged parental income—the only factor considered at present for illustration. The partial effect $\partial \pi_j(x)/\partial x$ serves to answer our question by quantifying the increased likelihood of an offspring being in income class j if their parents' income were increased by 1% from level x. This partial effect, contingent upon the functional form of π_j, is potentially heterogeneous across families with different parental income levels. If we employ a linear probability model as $\pi_j(x) = x\beta$, the partial effect implied is β, which is identical across all families with different parental income levels. If we employ the ordered probit or logit model with a linear g function, the partial effect is given by

$$\frac{\partial \pi_j(x)}{\partial x} = \big[f(\tau_{j-1} - x\beta) - f(\tau_j - x\beta)\big]\beta,$$

where f is the probability density function (PDF) of the standard normal distribution or the logistic distribution. This partial effect, although heterogeneous in x,

[2]The reader is referred to, e.g., Yan (2023) for a detailed discussion on the required identification condition for discrete choice models.

depends heavily on the shape of the PDF f under consideration. It could be the case that partial effects as functions of x with certain shapes cannot be generated from the probit or logit model. In contrast, our approach gives a partial effect

$$\frac{\partial \pi_j(x)}{\partial x} = [f(\tau_{j-1} - g(x)) - f(\tau_j - g(x)]\frac{\partial g(x)}{\partial x}.$$

By allowing for flexible forms of f and g, we are able to generate heterogeneous partial effects with no restrictions on their shapes if it is viewed as a function of the given covariates x.

Usually, the covariates consist of multiple factors, denoted as $x_i = (z_i, w_i)'$, where z_i is the factor whose heterogeneous impact is of primary interest, and w_i consists of all other factors considered under our study. The conditional average partial effect (CAPE) of z_i on income class probabilities π_j may be formally defined as

$$CAPE_j(z) = \mathbb{E}\left[\frac{\partial \pi_j(z_i, w_i)}{\partial z_i}\middle| z_i = z\right],$$

evaluated at a particular point z. As we vary the evaluation point z, we get the conditional average partial effect as a function of z. Once we obtain an estimator $\hat{\pi}_j(x)$ for $\pi_j(x)$, we may estimate the heterogeneous average partial effect by

$$\widehat{CAPE}_j(z) = \frac{1}{n}\sum_{i=1}^{n} \rho_i \frac{\partial \hat{\pi}_j(z, w_i)}{\partial z} K_h(z - z_i),$$

where (ρ_i) are the survey weights,[3] and $K_h(\cdot) = (1/h)K(\cdot/h)$ is defined with a kernel function K and bandwidth parameter $h > 0$. The kernel function is introduced here to take the local average of $\partial \hat{\pi}_j(z, w_i)/\partial z$ in a neighborhood of any given z. The standard normal density function is commonly used for the kernel function in this context.[4]

The same idea can be applied to study the heterogeneous treatment effect of certain treatments. For instance, consider a treatment such as a college degree for the parents. It is expected that there exists a disparity in the probability of belonging to

[3]These weights are provided by our data set and used in our empirical study to adjust for sample selection and non-random attrition, as will be explained later in Section 4.

[4]However, the uniform kernel, which is given by $K(z) = 1\{|z| \leq 1/2\}$ and $K_h(z) = (1/h)\{|z| \leq h/2\}$, makes it more clear what the kernel function does here. If it is used, we take the local average of $\partial \hat{\pi}_j(z, w_i)/\partial z$ to estimate $\widehat{CAPE}_j(z)$ over the values of (z_i) such that $z - h/2 \leq z_i \leq z + h/2$ for a small value of h.

a specific income class between children whose parents have or have not obtained a college degree. Moreover, it is plausible that such discrepancy in probabilities might exhibit variations among children raised in families with diverse parental income levels. Exploring these variations can offer insights into how parental college education and other family background factors can interact with each other to determine mobility.

To conduct such an analysis, we first partition our covariates into $x_i = (z_i, d_i, w_i)$. Here, z_i represents the contingency variable under investigation (in our example, logged parental income), d_i is the treatment variable (1 if at least one of the parents has a college degree, and 0 otherwise), and w_i includes all other factors considered in our study. We then compute the conditional average treatment effect (CATE) in the probability gap given a particular parental income level z, formally defined as

$$CATE_j(z) = \mathbb{E}\left[\pi_j(z_i, 1, w_i) - \pi_j(z_i, 0, w_i)\middle| z_i = z\right].$$

With a properly estimated income class probability function $\hat{\pi}_j(x)$, we may estimate the heterogeneous average treatment effect by

$$\widehat{CATE}_j(z) = \frac{1}{n}\sum_{i=1}^{n} \rho_i[\hat{\pi}_j(z, 1, w_i) - \hat{\pi}_j(z, 0, w_i)]K_h(z - z_i), \tag{2}$$

where we use a kernel function again for local averaging over (z_i) around a given z.

To sum up, our general and flexible framework enables us to design analytical tools to capture complex interactions among the factors without imposing functional forms or predefined interaction terms commonly employed in traditional regression techniques. Moreover, as will be shown later, machine learning techniques and tools empower us to estimate and conduct statistical inference in a framework as general and flexible as ours, even when we face a large number of potential factors and a large sample size.

3.3 Maximum Likelihood Estimation

The nonparametric function g in the systematic component, the PDF f of the random component, and the threshold values τ, $\tau = (\tau_1, \ldots, \tau_{m-1})'$, can be jointly estimated by maximum likelihood estimation for our nonparametric ordered multinomial choice model. We define the maximum likelihood estimators $\hat{g}, \hat{f}$ and $\hat{\tau}$,

$\hat{\tau} = (\hat{\tau}_1, \ldots, \hat{\tau}_{m-1})'$, for g, f and τ by

$$\left(\hat{g}, \hat{f}, (\hat{\tau}_j)_{j=1}^{m-1}\right) = \underset{\substack{g \in \mathcal{G}, f \in \mathcal{F}, \\ \tau \in \mathbb{R}^{m-1}}}{\arg\max} \sum_{i=1}^{n} \rho_i \ell(y_i, x_i, \theta), \tag{3}$$

where θ contains the parameters (g, f, τ), ρ_i is the survey weight for the i-th observation introduced earlier, the log-likelihood function ℓ is given by

$$\ell(y_i, x_i, \theta) = \sum_{j=1}^{m} 1\{y_i = j\} \left[F(\tau_j - g(x_i)) - F(\tau_{j-1} - g(x_i))\right] \tag{4}$$

with the convention $\tau_0 = -\infty$ and $\tau_m = \infty$, and F is the CDF of the distribution given by f. In the main text, we estimate the PDF f of the random component fully nonparametrically, as will be explained in detail below.[5]

The function g in the systematic component of our model is assumed to belong to the class $\mathcal{G}$ of functions that are given by any linear combination of a set of basis functions

$$K(\cdot, x_i) = \exp\left(-\kappa \| \cdot - x_i \|^2 \right) \tag{5}$$

for $i = 1, \ldots, n$, where $\kappa > 0$ is a scale parameter and $\|z\|^2 = z'z$ denotes the squared norm, in the so-called the *reproducing kernel Hilbert space* defined by K. The function K we use here to generate a functional basis is referred to as a *kernel function*.[6] The scale parameter κ in the kernel function K is a tuning parameter and has to be set a priori. We use a particular kernel function given in (5), which is most commonly used and called the *radial kernel*, though other choices are also possible. The class $\mathcal{G}$ of functions is known to be large enough to approximate any continuous function g arbitrarily well over any compact subset of its domain uniformly. Using a linear combination of the basis functions given by (5) to estimate the function g means that we obtain our estimate for g essentially by a linear combination of normal densities centered at $(x_i)_{i=1}^n$ with the same variance $\sigma^2 = 1/(2\kappa)$.

[5] For comparison, we also consider the case with f being the PDF of the standard normal distribution and the standard logistic distribution. See Appendix D.1 for the details. Their results are similar, quantitatively as well as qualitatively, to the benchmark results based on a fully nonparametric approach. Nevertheless, our fully nonparametric approach yields narrower confidence bands in most cases.

[6] The kernel function is used here to generate a space of functions defined as a reproducing kernel Hilbert space, and it is totally different from the kernel function we introduce earlier for local averaging.

We let

$$g(x) = \sum_{j=1}^{n} c_j K(x, x_j) \tag{6}$$

with a set of coefficients $(c_j)_{j=1}^n$. It is clear that there exists a set of coefficients $(c_j)_{j=1}^n$ such that

$$g(x_i) = \sum_{j=1}^{n} c_j K(x_i, x_j)$$

for all $i = 1, \ldots, n$. Indeed, if we define $g_\circ = \big(g(x_1), \ldots, g(x_n)\big)'$, $K_\circ = \big(K(x_i, x_j)\big)_{i,j=1}^n$ and $c = (c_1, \ldots, c_n)'$, then we have

$$g_\circ = K_\circ c, \tag{7}$$

from which we may easily obtain such c as $c = K_\circ^{-1} g_\circ$, since $K_\circ$ is invertible.

However, estimating the function g as in (6) with such c yields overfitting, and we need to reduce the dimension of c through an appropriate regularization method. Note that c includes n unknown parameters, i.e., as many as the sample size. To avoid the problem, we simply set

$$c = V\beta$$

with p-dimensional parameter vector β, where V is an $n \times p$ matrix whose columns are leading principal components of $K_\circ$, which are the p eigenvectors of $K_\circ$ corresponding to its p largest eigenvalues $(\lambda_i)_{i=1}^p$. This amounts to approximating (7) as

$$g_\circ \approx V\Lambda\beta, \tag{8}$$

where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$.[7] Our approach here is often used in machine learning. See Appendix A for a more detailed discussion.

For the PDF f of the random component, we follow Gallant and Nychka (1987) and choose f in the class $\mathcal{F}$ of density functions given by

$$f(u) = \frac{1}{w}\left(1 + \sum_{k=1}^{q} \alpha_k u^k\right)^2 \phi(u), \tag{9}$$

[7]Since $K_\circ$ is a symmetric matrix, we may represent it as $K_\circ = V_\circ \Lambda_\circ V_\circ'$, where $\Lambda_\circ$ is the diagonal matrix of the eigenvalues $(\lambda_i)_{i=1}^n$ of $K_\circ$ and $V_\circ$ is the $n \times n$-orthogonal matrix of the eigenvectors of $K_\circ$ associated with the eigenvalues $(\lambda_i)_{i=1}^n$. The matrices V and Λ introduced here are $n \times p$ and $p \times p$ leading submatrices of $V_\circ$ and $\Lambda_\circ$, respectively.

where $(\alpha_k)_{k=1}^q$ are the coefficients of polynomial terms, ϕ is the standard normal PDF, and w is a normalization constant given as a function of $(\alpha_k)_{k=1}^q$ introduced to make f a proper PDF. A wide variety of densities can be approximated arbitrarily well by a function of the form in (9). The class $\mathcal{F}$ of PDFs we consider here is broad and includes, for instance, all Hermite polynomials of finite order. Hermite polynomial approximation of the PDF f is particularly suitable in our model, where we let f have unbounded support. We impose the mean zero restriction

$$\int_{-\infty}^{\infty} uf(u)du = 0 \tag{10}$$

for the PDF f. Our specification of f in (9) with the zero mean restriction in (10) defines the likelihood function ℓ in (4) explicitly as a function of $(\alpha_k)_{k=1}^q$. The interested reader is referred to Appendix B for more details.

Consequently, our problem of maximizing likelihood function in (3) reduces to

$$\hat{\theta} = \underset{\substack{\beta\in\mathbb{R}^p,\alpha\in\mathbb{R}^q,\\ \tau\in\mathbb{R}^{m-1}}}{\arg\max} \sum_{i=1}^n \rho_i \ell(y_i, x_i, \theta),$$

where θ contains $p + q + (m - 1)$ parameters in $\big(\beta, \alpha, \tau\big)$. This is a completely standard problem. Our approach is thus able to handle without extra difficulty the situation when the dimension of the covariate is large and/or the sample size is large. Regardless of how large the dimension of the covariate (x_i) is, the dimensionality of (x_i) does not pose any problem to our approach. Note that we only need the covariate (x_i) in the evaluation of the kernel $K(\cdot, \cdot)$ in our approach, and the value of the kernel function is dependent only on the norm $\|x_i - x_j\|$ of the data pairs (x_i, x_j).

To select the tuning parameters including the dimension p of the parameter β the dimension q of the parameter α, which are needed to regularize our estimator for g and estimate the error density function f, respectively, we use the cross-validation, which is a standard method for selecting tuning parameters in non-parametric statistics. In the i-th iteration of the cross-validation procedure, we construct a sub-sample by leaving out the i-th observation, estimate the model with this sub-sample, make predictions for the i-th observation based on the estimated model, obtain the predicted probabilities $\big(\hat{\pi}_{ij}\big)_{j=1}^m$ for the m classes we consider, and finally calculate the sum of squared errors of the predicted probabilities across m

classes as

$$\sum_{j=1}^{m} \left(\hat{\pi}_{ij} - \pi_{ij}\right)^2,$$

where $\left(\pi_{ij}\right)_{j=1}^{m}$ is a degenerate distribution that reflects the true class probabilities of the i-th observation. Then we obtain the average of the above squared loss computed for each i and select the parameter combination that yields the smallest average loss. We search within the range of $p = 1, \ldots, 10$ and $q = 1, \ldots, 5$ and end up with $p = 7$ and $q = 2$. We set the scale parameter $\kappa = 1/2$ for the kernel function. This seems to be a reasonable choice, given that we follow the usual practice of standardizing the covariates so that they have mean zero and variance one. Setting $\kappa = 1/2$ means that we use the standard normal densities as basis functions to estimate g. Finally, we use a bootstrap procedure to obtain confidence intervals/bands of our estimates. Details of our bootstrap procedure are presented in Appendix C. The asymptotic distribution of our nonparametric estimator is not available.

4 Data

Our sample is constructed from the Panel Study of Income Dynamics (PSID). PSID is a comprehensive longitudinal household survey in the United States, tracking individuals and their descendants over several decades and containing variables on the economic, health, educational, and social behavior of individuals and families. Given the survey's time span and the fact that it tracks families across generations, it is one of the most widely used data sets in the study of intergenerational mobility.

The PSID was initiated in 1968 and contains annual data from year 1968 to 1997. Data is available biannually after 1997. We focus on a sample from the years 1968 to 1997 to avoid any inconsistency due to the change in the survey design. Our sample includes individuals who reached an age between 30 to 35 years old (inclusive) during any of our sample periods (1968-1997). We also track their parents and ultimately we consider child-parent pairs in conducting our analysis. To reflect the fact that we are studying such pairs, we shall refer to the individuals in our study as the child from now on.

Due to sample size constraints, we use the logged average household income of the head and spouse within the age range of 30 to 35 (inclusive) for the child as a measure of the child's overall economic status during adulthood. We use household income instead of personal income due to the economic partnership and risk-sharing function of marriage, by which we think that an individual's economic status is better reflected by the household income instead of his or her personal income. We adopt the Pew Research Center's methodology to categorize children into low, middle, and high-income classes (Pew Research Center, 2020). Specifically, we calculate the median income of the children and set the threshold for low income at two-thirds of this median income and for high income at twice the median income.[8] The highest income for a low-income family is \$14,147 and the lowest income for a high-income family is \$42,442, both in 1977 dollars.

The factors we propose that may affect the economic status of an individual are parental income, parental age at childbirth, parental college education, child gender, child race, child college education, and child health condition at birth. We measure parental income by the parents' logged average income during the period when the child is between 15 and 20 years old (inclusive). We choose this time span for two reasons. First, it reflects the parents' economic situation during the child's period of dependency, which significantly influences the child's economic status rather than the parents' economic condition after the child becomes economically independent. Given that many children leave home for college or work after adolescence, we focus on income up until the child reaches the age of 20. Second, due to sample size considerations, we are not able to use the parents' income throughout the entire childhood of the individual. We therefore strike a balance and consider this specific time span.[9]

We also consider parents' average age when the child was born. Reasons we consider this include parental maturity and location of parents in the life cycle of income and overall family resources.

It should be noted that income is top-coded in the PSID. Also, there are instances of zero and negative incomes in the data, which could potentially indicate measurement errors. Our method is robust in handling this censored data issue for the dependent variable as we categorize children into income classes rather than analyzing specific income levels. For parental income used as one of the covariates, we simply set the zero or negative incomes to one in our empirical analysis, as often

[8] In Appendix D.3, we also present results using tertiles, quartiles, and quintiles to classify child income groups. Classification based on quantiles implies constant shares of income groups across generations, whereas the classification used here allows the shares of income groups to vary across generations.

[9] We also present in Appendix D.2 the results using the age-10 to 15 logged parental income as an alternative measure of parental income. The two measures of parental income generate similar results.

Table 1: Summary Statistics of Variables

variable	**type**	**mean**	**std**	**min**	**max**
log child income	continuous	9.833	0.904	0.000	11.724
low income class	dummy	0.260	0.438	0.000	1.000
middle income class	dummy	0.664	0.472	0.000	1.000
high income class	dummy	0.076	0.265	0.000	1.000
log parental income	continuous	9.813	1.275	0.000	13.016
male	dummy	0.508	0.500	0.000	1.000
child college degree	dummy	0.393	0.488	0.000	1.000
parental college degree	dummy	0.312	0.463	0.000	1.000
black	dummy	0.060	0.237	0.000	1.000
parental age at birth	continuous	28.030	5.354	12.000	46.000
underweight at birth	dummy	0.049	0.216	0.000	1.000

done in the studies of intergenerational mobility.

Our benchmark sample consists of a total of 962 child-parent pairs. Survey weights provided by PSID are also used to adjust for sample selection (oversampling of low-income families) and non-random attrition in the PSID survey. Income is deflated by the Consumer Price Index for All Urban Consumers (CPI-U-RS, 1977 = 100), following the usual practice. Table 1 provides the summary statistics of the variables we use in our study.

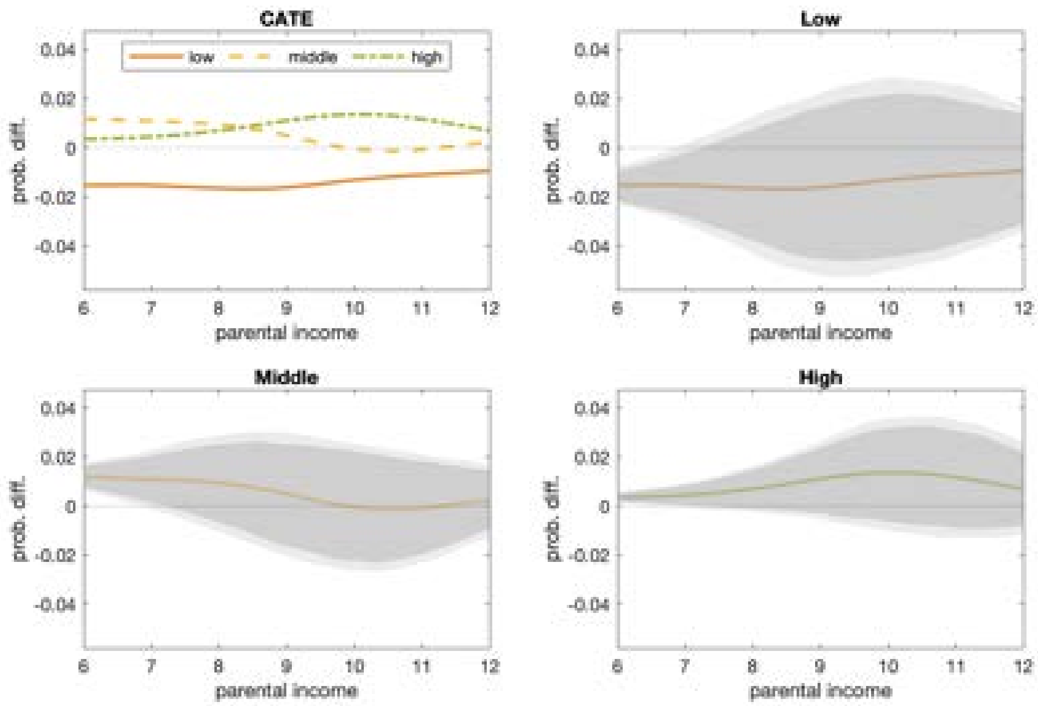
5 Empirical Results

5.1 Effects of Single Factors

We first employ our framework to analyze heterogeneous effects on children by considering gender, race, and parental education as separate factors, conditioning on different parental income levels. We do this by estimating a single nonparametric multinomial choice model and then integrating out each variable except income and the factor under consideration using (2).

Figure 1 plots the probability differentials between male and female children, as functions of parental income. In each figure, the vertical axis measures the probability differential between male and female children, and the horizontal axis measures parental income. Each parental income/gender pair produces a probability differential for membership, by the child, in each of the income classes. The upper-left panel plots the three probability differentials of the child's membership in the low-,

Figure 1: Probability Differentials Between Males and Females, as Functions of Parental Income

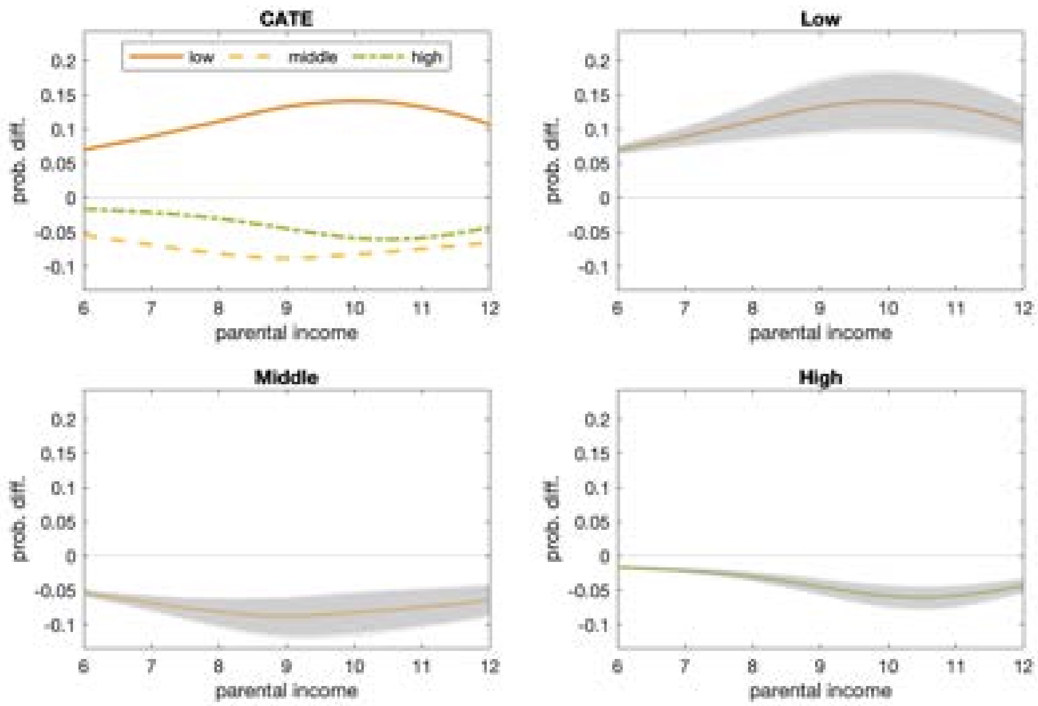


Notes: This figure presents the probability differentials of being in various income classes between male and female children, as functions of parental income. The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

middle-, and high-income classes for comparisons based on their parents' incomes. The next three panels plot the three differentials separately with confidence bands. The light and dark gray areas correspond to the 95% and 90% confidence bands, respectively.

The plots show that males, compared to females, exhibit slightly lower probabilities of entering the low-income class and slightly higher probabilities of entering the high-income class. However, these differentials are generally not statistically significant at the 0.1 significance level, except for children from the poorest families. For those, we observe that males are less (more) likely than females to be in the low (middle) category when they were born into a very poor family as shown in the upper-right and lower-left panels in Figure 1. The effects are small, but statistically

Figure 2: Probability Differentials Between Blacks and Non-Blacks, as Functions of Parental Income

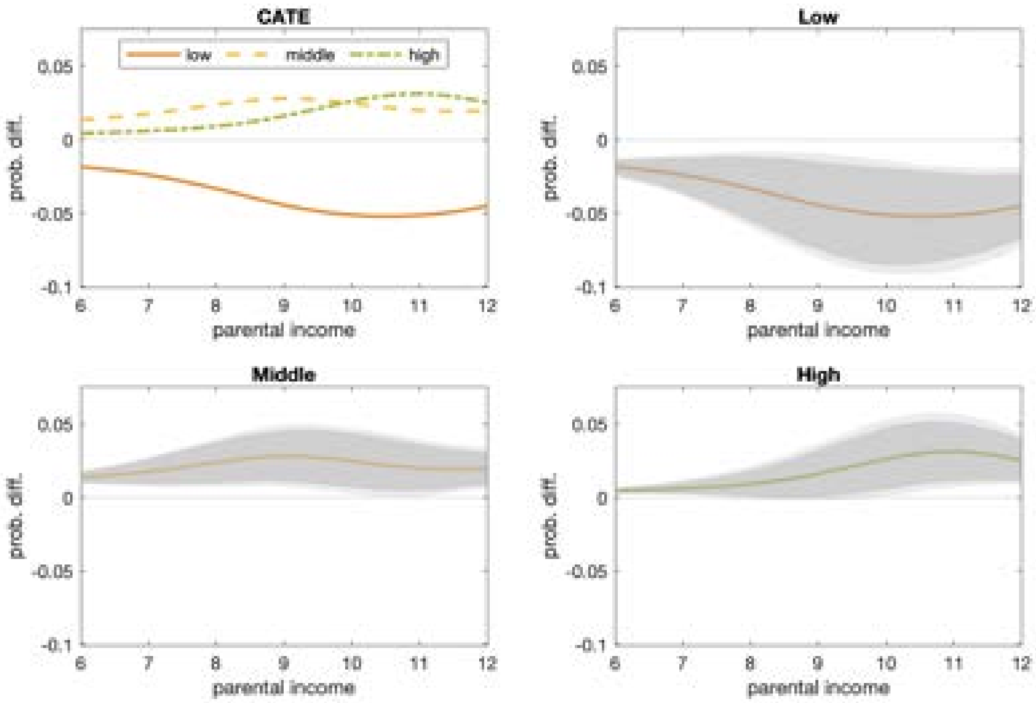


Notes: This figure presents the probability differentials of being in various income classes between Black and non-Black children, as functions of parental income. The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

significant. For the probability differentials of being in the high-income category, the gender effect is positive for all parental income levels but the magnitude is bigger for those with richer parents. However, the effect is not statistically significant at the 0.1 level. Overall, at best we find weak evidence that parental income has differential effects on the income class of offspring of different genders.

Figure 2 illustrates the probability differentials between Black and non-Black children across different parental income levels. The plots show that Black individuals have a higher likelihood of falling into the low-income class and face a comparative disadvantage in accessing the middle- and high-income classes compared to their non-Black counterparts. All these racial differentials are statistically significant. The summary of the differences is straightforward. First, for all parental

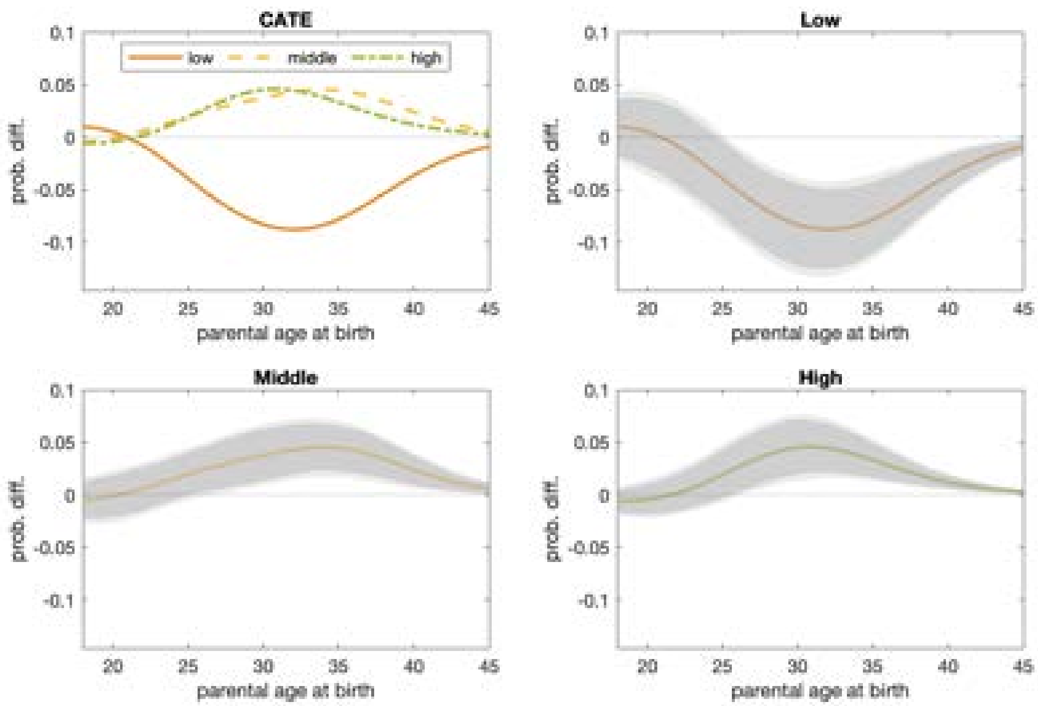
Figure 3: Probability Differentials Between Children with Parents as College and Non-college Graduates, as Functions of Parental Income



Notes: This figure presents the probability differentials of being in various income classes between children whose parents have a college degree and those whose parents do not, as functions of parental income. The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

income levels, Black children are substantially more likely to reside in the low-income category than comparable non-Blacks and are less likely to reside in either the middle or higher-income categories than non-Blacks. By implication, Black families are less able to lock in middle or high incomes than non-Black families while the low-income category is harder to escape for Blacks. These results on lower rates of upward mobility for Blacks are qualitatively similar to Bhattacharya and Mazumder (2011) and the results on relatively higher rates of downward mobility for Blacks are qualitatively similar to Chetty et al. (2020). Our ability to generate similar findings when one allows for distinct heterogeneity variables across families is an important corroboration of the salience of race as a distinct source of disparities in mobility.

Figure 4: Probability Differentials Between Children with Parents as College and Non-College Graduates, as Functions of Parental Age at Childbirth

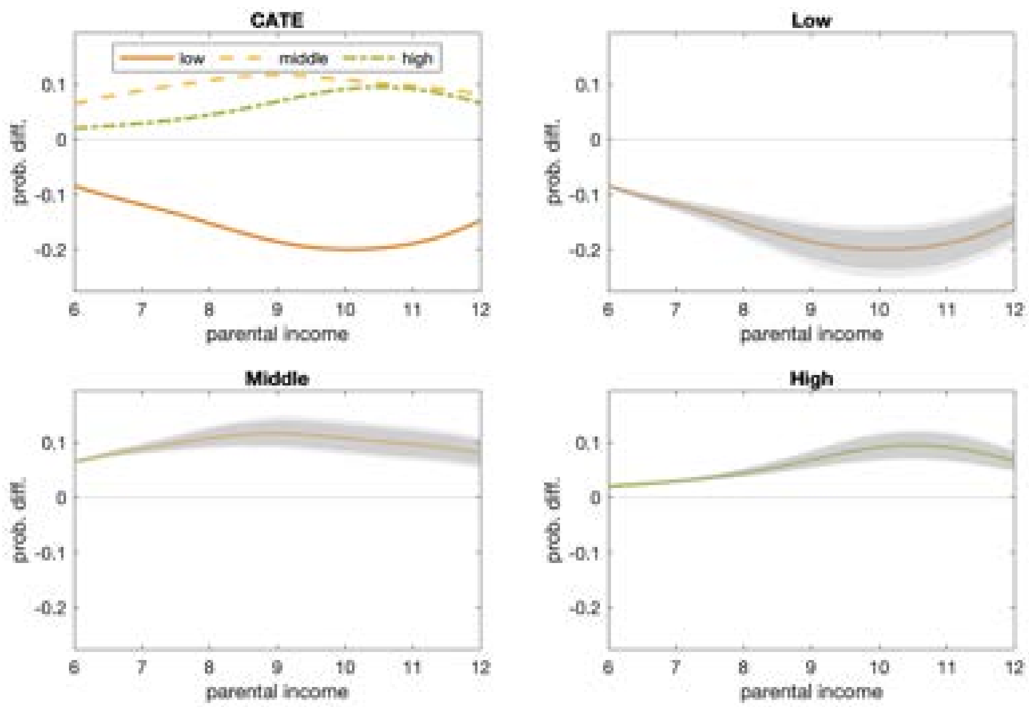


Notes: This figure presents the probability differentials of being in various income classes between children whose parents have a college degree and those whose parents do not, as functions of parental age at childbirth. The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 3 illustrates the probability differentials based on parents' possession of a college degree, relative to those without, as a function of parental income. These results reveal a significant contrast in the probability of a child ending up in the lower income category when the parents have not attended college versus when they have. This effect is especially large for families whose incomes lie in the middle-to-high range of income support, with college degrees making low income among children 5 percent less likely than otherwise. This result suggests a complementarity between parental income and parental education.

Figure 4 illustrates the probability differentials between children whose parents have or have not obtained a college degree when the average age of parents at

Figure 5: Probability Differentials (Multiple Treatments), as Functions of Parental Income



Notes: This figure presents the probability differentials of being in various income classes between children from non-Black families whose parents have a college degree and were aged 30 at childbirth, and those from Black families whose parents do not have a college degree and were aged 18 at childbirth, as functions of parental income. The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

childbirth is allowed to vary. Complementing Figure 3, Figure 4 reveals that the disparities in income class probabilities due to parental college education are most pronounced when parents give birth during their late twenties to mid-thirties.

5.2 Combining Factors

Combining our single factor analyses in the previous section reveals the existence of family background configurations that make it challenging for a child to avoid the low-income category. Our findings from these analyses indicate the presence of a privileged group of children, originating from non-Black families with college-educated parents, and born when their parents are around 30. This group of children contrasts sharply with children who are Black, born to parents at the age of 18, and without college degrees.

Figure 5 presents the probability differentials as a function of parental incomes. Particularly striking are the results in the upper-right panel, where the probability of a child being in the low-income category consistently remains 10 percent higher for our disadvantaged category across all family income levels. For middle-to-high-income categories, this disparity exceeds 20 percent. These findings provide further insight into how the probability of downward mobility varies across different demographic groups.

6 Conclusions

This paper proposes a fully nonparametric multinomial outcome model to study intergenerational income mobility. Our approach effectively captures the nonlinear and interactive effects of various factors on personal income status and societal mobility levels. It demonstrates strong computational efficiency and robustness, particularly suitable for analyzing large datasets with high-dimensional covariates. We affirm race, parental education, and parental childbearing age as crucial determinants influencing intergenerational mobility. Each of these factors significantly impacts the predictive power of parental income for the incomes of children. These findings are all consistent with the overall state of the mobility literature. The robustness of these claims to the general nonparametric framework we set up reinforces the vision that they simultaneously matter.

Our findings, which highlight the distinct relationships shaped by race, parental education, and parental childbearing age, underscore the importance of systematically investigating *bottlenecks* in intergenerational mobility dynamics. By bottlenecks, we refer to a set of family background variables that perpetuate low incomes across generations, where higher incomes alone may not suffice to break such persistence. The differences between children of non-Black families with college-educated parents, and born to parents in their early 30s as opposed to children who are Black, born to parents around 18, without college degrees are stark. These speak to the idea of bottlenecks in the income dynamics where some set of conditions during childhood makes socioeconomic successes in income and education highly unlikely. These phenomena represent a natural stochastic generalization of poverty trap models. Currently, we are actively pursuing further research on this topic.

While our analysis is statistical and so does not directly identify underlying mechanisms, we note that it does speak to general aspects of theorizing about intergenerational mobility. First, if one contrasts the classical economic models of mobility due to Becker and Tomes (1979, 1986) with the classic sociological perspective associated with the Wisconsin Status Attainment Model, Sewell and Portes (1969) and Sewell and Ohlendorf (1970), we think the interactions between educational status, ethnicity and age suggest the importance of placing multiple channels at the heart of the sociology approach. This highlights the value of recent advances in formal economic mobility models that include social and psychological factors, see Durlauf et al. (2022) for elaboration. Second, the stark contrasts we estimate between the effects of parental income on offspring for parents with different ethnicities, educations, and ages are suggestive of mechanisms that apply to groups rather than individuals. By this, theories of categorical inequality (Tilly, 1998; Massey, 2007), or theories in economics based on group memberships (Durlauf, 1999, 2006; Darity, 2022) are suggestive of the types of patterns we find. To be clear, to say more will require even richer models than we consider here. One natural path involves explicit attention to neighborhood effects along the lines pursued in Wodtke et al. (2016). Expanding our tools in these directions is underway.

References

Abadie, A., Angrist, J., and Imbens, G. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, 70(1):91–117.

Bacic, R. and Zheng, A. (2024). Race and the income-achievement gap. *Economic Inquiry*, 62(1):5–23.

Becker, G. S. and Tomes, N. (1979). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy*, 87(6):1153–1189.

Becker, G. S. and Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of Labor Economics*, 4(3):1–39.

Bhattacharya, D. and Mazumder, B. (2011). A nonparametric analysis of Black–white differences in intergenerational income mobility in the United States. *Quantitative Economics*, 2:335–379.

Bloome, D. (2014). Racial inequality trends and the intergenerational persistence of income and family. *American Sociological Review*, 79(6):1196–1225.

Bloome, D. (2017). Childhood family structure and intergenerational income mobility in the United States. *Demography*, 54(2):541–569.

Brand, J. E. and Xie, Y. (2010). Who benefits most from college?: Evidence for negative selection in heterogeneous economic returns to higher education. *American Sociological Review*, 75(2):273–302.

Chen, J. and Roth, J. (2024). Logs with zeros? Some problems and solutions. *Quarterly Journal of Economics*, 139(2):891–936.

Chetty, R., Grusky, D., Hell, M., Hendren, N., Manduca, R., and Narang, J. (2017). The fading American Dream: Trends in absolute income mobility since 1940. *Science*, 356:398–406.

Chetty, R., Hendren, N., Jones, M., and Porter, S. (2020). Race and economic opportunity in the United States: An intergenerational perspective. *Quarterly Journal of Economics*, 135(2):711–783.

Dahl, G. and Lochner, L. (2012). The impact of family income on child achievement: Evidence from the earned income tax credit. *American Economic Review*, 102(5):1927–1956.

Darity, W. (2022). Position and possession: Stratification economics and intergroup inequality. *Journal of Economic Literature*, 60:713–736.

Duncan, O. D. (1968). Patterns of occupational mobility among Negro men. *Demography*, 5:11–22.

Durlauf, S. N. (1999). The memberships theory of inequality: Ideas and implications. In Brezis, E. and Temin, P., editors, *Elites, Minorities, and Economic Growth*. North Holland.

Durlauf, S. N. (2006). Groups, social influences, and inequality: A memberships theory perspective on poverty traps. In Bowles, S., Durlauf, S., and Hoff, K., editors, *Poverty Traps*. Princeton University Press.

Durlauf, S. N., Kourtellos, A., and Tan, C. M. (2022). The Great Gatsby curve. *Annual Review of Economics*, 14:571–605.

Gallant, A. R. and Nychka, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica*, 55(2):363–390.

Hout, M. (1984). Occupational mobility of Black men: 1962 to 1973. *American Sociological Review*, 49:308–322.

Lopoo, L. M. and DeLeire, T. (2014). Family structure and the economic wellbeing of children in youth and adulthood. *Social Science Research*, 43:30–44.

Maralani, V. (2013). The demography of social mobility: Black-white differences in the process of educational reproduction. *American Journal of Sociology*, 118(6):1509–1558.

Massey, D. (2007). *Categorically Unequal*. Russell Sage Foundation Press.

Pew Research Center (2020). Most Americans say there is too much economic inequality in the U.S., but fewer than half call it a top priority. Research Report.

Sewell, W., A. H. and Ohlendorf, G. (1970). The educational and early occupational status attainment process: Replication and revision. *American Sociological Review*, 35:1014–1024.

Sewell, W., A. H. and Portes, A. (1969). The educational and early occupational attainment process. *American Sociological Review*, 34:82–92.

Stewart, M. B. (2005). A comparison of semiparametric estimators for the ordered response model. *Computational Statistics & Data Analysis*, 49(2):555–573.

Tilly, C. (1998). *Durable Iequality.* University of California Press.

Wodtke, G., Harding, D., and Elwert, F. (2016). Neighborhood effect heterogeneity by family income and developmental period. *American Journal of Sociology*, 121:1168–1222.

Yan, G. (2023). Nonparametric estimation of large dimensional binary choice models. Manuscript.

A Estimation of Systematic Component

Let our covariates (x_i) be r-dimensional and take values in a subset $\mathcal{D}$ of $\mathbb{R}^r$. The function g on $\mathcal{D}$ in the systematic component of our ordered choice model is estimated as a function in the *reproducing kernel Hilbert space* (RKHS) H_K defined by a kernel $K : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$

$$K(x, y) = \exp\left(-\kappa \|x - y\|^2\right),$$

where $\kappa > 0$ is the scale parameter, $x, y \in \mathcal{D}$ and $\|x - y\|^2 = (x - y)'(x - y)$ denotes the squared distance between x and y in $\mathcal{D}$. This kernel function is symmetric, i.e., $K(x, y) = K(y, x)$ for all $x, y \in \mathcal{D}$, and positive definite, i.e.,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j K(x_i, x_j) > 0$$

for any $c_1, \ldots, c_n$ not all identically zero and for all (x_i) in $\mathcal{D}$. These two properties are essential in the sense that we may use any continuous function to define a RKHS if it satisfies these properties. Here we choose the most commonly used kernel, which is often called the radial kernel, although many other choices of kernel are also possible.

The RKHS H_K defined by the kernel K is a vector space involving all functions given as linear combinations of

$$K(\cdot, x_1), \ldots, K(\cdot, x_n) \tag{11}$$

for all choices of n and $x_1, \ldots, x_n \in \mathcal{D}$, which is endowed with the inner product $\langle \cdot, \cdot \rangle_K$ defined by

$$\langle K(\cdot, x), K(\cdot, y) \rangle_K = K(x, y) \tag{12}$$

for any $x, y \in \mathcal{D}$. The value $K(x, y)$ of kernel function K may thus be obtained by taking the inner product of two functions $K(\cdot, x)$ and $K(\cdot, y)$ for each $(x, y) \in \mathcal{D} \times \mathcal{D}$, and therefore, the kernel function K may be *reproduced* from the inner product of functions in H_K. For this reason, H_K is called a RKHS. The RKHS H_K defined by the radial kernel K introduced above includes a wide range of functions. It is indeed known that any continuous function can be approximated arbitrarily well by a function in this RKHS uniformly on any compact subset of $\mathcal{D}$.

To estimate the function g defining the systematic component of our ordered choice model, we assume $g \in H_K$ and write it as

$$g(x) = \sum_{j=1}^{n} c_j K(x, x_j) \tag{13}$$

for $x \in \mathcal{D}$, where $(c_j)_{j=1}^n$ are a set of unknown parameters. This is the most flexible specification of g. Since $g(x)$ is observed only at n-number of x's given by $(x_i)_{i=1}^n$, we may choose n-unknown parameters $(c_j)_{j=1}^n$ appropriately to have a perfect fit for $\big(g(x_i)\big)$. Note that $g(x_i) = \sum_{j=1}^n c_j K(x_i, x_j)$ for $i = 1, \ldots, n$, which we may write as

$$g_\circ = K_\circ c$$

in matrix form, where $g_\circ = \big(g(x_1), \ldots, g(x_n)\big)'$, $c = (c_1, \ldots, c_n)'$ and $K_\circ$ is an $n \times n$ invertible matrix defined as

$$K_\circ = \begin{pmatrix} K(x_1, x_1) & \cdots & K(x_1, x_n) \\ \vdots & \ddots & \vdots \\ K(x_n, x_1) & \cdots & K(x_n, x_n) \end{pmatrix}.$$

Here the entries of $K_\circ$ are given by the inner products of basis functions $\big(K(\cdot, x_j)\big)_{j=1}^n$ for an n-dimensional subspace of H_K, and such a matrix is generally referred to as a *Gram matrix*.

Let

$$g^*(x) = \big(K(x, x_1), \ldots, K(x, x_n)\big)c^* \tag{14}$$

with $c^* = K_\circ^{-1} g_\circ$, so that $g^*(x_i) = g(x_i)$ for all $i = 1, \ldots, n$. Then it follows from (12) that

$$\big\langle K(\cdot, x_i), g(\cdot) - g^*(\cdot)\big\rangle_K = g(x_i) - g^*(x_i) = 0$$

for all $i = 1, \ldots, n$, which implies that $g - g^*$ is orthogonal to the n-dimensional subspace V_K of H_K spanned by the basis $\big(K(\cdot, x_i)\big)_{i=1}^n$ introduced in (11). Therefore, g^* is the orthogonal projection of g on V_K in H_K.

However, the specification g^* of g in (14) is too flexible, which needs to be regularized. There are several ways of regularizing, one of which is to introduce a penalty term given by

$$\lambda \|g\|_K^2 = \lambda \left\langle \sum_{i=1}^n c_i K(\cdot, x_i), \sum_{j=1}^n c_j K(\cdot, x_j) \right\rangle_K = \lambda c' K_\circ c$$

obtained from (12) and (13) with an appropriately chosen penalty parameter $\lambda > 0$. This is usually done in the regression model. For our discrete choice model, we use a simpler, but known to be equally effective, method based on rank reduction of the Gram matrix $K_\circ$ defined above. The symmetric matrix $K_\circ$ admits the spectral representation given by

$$K_\circ = V_\circ \Lambda_\circ V_\circ',$$

where $\Lambda_\circ$ is a diagonal matrix of the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n > 0$ of $K_\circ$ and $V_\circ$ is an orthogonal matrix with columns given by the corresponding eigenvectors $v_1, \ldots, v_n$ of K. The matrix $K_\circ$ of rank n can be best approximated by the matrix

$$K_\bullet = V \Lambda V'$$

of rank p, $p < n$, where V is a semi-orthogonal matrix given by the $n \times p$ leading submatrix of $V_\circ$ and Λ is a diagonal matrix given by the $p \times p$ leading submatrix of $\Lambda_\circ$. Accordingly, we restrict the unknown parameter c introduced earlier to be in a p-dimensional subspace of $\mathbb{R}^n$ spanned by $v_1, \ldots, v_p$ and write $c = V\beta$ for a newly defined unknown parameter β in $\mathbb{R}^p$. Then we have

$$g_\circ = K_\circ c \approx K_\bullet c = V \Lambda \beta$$

with an p-dimensional unknown parameter β. We may easily obtain the maximum likelihood estimator $\hat{\beta}$ of β along with the maximum likelihood estimator $\hat{\alpha}$ of the other parameter α defined in the next section. Finally, we have

$$g(x) = \big(K(x, x_1), \ldots, K(x, x_n)\big) c = \big(K(x, x_1), \ldots, K(x, x_n)\big) V \beta,$$

which may be estimated by

$$\hat{g}(x) = \big(K(x, x_1), \ldots, K(x, x_n)\big) V \hat{\beta}$$

for any $x \in \mathcal{D}$. In our application, p is chosen using the standard leave-one-out cross-validation. Typically, p is chosen to be substantially smaller than n.

B Estimation of Random Component Distribution

Stewart (2005) and Yan (2023) show that the CDF F in the likelihood function ℓ in (4) and the zero mean restriction on the PDF f in (10) can be written explicitly as functions of $\alpha = (\alpha_1, \ldots, \alpha_q)'$. In fact, Let

$$m_k(u) = u^k \phi(u) \quad \text{and} \quad m_k = \int_{-\infty}^{\infty} m_k(u) du,$$

where m_k is the k-th moment of the standard normal distribution which is given explicitly as

$$m_0 = 1,\ m_1 = 0 \ \text{ and } \ m_k = (k-1) m_{k-2} \ \text{ for } \ k \geq 2.$$

Also, define the cumulative k-th moment function of the standard normal distribution as

$$M_k(u) = \int_{-\infty}^{u} m_k(v) dv,$$

which is given explicitly as

$$\begin{aligned}
&M_0(u) = \Phi(u),\ M_1(u) = -\phi(u),\ M_2(u) = -u\phi(u) + \Phi(u) \\
&M_k(u) = u\big[M_{k-1}(u) - (k-2) M_{k-3}(u)\big] + (k-1) M_{k-2}(u) \ \text{ for } \ k \geq 3
\end{aligned}$$

recursively, where Φ is the standard normal CDF. Finally, we let

$$c_k(\alpha) = \sum_{\ell = 0 \vee (k-q)}^{k \wedge q} \alpha_k \alpha_{k-\ell},$$

where $\vee$ and $\wedge$ denote the maximum and minimum, respectively.

Now we may rewrite f as

$$f(u) = \left[\sum_{k=0}^{2q} c_k(\alpha) m_k \right]^{-1} \sum_{k=0}^{2q} c_k(\alpha) m_k(u),$$

from which it follows that

$$F(u) = \left[\sum_{k=0}^{2q} c_k(\alpha) m_k\right]^{-1} \sum_{k=0}^{2q} c_k(\alpha) M_k(u),$$

and the zero mean restriction as

$$\sum_{k=0}^{2q} c_k(\alpha) m_{k+1} = 0.$$

These closed-form representations of the CDF F and the zero mean restriction on the PDF f make our maximum likelihood procedure extremely simple and straightforward. In particular, our maximum likelihood procedure does not require any numerical integration, which is generally necessary for the nonparametric estimation of discrete choice models.

C Bootstrap Details

We use bootstrap to obtain confidence bands for the heterogeneous treatment effects presented in Section 5. In the following, we describe some details of our bootstrap procedure.

Let θ be a parameter of interest and $\hat{\theta}$ its estimates from the data. To obtain its bootstrap confidence interval, in the bootstrap iteration b, we first resample from the original data, and estimate θ with the resampled data using the same procedures as those used for the estimation with the original data. Denote the b-th estimate with the resampled data by $\hat{\theta}^*_b$. Repeat this procedure B times and we obtain a vector $\hat{\theta}^* = (\hat{\theta}^*_1, \hat{\theta}^*_2, \ldots, \hat{\theta}^*_B)$ of bootstrap estimates for θ.

For $r \in (0, 1/2)$, if θ is one-dimensional, we obtain the $r/2$ and $(1-r/2)$ quantiles of $\hat{\theta}^*$, denoted by $\hat{\theta}^*_l$ and $\hat{\theta}^*_u$, respectively. Also, we calculate the mean of $\hat{\theta}^*$, and denote it by $\bar{\theta}^*$. We then construct

$$\left[\hat{\theta} + \hat{\theta}^*_l - \bar{\theta}^*,\ \hat{\theta} + \hat{\theta}^*_u - \bar{\theta}^*\right]$$

as our bootstrapped $100(1-r)\%$ confidence interval for θ.

If θ is a function as in our case of heterogeneous treatment effect, we obtain from $\hat{\theta}^*(z) = (\hat{\theta}^*_1(z), \hat{\theta}^*_2(z), \ldots, \hat{\theta}^*_B(z))$ the pointwise quantiles $\hat{\theta}^*_l(z)$ and $\hat{\theta}^*_u(z)$ for each z, and construct pointwise confidence interval as

$$\left[\hat{\theta}(z) + \hat{\theta}_l^*(z) - \bar{\theta}^*(z),\ \hat{\theta}(z) + \hat{\theta}_u^*(z) - \bar{\theta}^*(z)\right].$$

D Robustness Checks

D.1 Alternative Specifications of Random Term Distribution

Figures 6 to 10 show the probability differentials when the density of the random term u_i is set to be the density of the standard Gaussian distribution as in the probit model. Figures 11 to 15 show the probability differentials when the density of u_i is set to be the density of the standard logistic distribution as in the logit model. In both alternative settings of the distribution of u_i, the results are very similar, both qualitatively and quantitatively, to the benchmark results obtained with our fully nonparametric approach. All conclusions in the main text therefore remain intact. We note that our approach yields narrower confidence bands in most cases, suggesting that our fully nonparametric approach can provide more precise interval estimates.

D.2 Alternative Measures of Parental Income

Figures 16 to 20 present the probability differentials using average parental income during middle childhood (age 10 to 15, both inclusive) as an alternative measure of parental income in the model. The results from using the two measures are similar to each other. All conclusions in the main text are not changed.

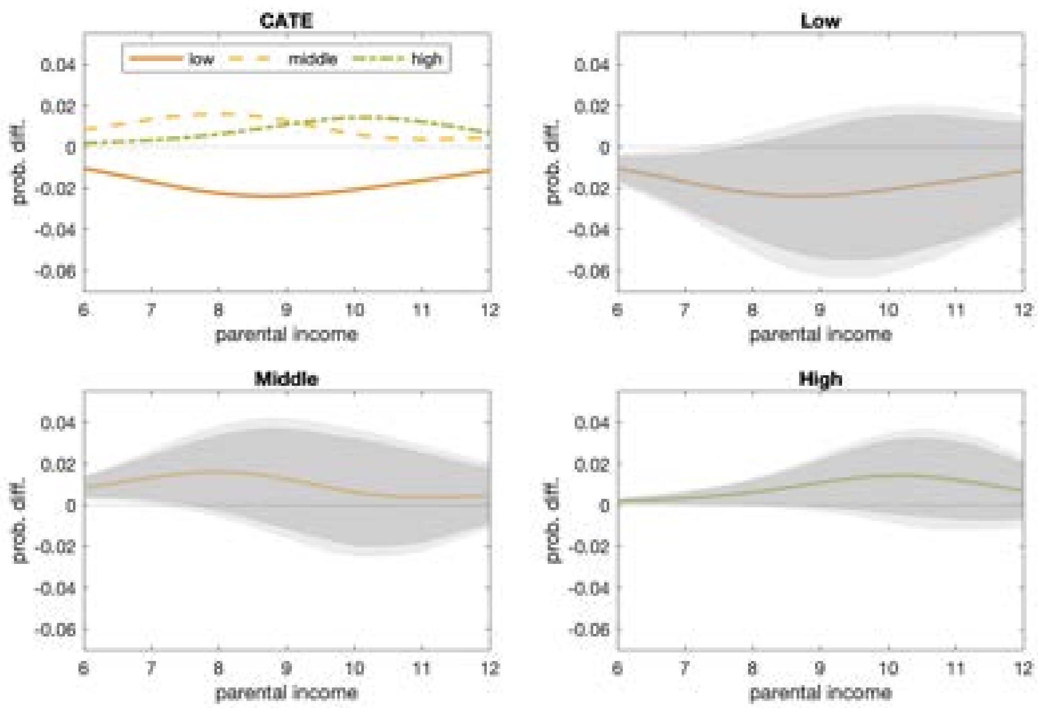
D.3 Alternative Income Class Thresholds

We also consider using alternative thresholds to classify child adult incomes. In particular, we classify the observed child income into three classes using tertiles. The main difference between this classification and the one in our main text is that classification by tertiles leads to a constant relative aggregate size of all three income classes (each constituting one-third), while the classification in our main text allows the relative size of each class to change as one generation transitions to the next. We also follow one of the referees' suggestions to consider classification based on quartiles and quintiles for finer categorizations.

Figures 21 through 25 present the probability differentials using tertiles to classify observed child income. The patterns for the estimated probability differentials of being in the low-income class remain roughly the same as those reported in the main text. However, the magnitude of the probability differentials for being in the middle-income class diminishes to near zero, while the magnitude of the probability differentials for being in the high-income class increases significantly. This is likely because part of the middle-income children under the benchmark classification are reclassified as high income under this new classification. In our judgment, the benchmark classification in the text better captures the substantive distinctions between lower, middle, and upper income classes than equal divisions of the income distribution. Specifically, the large-middle income category in the benchmark specification implicitly allows the data to reveal distinctions between affluence and disadvantage that are obscured under equal income distribution divisions.

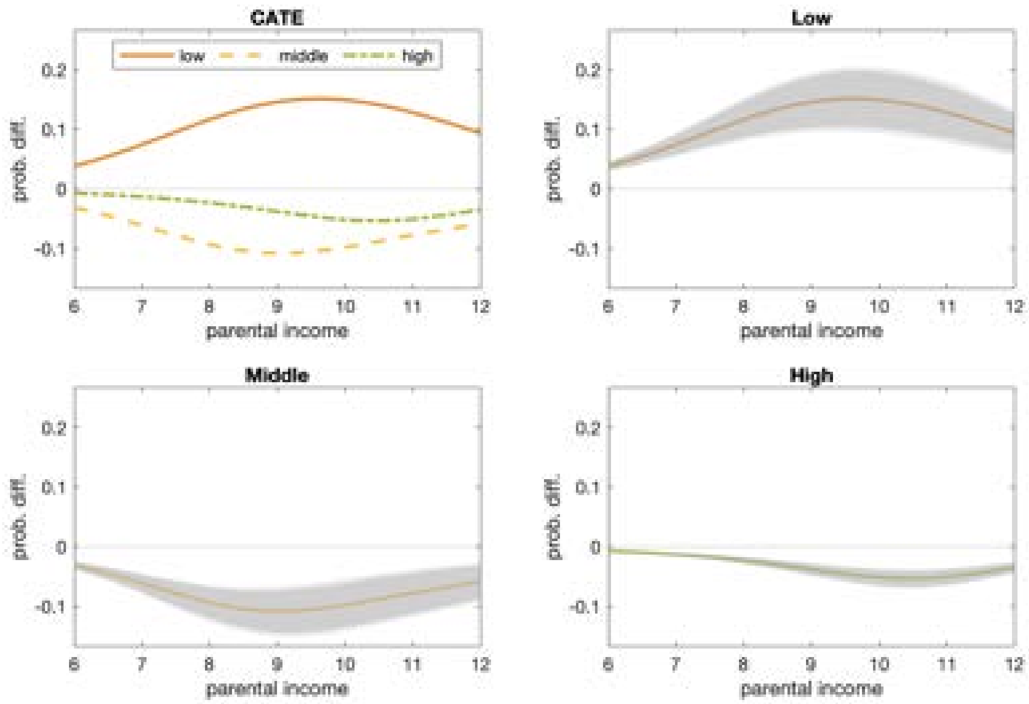
Similar patterns appear in the results using quartiles and quintiles for classification. Figures 26 through 30 present the probability differentials using quartiles to classify observed child income, and Figures 31 through 35 present the probability differentials using quintiles. In both cases, the magnitude of the probability differentials for being in the lowest and highest income classes increases, while the magnitude of the probability differentials for being in the middle two or three income classes decreases.

Figure 6: Probability Differentials Between Males and Females, as Functions of Parental Income



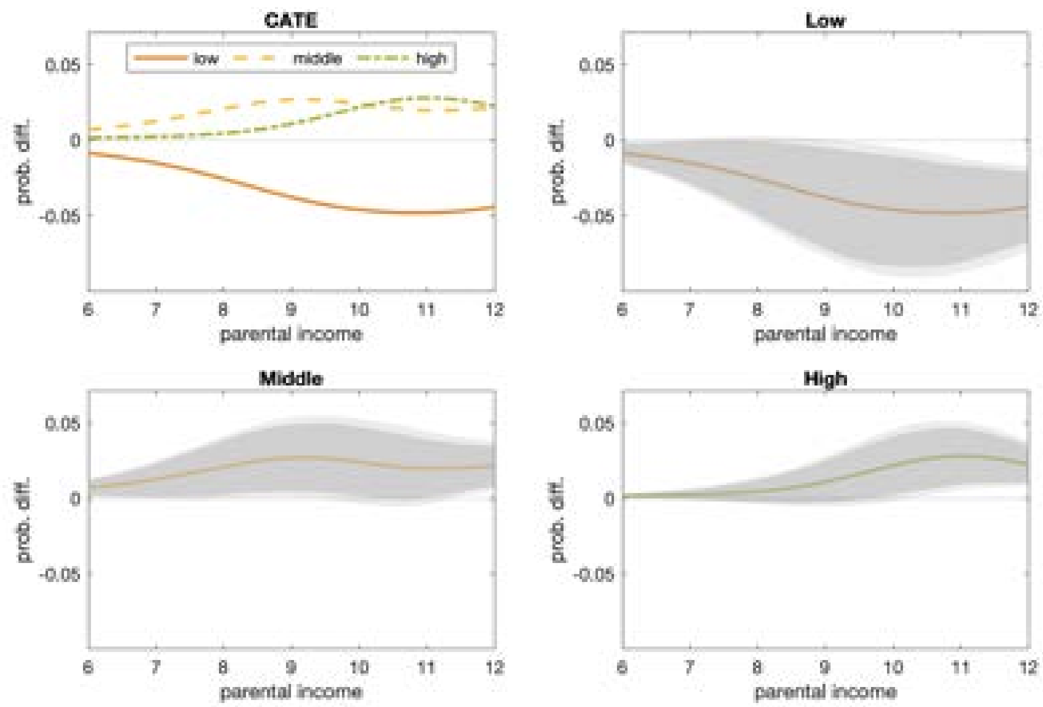
Notes: This figure presents the probability differentials of being in various income classes between male and female children, as functions of parental income, *with probit random term.* The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 7: Probability Differentials Between Blacks and Non-Blacks, as Functions of Parental Income



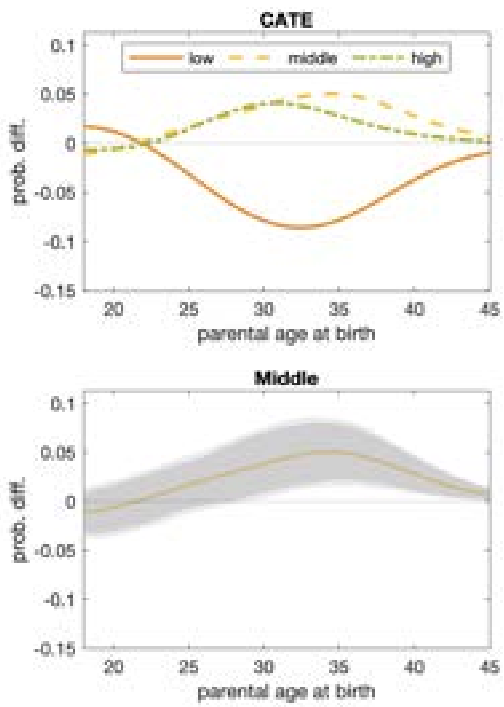
Notes: This figure presents the probability differentials of being in various income classes between Black and non-Black children, as functions of parental income, *with probit random term.* The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 8: Probability Differentials Between Children with Parents as College and Non-college Graduates, as Functions of Parental Income



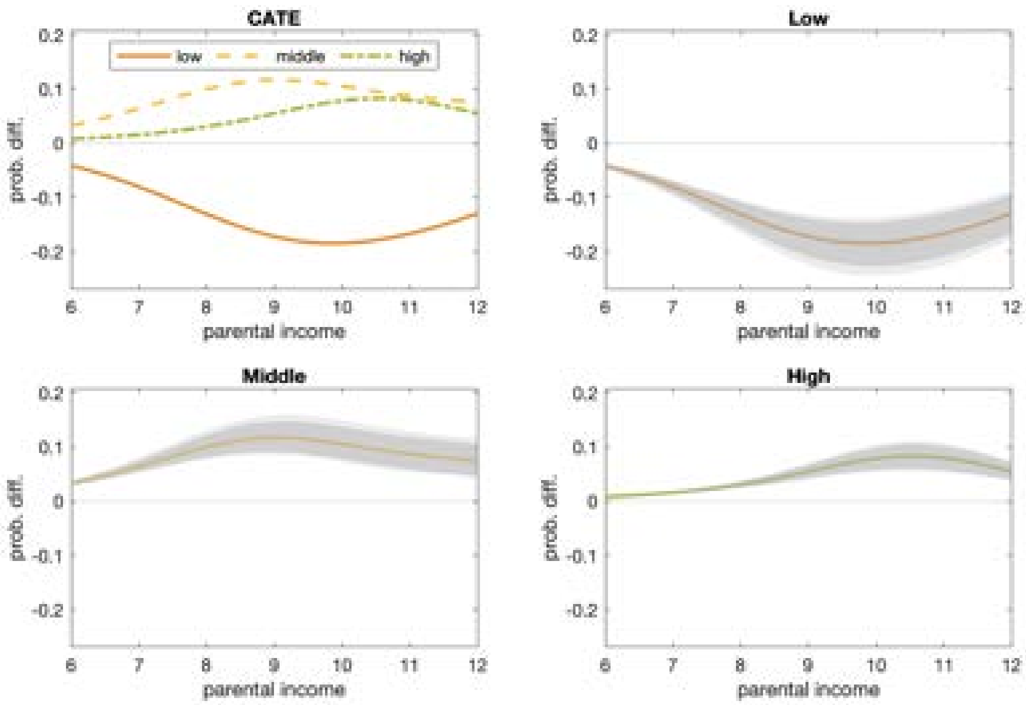
Notes: This figure presents the probability differentials of being in various income classes between children whose parents have a college degree and those whose parents do not, as functions of parental income, *with probit random term*. The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 9: Probability Differentials Between Children with Parents as College and Non-College Graduates, as Functions of Parental Age at Childbirth



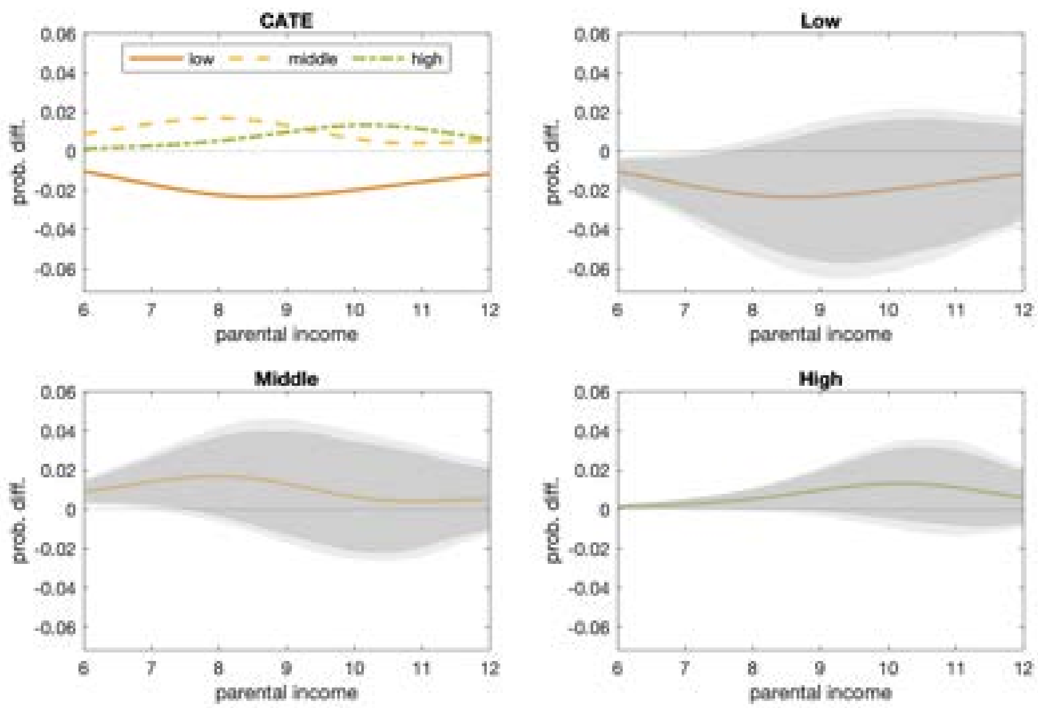
Notes: This figure presents the probability differentials of being in various income classes between children whose parents have a college degree and those whose parents do not, as functions of parental age at childbirth, *with probit random term*. The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 10: Probability Differentials (Multiple Treatments), as Functions of Parental Income



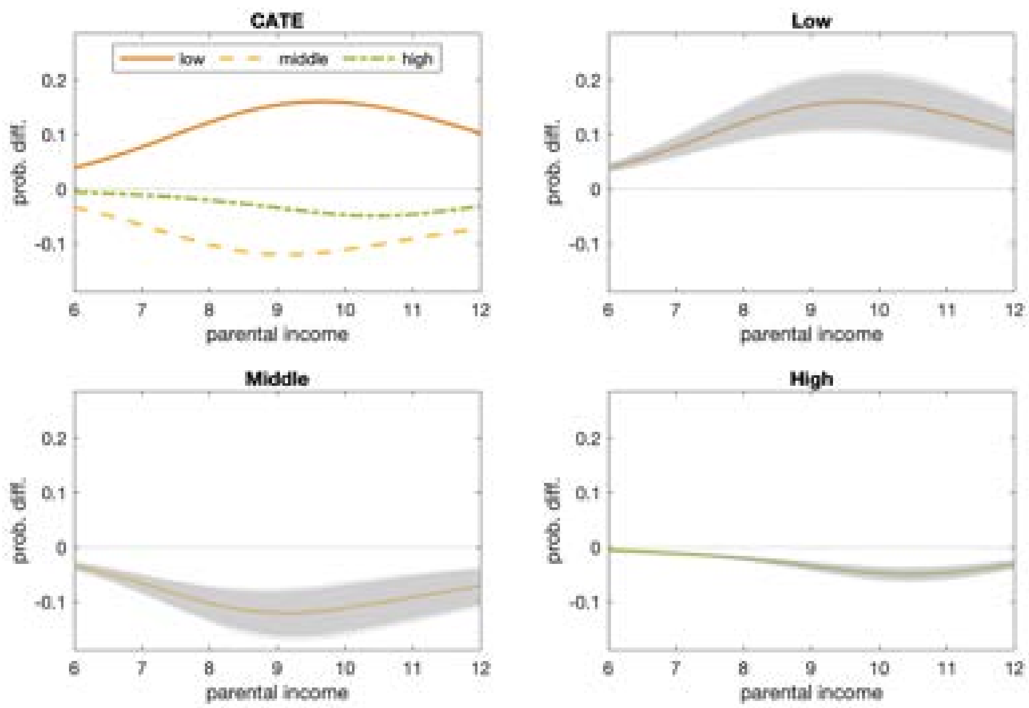
Notes: This figure presents the probability differentials of being in various income classes between children from non-Black families whose parents have a college degree and were aged 30 at childbirth, and those from Black families whose parents do not have a college degree and were aged 18 at childbirth, as functions of parental income, *with probit random term*. The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 11: Probability Differentials Between Males and Females, as Functions of Parental Income



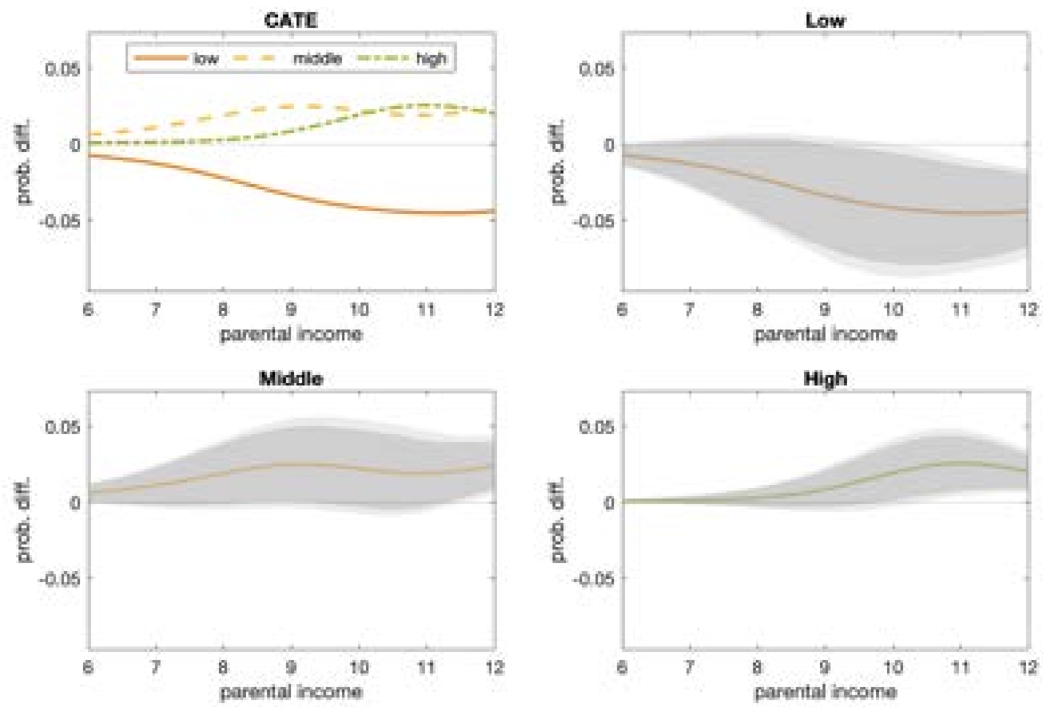
Notes: This figure presents the probability differentials of being in various income classes between male and female children, as functions of parental income, *with logit random term.* The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 12: Probability Differentials Between Blacks and Non-Blacks, as Functions of Parental Income



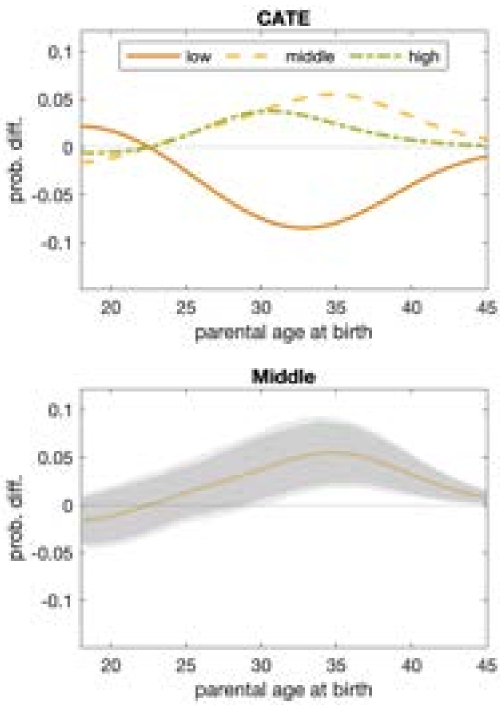
Notes: This figure presents the probability differentials of being in various income classes between Black and non-Black children, as functions of parental income, *with logit random term.* The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 13: Probability Differentials Between Children with Parents as College and Non-college Graduates, as Functions of Parental Income



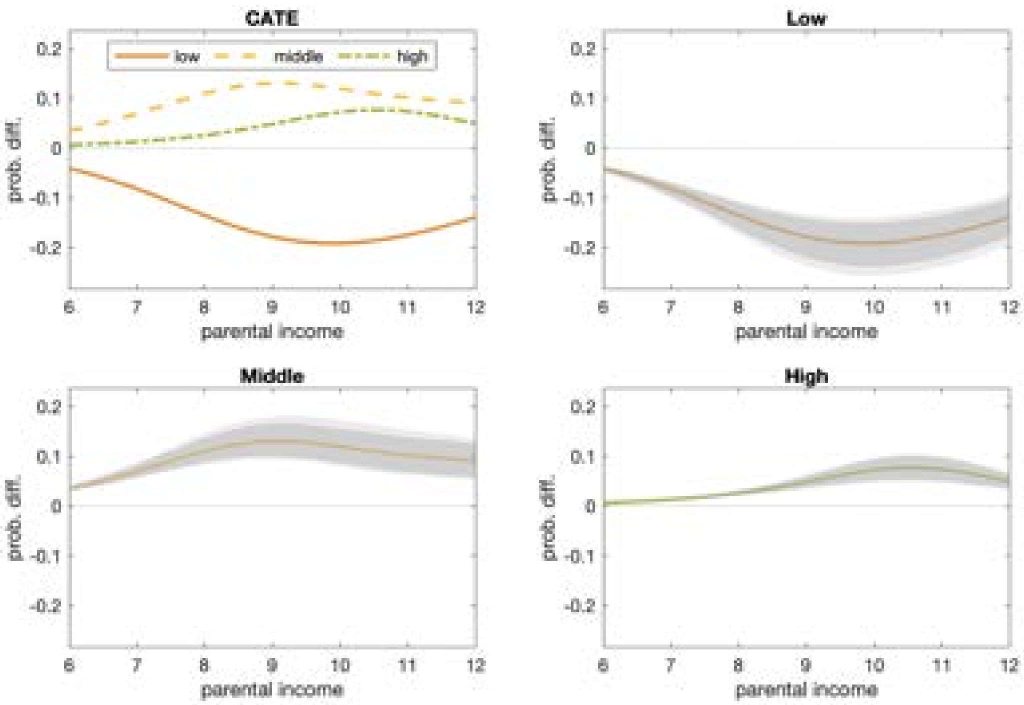
Notes: This figure presents the probability differentials of being in various income classes between children whose parents have a college degree and those whose parents do not, as functions of parental income, *with logit random term*. The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 14: Probability Differentials Between Children with Parents as College and Non-College Graduates, as Functions of Parental Age at Childbirth



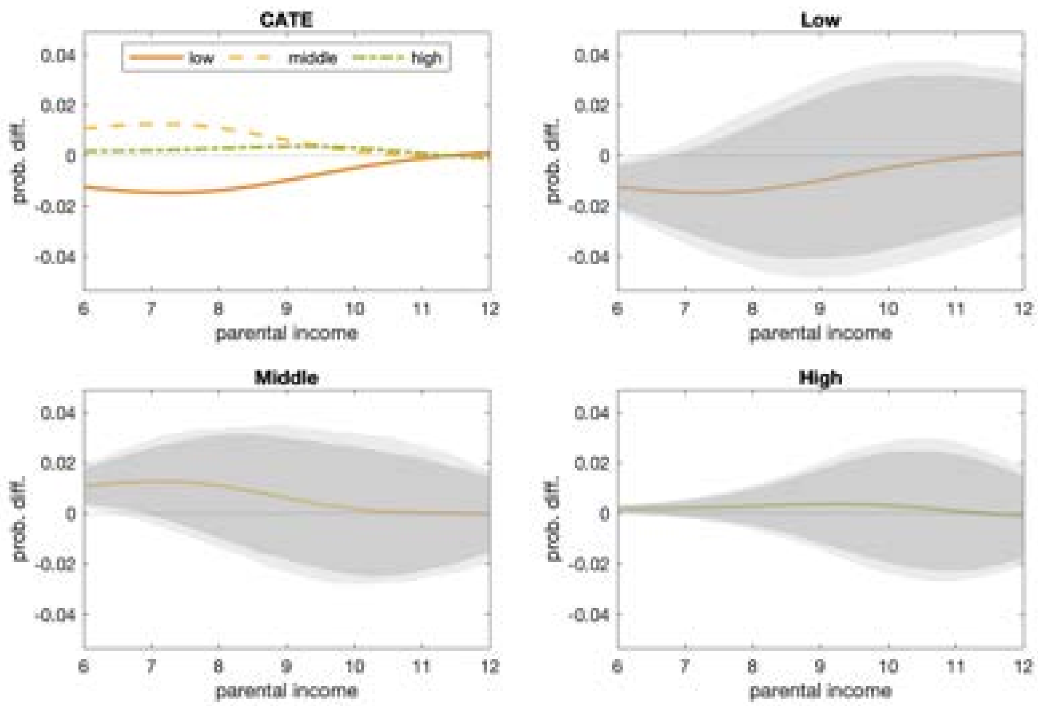
Notes: This figure presents the probability differentials of being in various income classes between children whose parents have a college degree and those whose parents do not, as functions of parental age at childbirth, *with logit random term*. The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 15: Probability Differentials (Multiple Treatments), as Functions of Parental Income



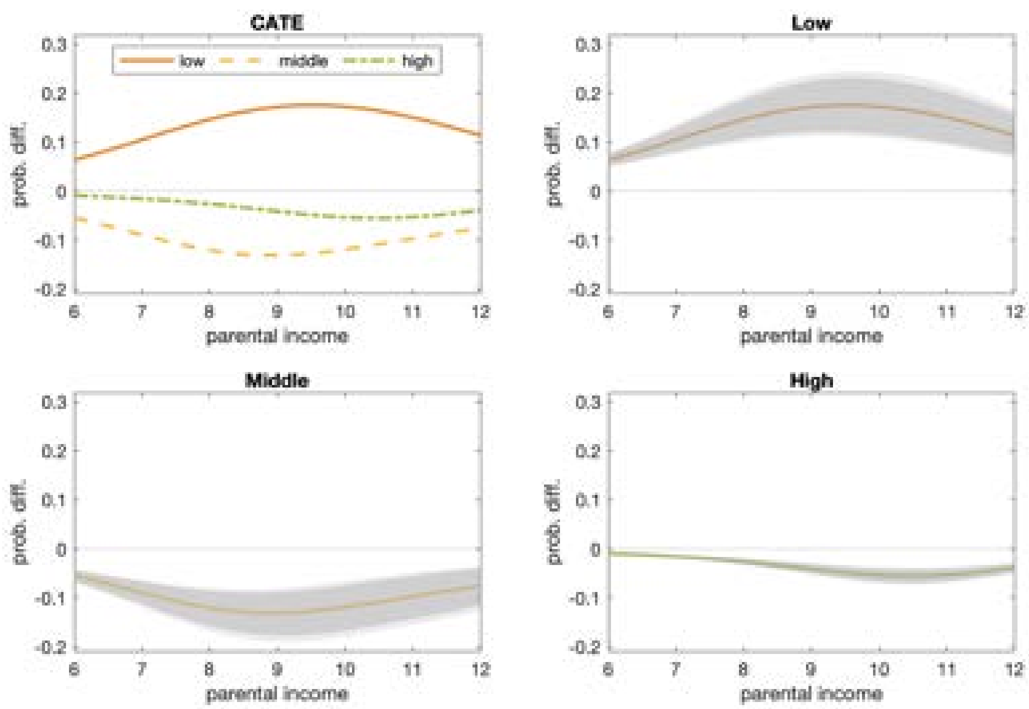
Notes: This figure presents the probability differentials of being in various income classes between children from non-Black families whose parents have a college degree and were aged 30 at childbirth, and those from Black families whose parents do not have a college degree and were aged 18 at childbirth, as functions of parental income, *with logit random term.* The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 16: Probability Differentials Between Males and Females, as Functions of Parental Income



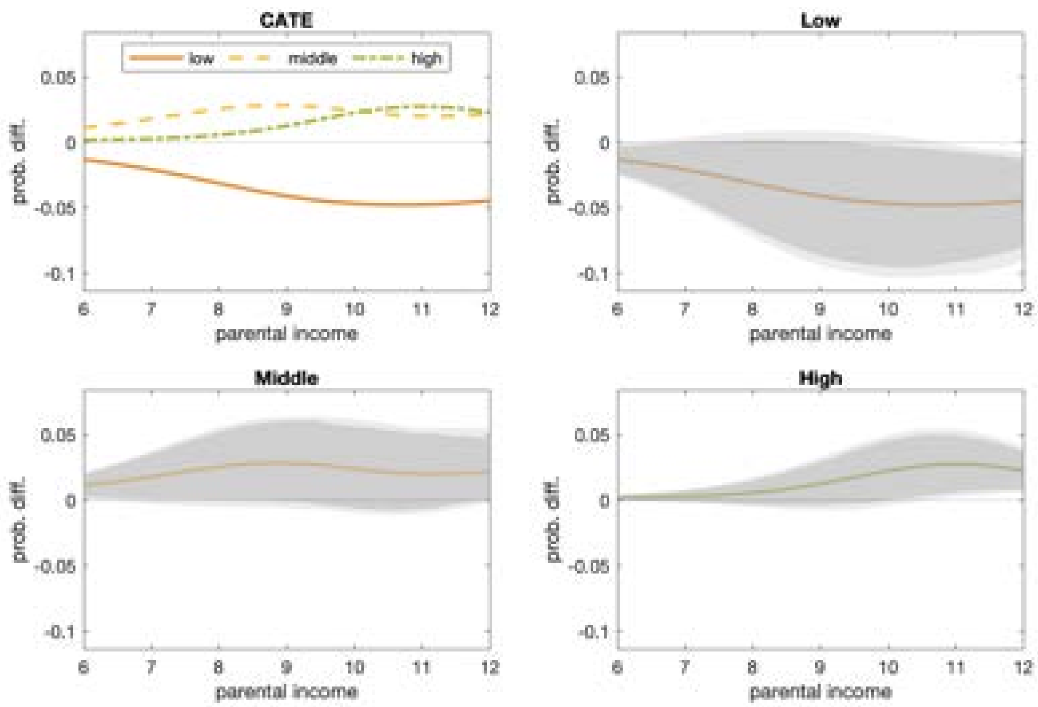
Notes: This figure presents the probability differentials of being in various income classes between male and female children, as functions of parental income, using *parents' average logged income during the period when the child is between 10 and 15 years old (inclusive)* as an alternative measure of parental income. The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 17: Probability Differentials Between Blacks and Non-Blacks, as Functions of Parental Income



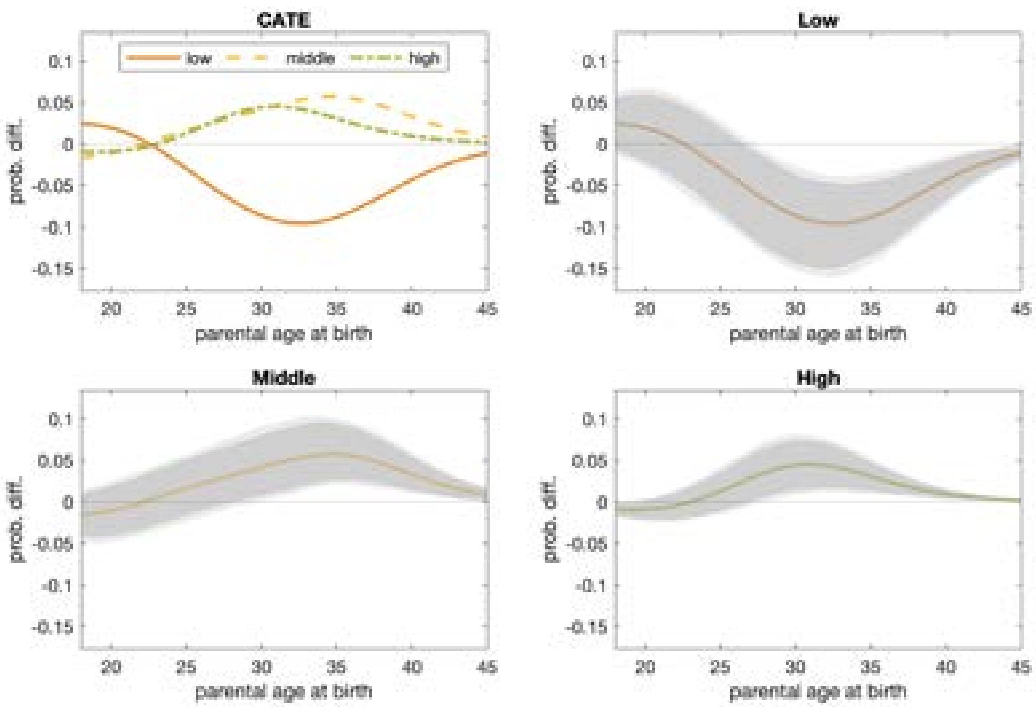
Notes: This figure presents the probability differentials of being in various income classes between Black and non-Black children, as functions of parental income, using *parents' average logged income during the period when the child is between 10 and 15 years old (inclusive)* as an alternative measure of parental income. The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 18: Probability Differentials Between Children with Parents as College and Non-college Graduates, as Functions of Parental Income



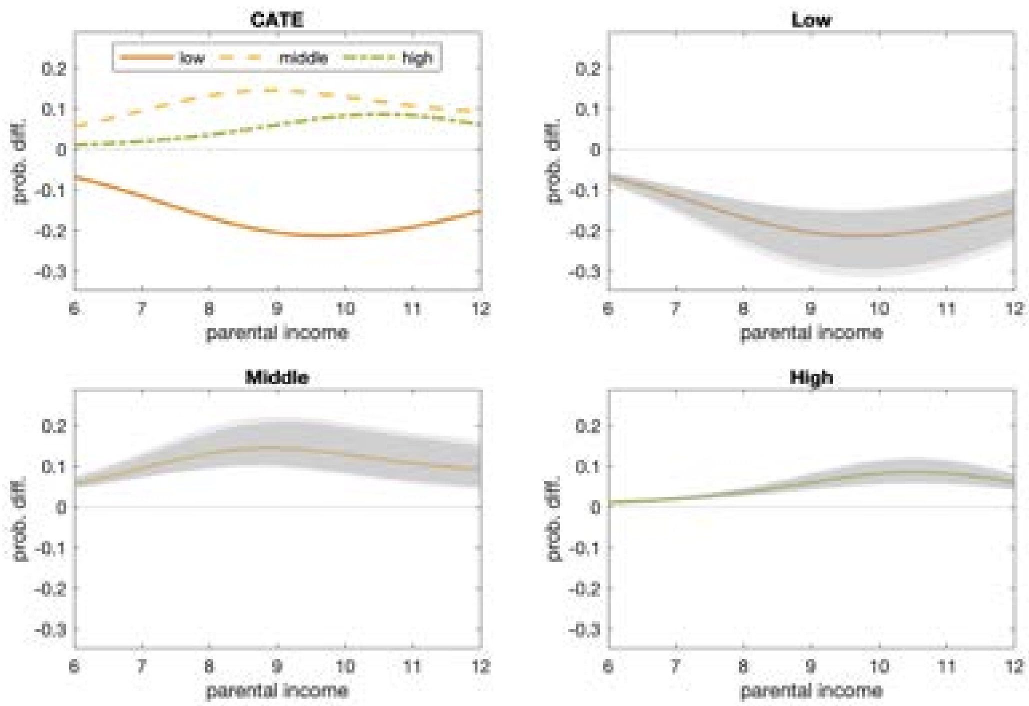
Notes: This figure presents the probability differentials of being in various income classes between children whose parents have a college degree and those whose parents do not, as functions of parental income, using *parents' average logged income during the period when the child is between 10 and 15 years old (inclusive)* as an alternative measure of parental income. The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 19: Probability Differentials Between Children with Parents as College and Non-College Graduates, as Functions of Parental Age at Childbirth



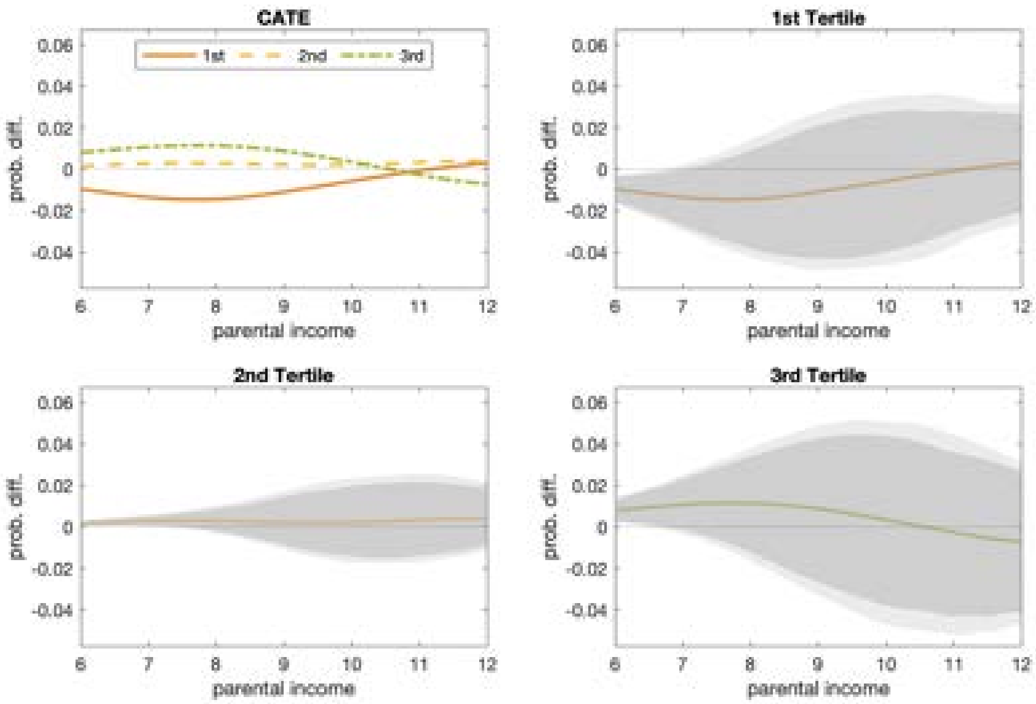
Notes: This figure presents the probability differentials of being in various income classes between children whose parents have a college degree and those whose parents do not, as functions of parental age at childbirth, using *parents' average logged income during the period when the child is between 10 and 15 years old (inclusive)* as an alternative measure of parental income. The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 20: Probability Differentials (Multiple Treatments), as Functions of Parental Income



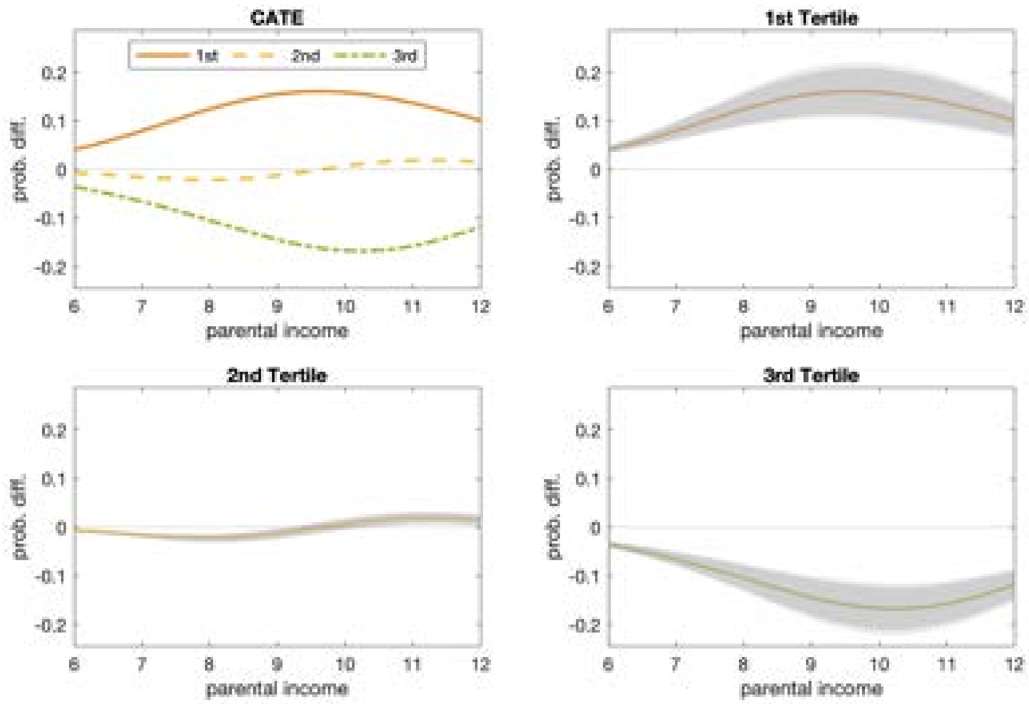
Notes: This figure presents the probability differentials of being in various income classes between children from non-Black families whose parents have a college degree and were aged 30 at childbirth, and those from Black families whose parents do not have a college degree and were aged 18 at childbirth, as functions of parental income, using *parents' average logged income during the period when the child is between 10 and 15 years old (inclusive)* as an alternative measure of parental income. The upper-left panel displays the probability differentials of being in the low-, middle-, and high-income classes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 21: Probability Differentials Between Males and Females, as Functions of Parental Income



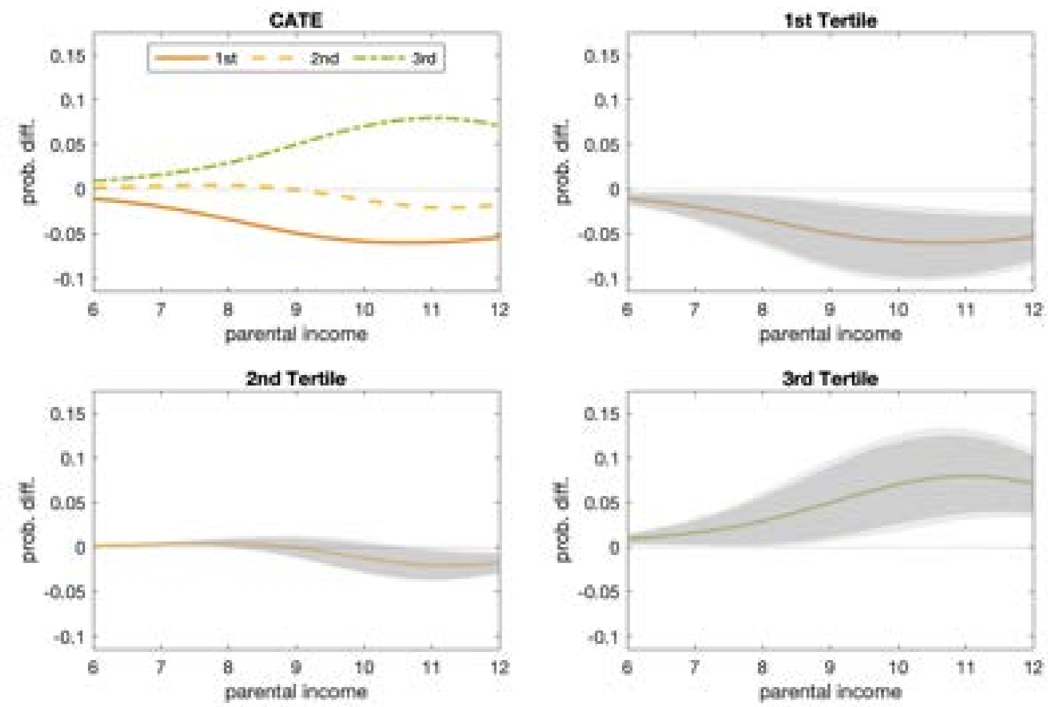
Notes: This figure presents the probability differentials of being in various income classes between male and female children, as functions of parental income, *with income groups classified using tertiles*. The upper-left panel displays the probability differentials of being in the 1st-, 2nd-, and 3rd-tertiles of incomes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 22: Probability Differentials Between Blacks and Non-Blacks, as Functions of Parental Income



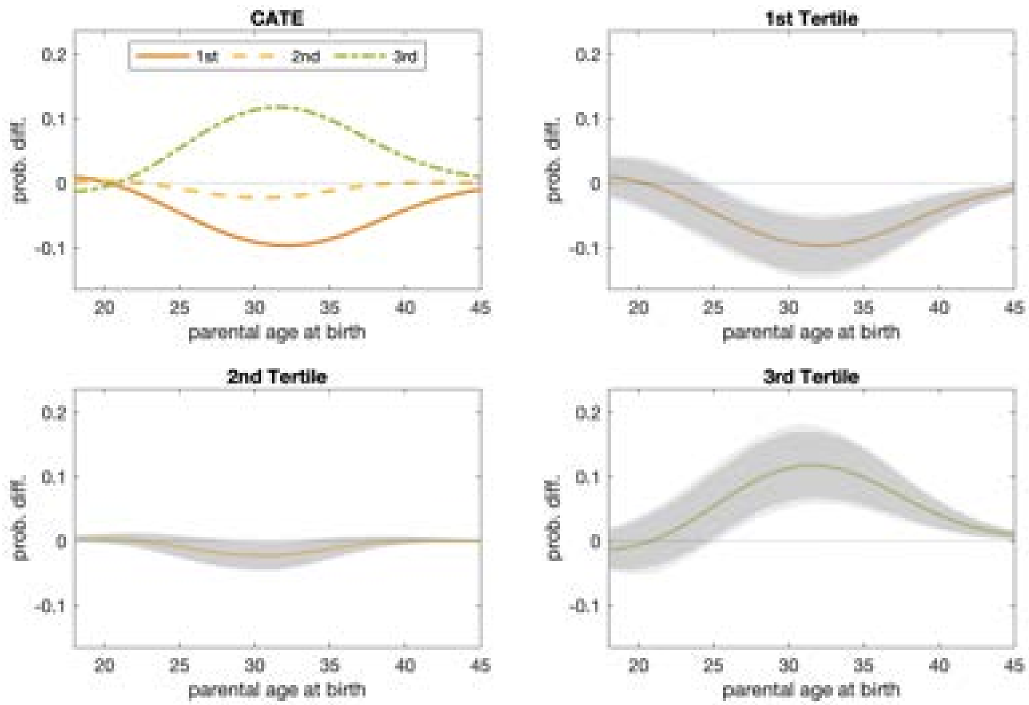
Notes: This figure presents the probability differentials of being in various income classes between Black and non-Black children, as functions of parental income, *with income groups classified using tertiles*. The upper-left panel displays the probability differentials of being in the 1st-, 2nd-, and 3rd-tertiles of incomes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 23: Probability Differentials Between Children with Parents as College and Non-college Graduates, as Functions of Parental Income



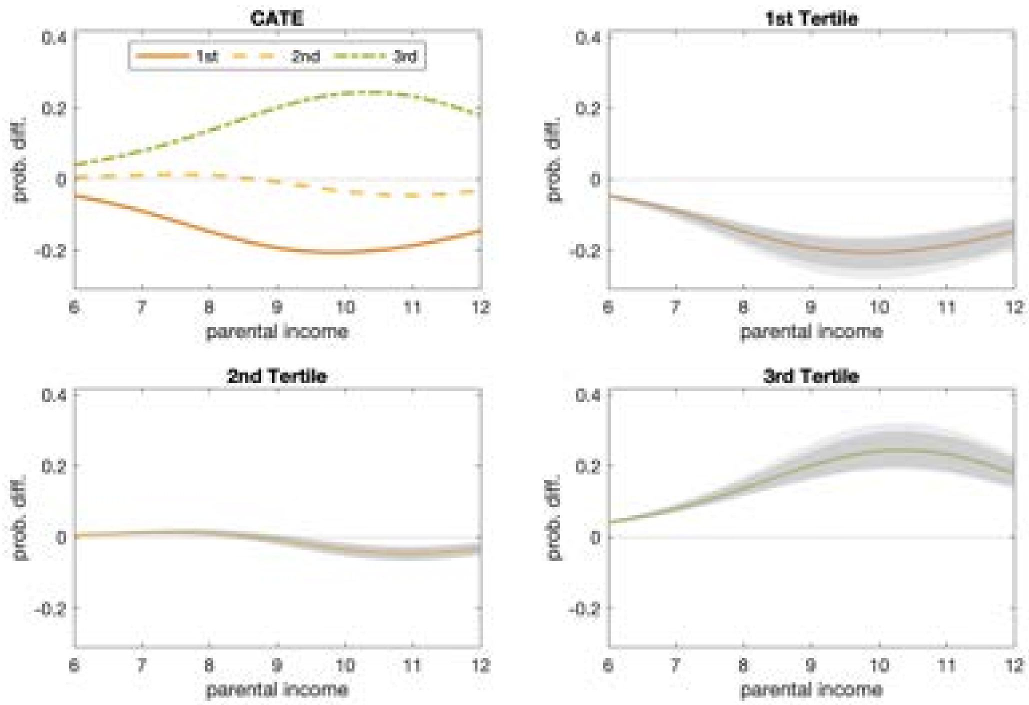
Notes: This figure presents the probability differentials of being in various income classes between children whose parents have a college degree and those whose parents do not, as functions of parental income, *with income groups classified using tertiles*. The upper-left panel displays the probability differentials of being in the 1st-, 2nd-, and 3rd-tertiles of incomes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 24: Probability Differentials Between Children with Parents as College and Non-College Graduates, as Functions of Parental Age at Childbirth



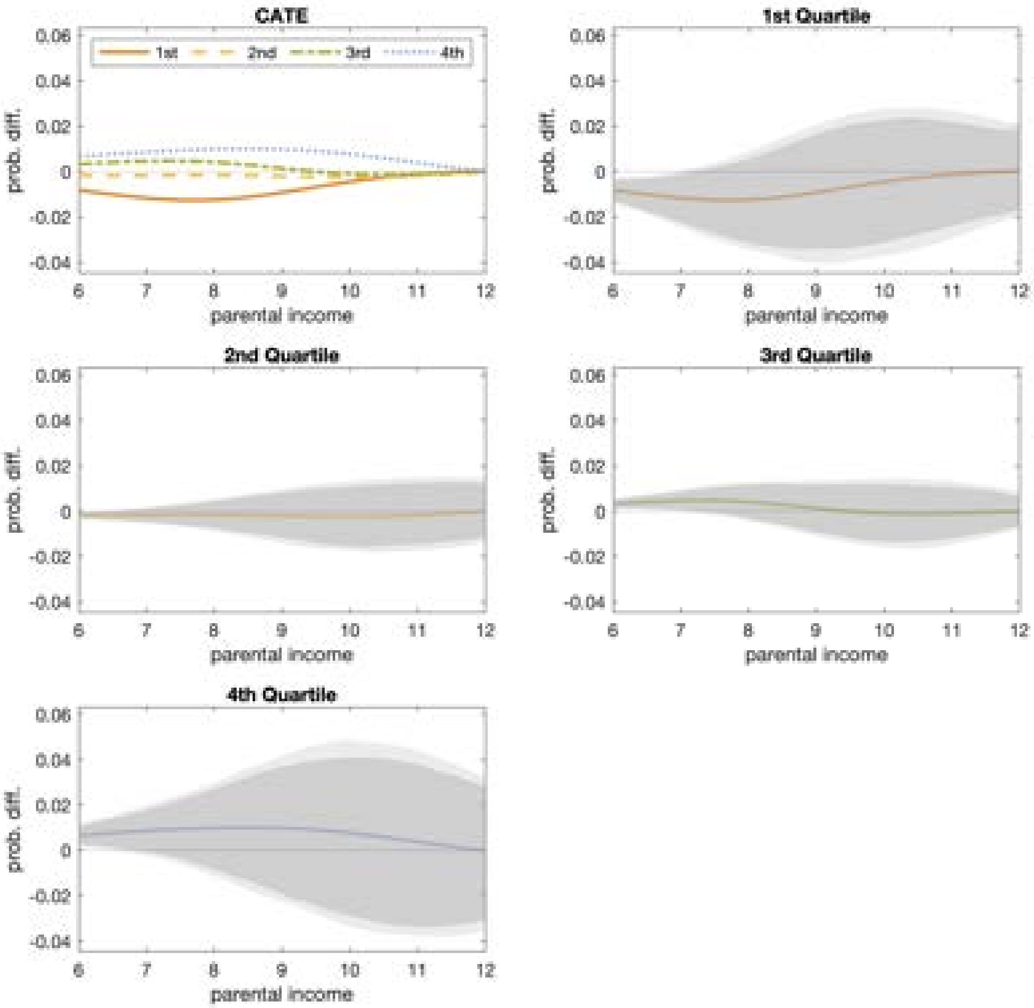
Notes: This figure presents the probability differentials of being in various income classes between children whose parents have a college degree and those whose parents do not, as functions of parental age at childbirth, *with income groups classified using tertiles*. The upper-left panel displays the probability differentials of being in the 1st-, 2nd-, and 3rd-tertiles of incomes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 25: Probability Differentials (Multiple Treatments), as Functions of Parental Income



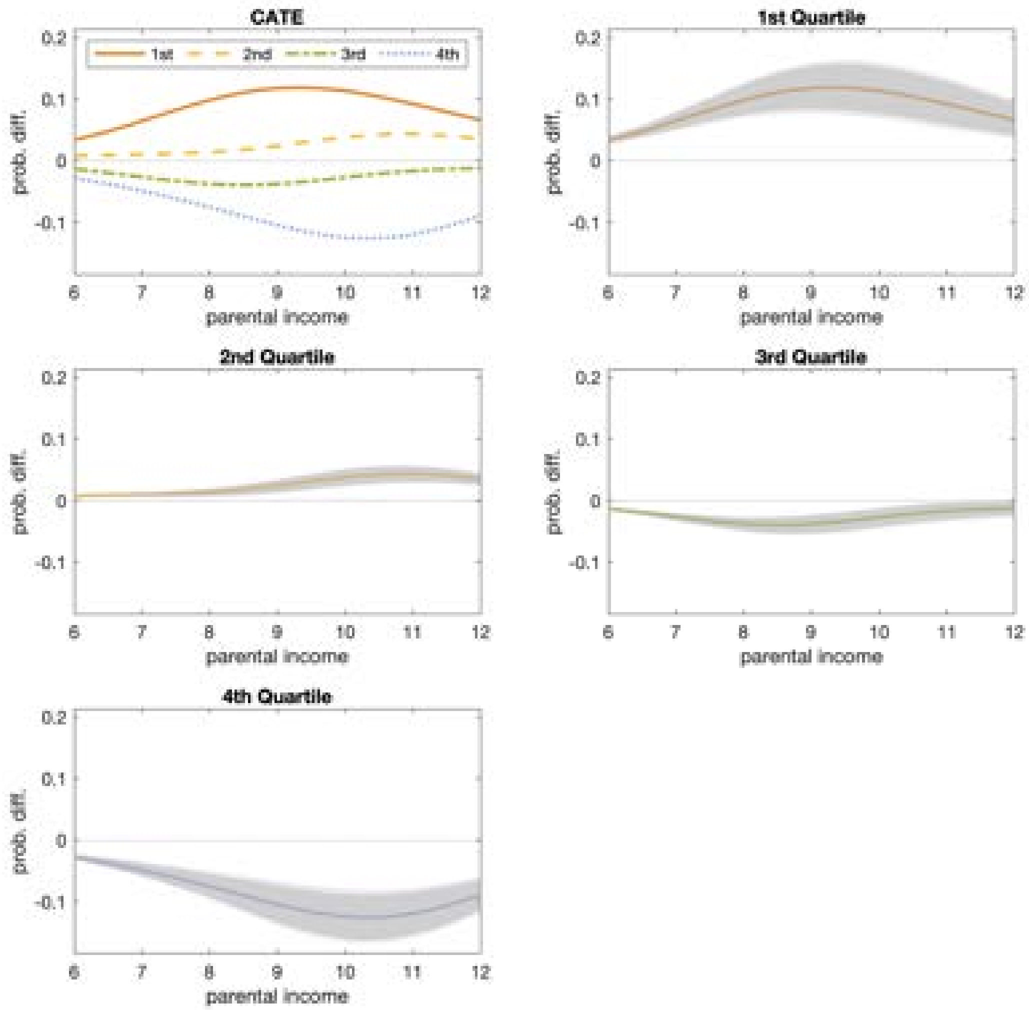
Notes: This figure presents the probability differentials of being in various income classes between children from non-Black families whose parents have a college degree and were aged 30 at childbirth, and those from Black families whose parents do not have a college degree and were aged 18 at childbirth, as functions of parental income, *with income groups classified using tertiles.* The upper-left panel displays the probability differentials of being in the 1st-, 2nd-, and 3rd-tertiles of incomes within one single plot. The upper-right, lower-left, and lower-right panels depict the three differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 26: Probability Differentials Between Males and Females, as Functions of Parental Income



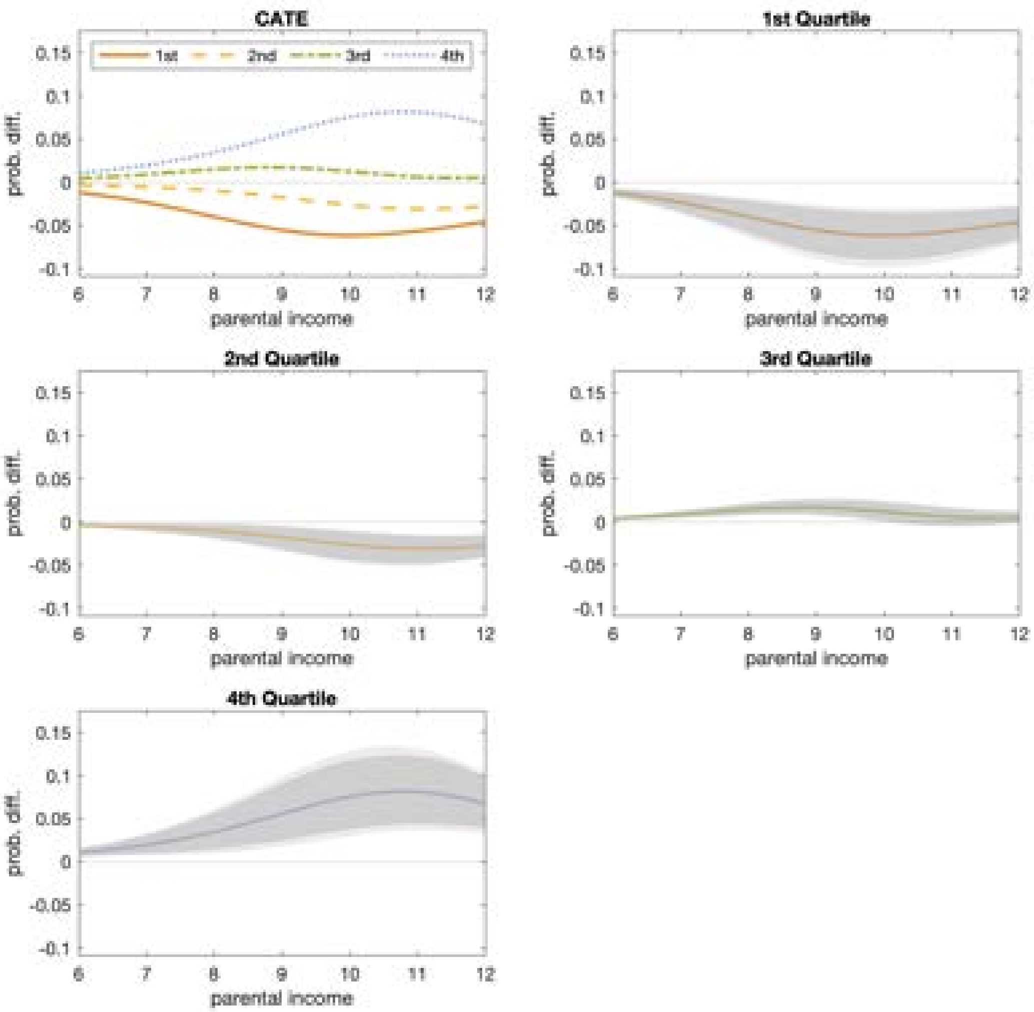
Notes: This figure presents the probability differentials of being in various income classes between male and female children, as functions of parental income, *with income groups classified using quartiles.* The upper-left panel displays the probability differentials of being in the 1st-, 2nd-, 3rd-, and 4th-quartiles of incomes within one single plot. The rest panels depict the four differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 27: Probability Differentials Between Blacks and Non-Blacks, as Functions of Parental Income



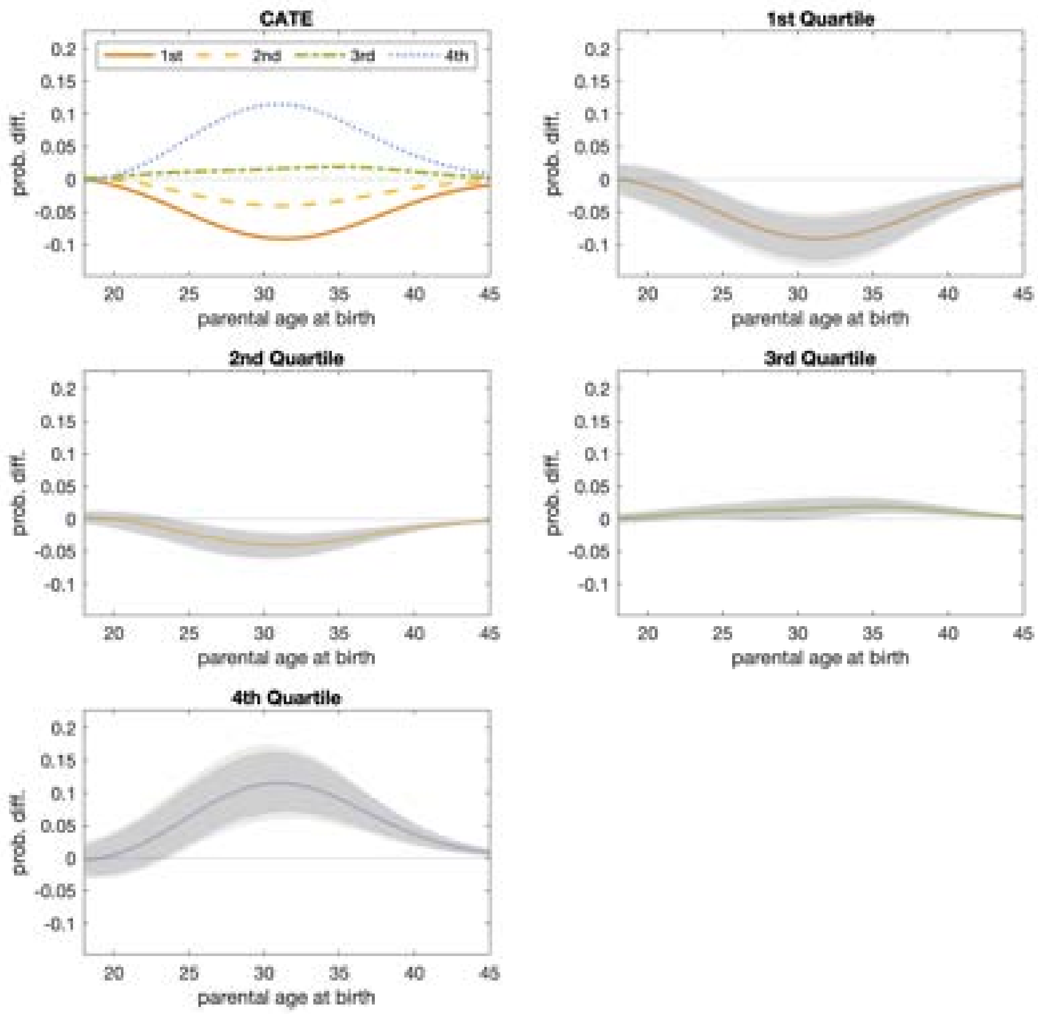
Notes: This figure presents the probability differentials of being in various income classes between Black and non-Black children, as functions of parental income, *with income groups classified using quartiles.* The upper-left panel displays the probability differentials of being in the 1st-, 2nd-, 3rd-, and 4th-quartiles of incomes within one single plot. The rest panels depict the four differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 28: Probability Differentials Between Children with Parents as College and Non-college Graduates, as Functions of Parental Income



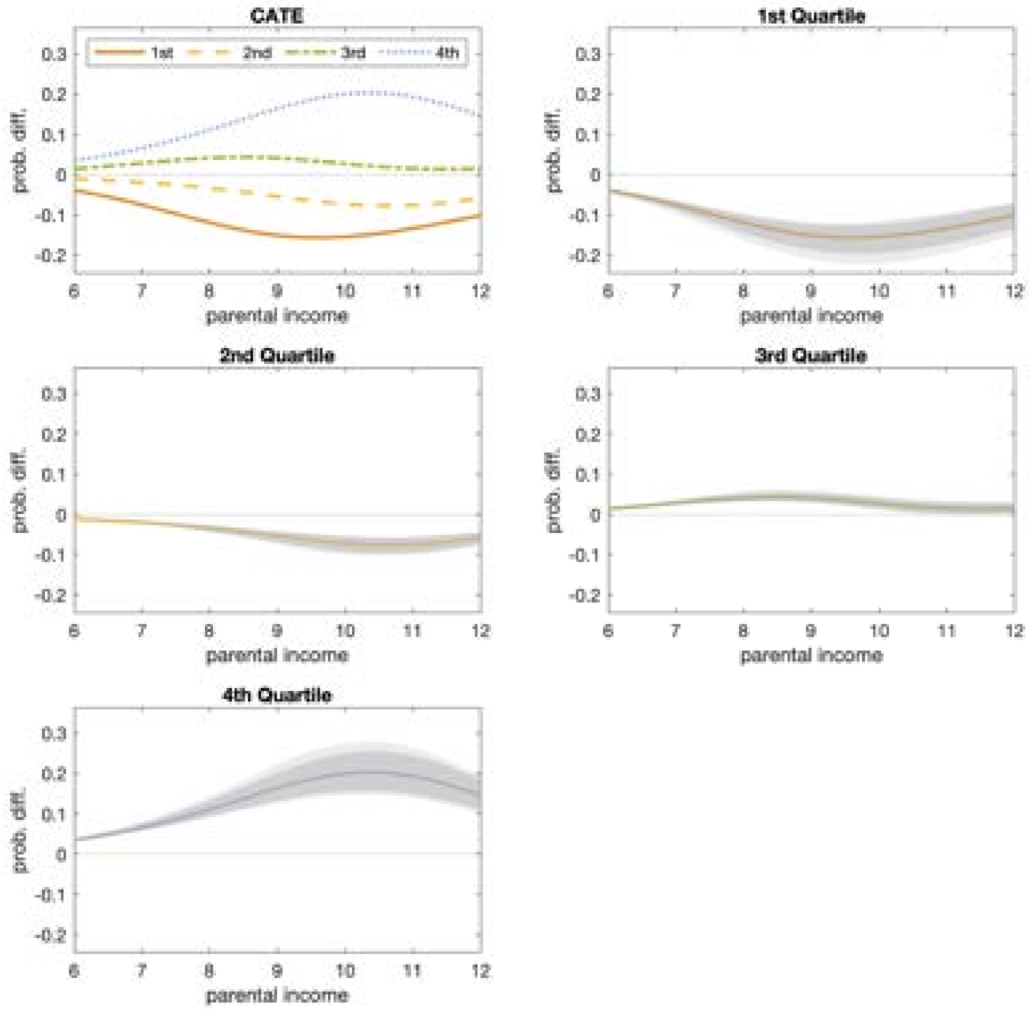
Notes: This figure presents the probability differentials of being in various income classes between children whose parents have a college degree and those whose parents do not, as functions of parental income, *with income groups classified using quartiles*. The upper-left panel displays the probability differentials of being in the 1st-, 2nd-, 3rd-, and 4th-quartiles of incomes within one single plot. The rest panels depict the four differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 29: Probability Differentials Between Children with Parents as College and Non-College Graduates, as Functions of Parental Age at Childbirth



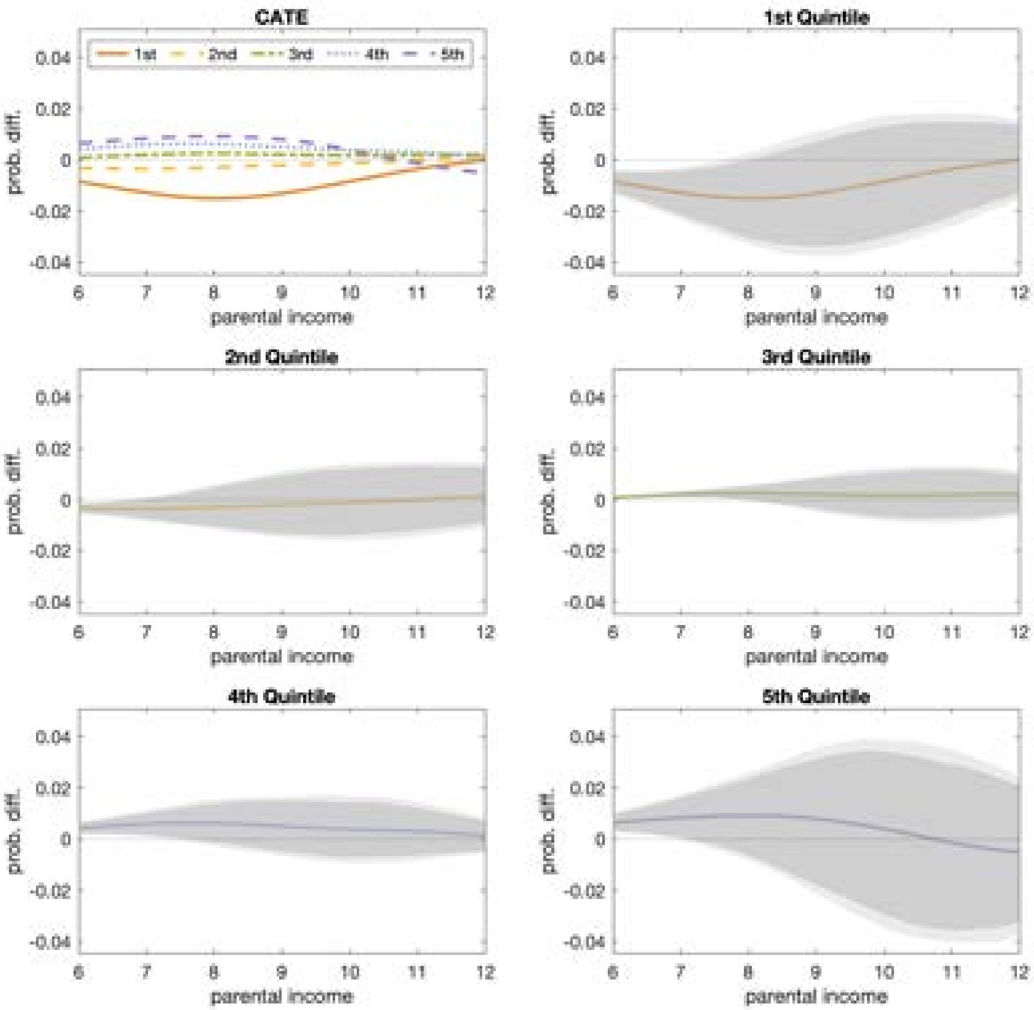
Notes: This figure presents the probability differentials of being in various income classes between children whose parents have a college degree and those whose parents do not, as functions of parental age at childbirth, *with income groups classified using quartiles.* The upper-left panel displays the probability differentials of being in the 1st-, 2nd-, 3rd-, and 4th-quartiles of incomes within one single plot. The rest panels depict the four differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 30: Probability Differentials (Multiple Treatments), as Functions of Parental Income



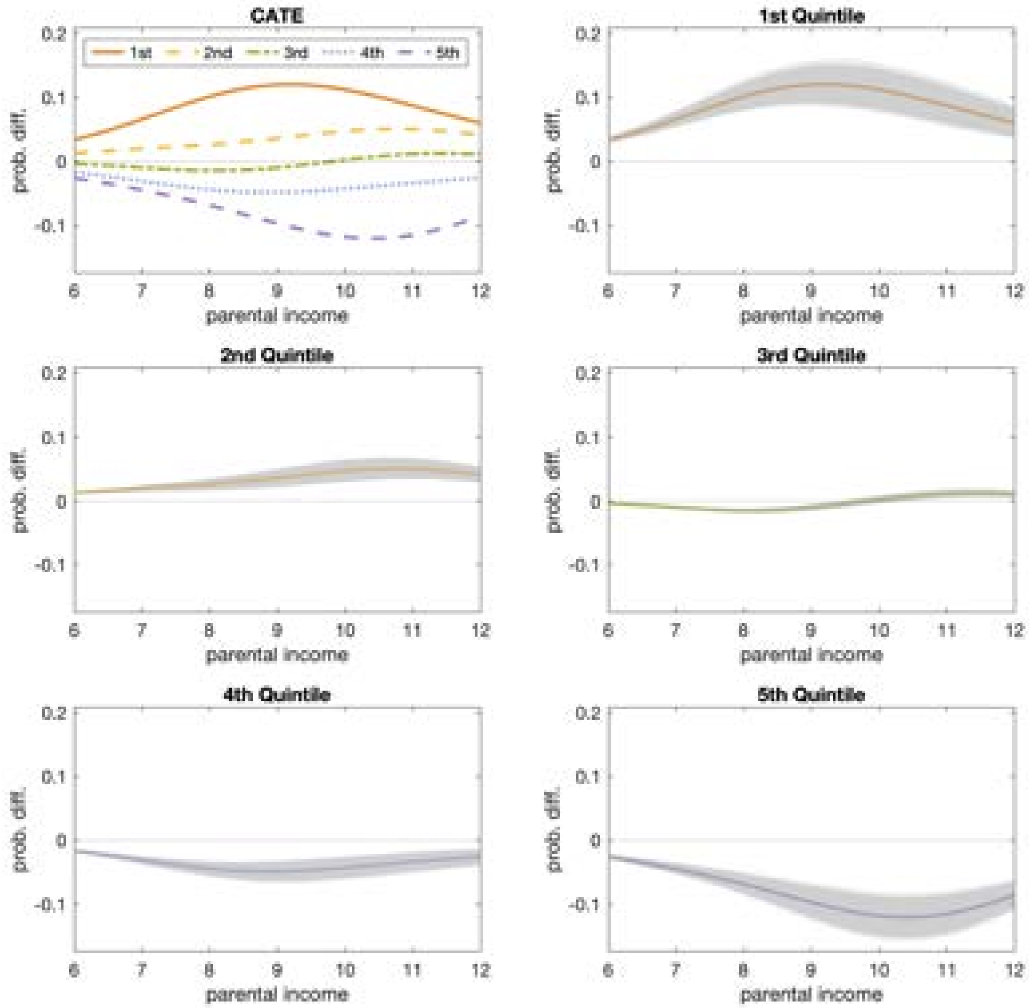
Notes: This figure presents the probability differentials of being in various income classes between children from non-Black families whose parents have a college degree and were aged 30 at childbirth, and those from Black families whose parents do not have a college degree and were aged 18 at childbirth, as functions of parental income, *with income groups classified using quartiles.* The upper-left panel displays the probability differentials of being in the 1st-, 2nd-, 3rd-, and 4th-quartiles of incomes within one single plot. The rest panels depict the four differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 31: Probability Differentials Between Males and Females, as Functions of Parental Income



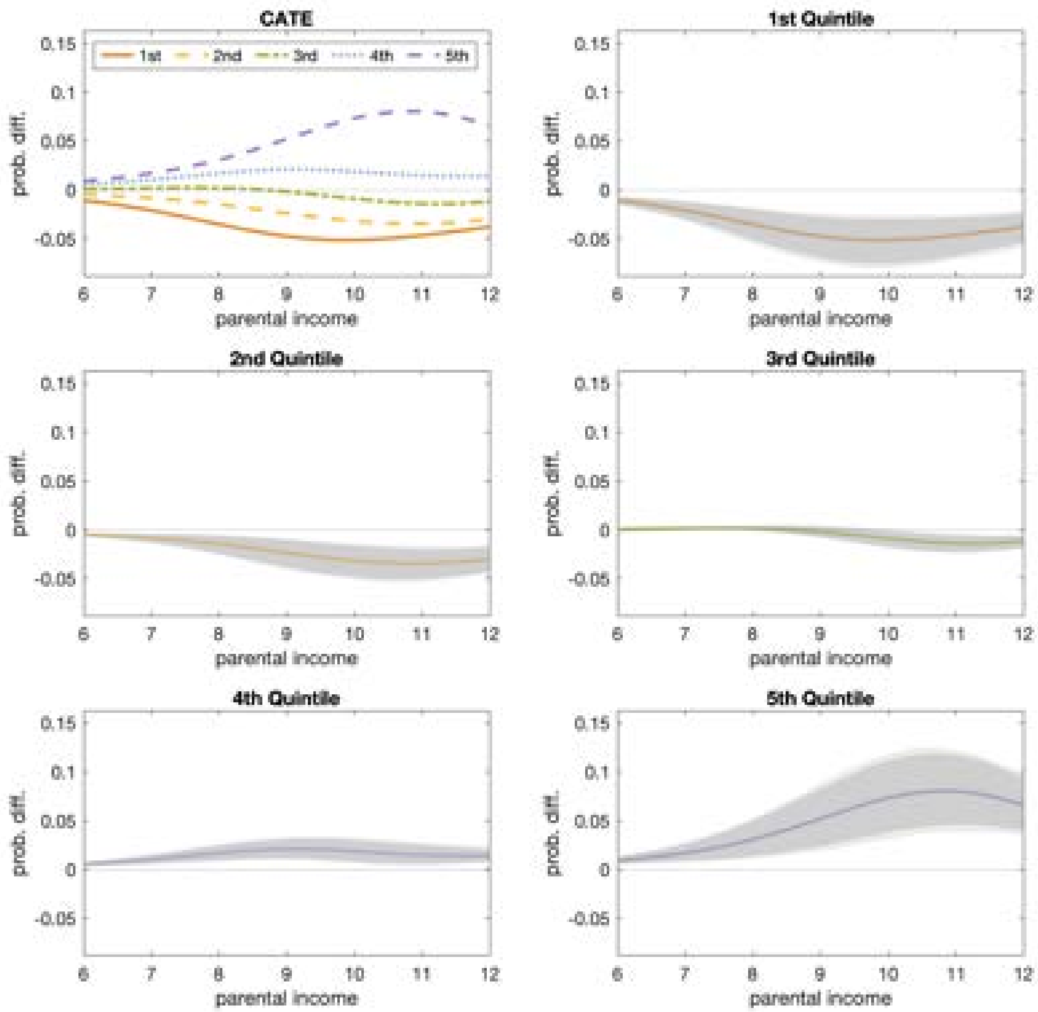
Notes: This figure presents the probability differentials of being in various income classes between male and female children, as functions of parental income, *with income groups classified using quintiles*. The upper-left panel displays the probability differentials of being in the 1st-, 2nd-, 3rd-, 4th-, and 5th-quintiles of incomes within one single plot. The rest panels depict the five differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 32: Probability Differentials Between Blacks and Non-Blacks, as Functions of Parental Income



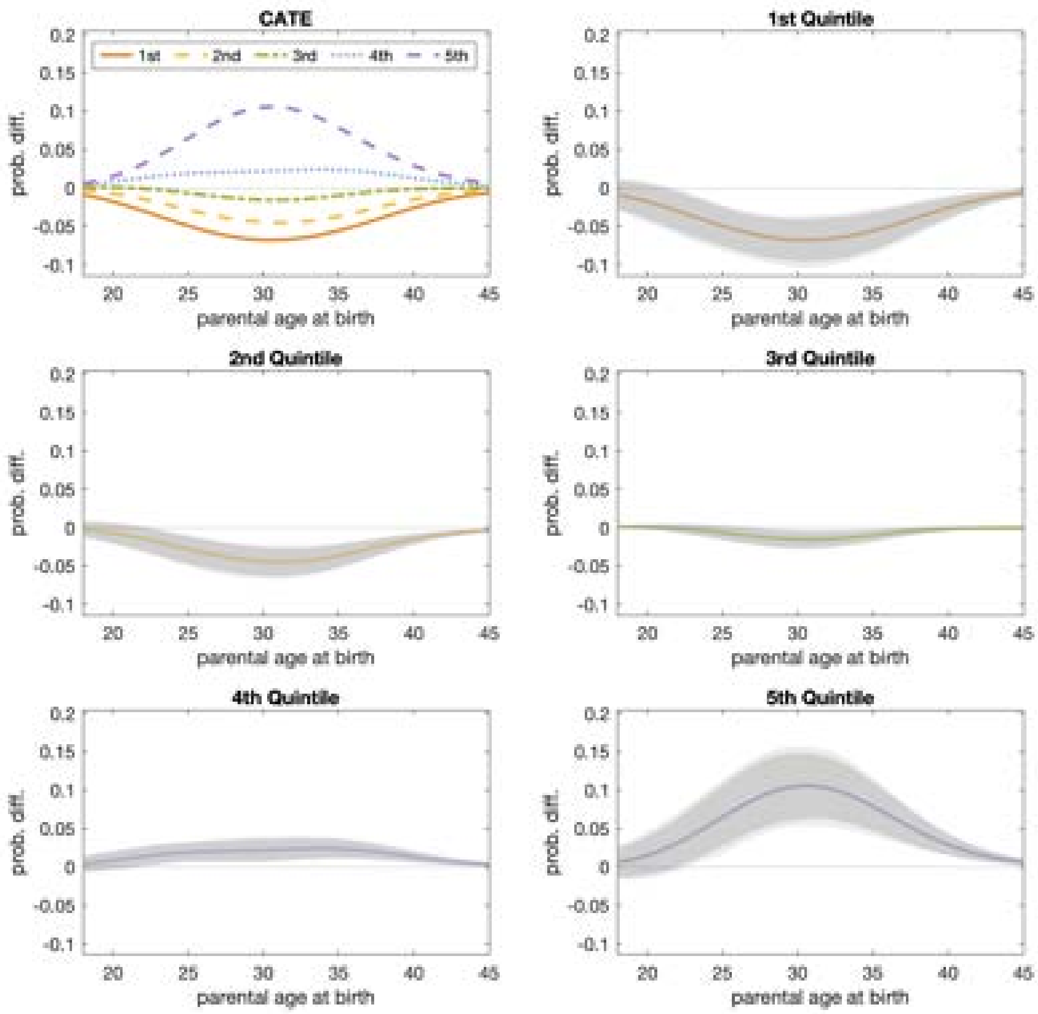
Notes: This figure presents the probability differentials of being in various income classes between Black and non-Black children, as functions of parental income, *with income groups classified using quintiles*. The upper-left panel displays the probability differentials of being in the 1st-, 2nd-, 3rd-, 4th-, and 5th-quintiles of incomes within one single plot. The rest panels depict the five differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 33: Probability Differentials Between Children with Parents as College and Non-college Graduates, as Functions of Parental Income



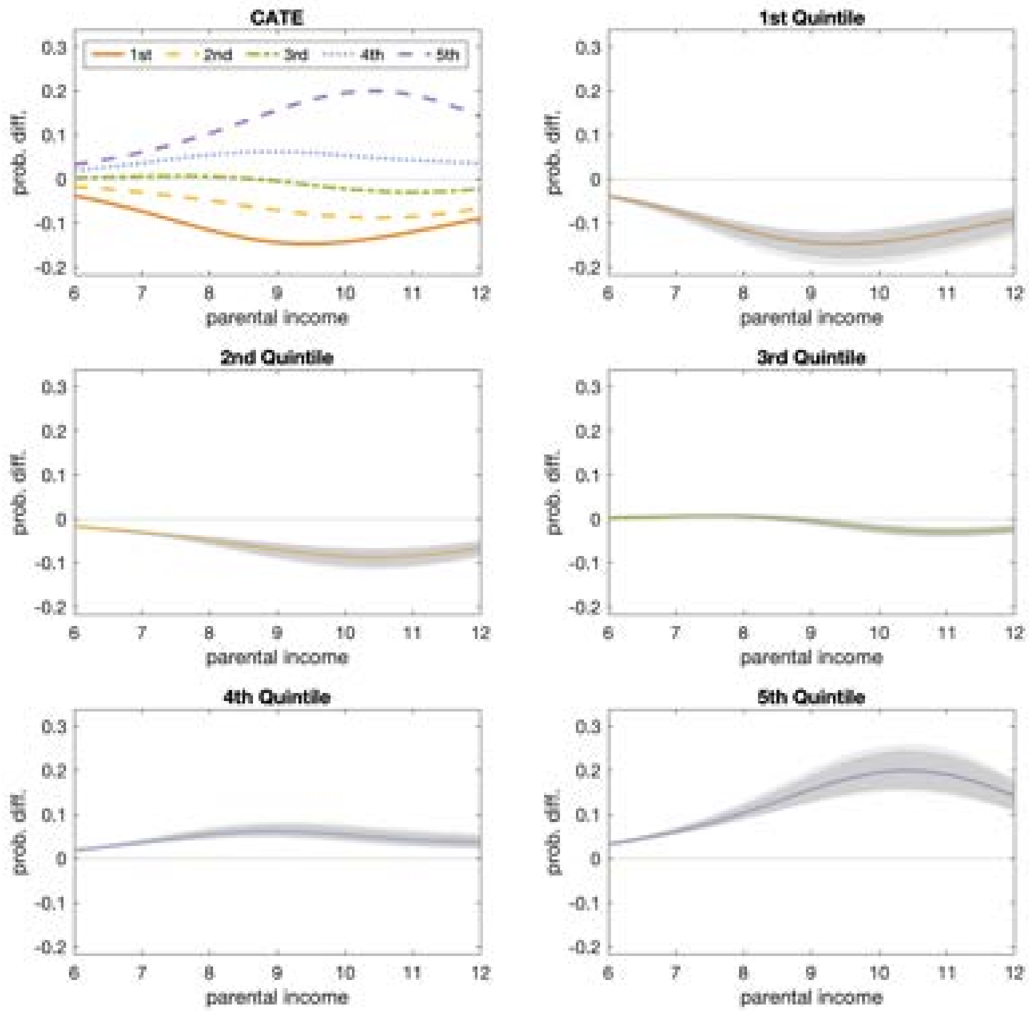
Notes: This figure presents the probability differentials of being in various income classes between children whose parents have a college degree and those whose parents do not, as functions of parental income, *with income groups classified using quintiles.* The upper-left panel displays the probability differentials of being in the 1st-, 2nd-, 3rd-, 4th-, and 5th-quintiles of incomes within one single plot. The rest panels depict the five differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 34: Probability Differentials Between Children with Parents as College and Non-College Graduates, as Functions of Parental Age at Childbirth



Notes: This figure presents the probability differentials of being in various income classes between children whose parents have a college degree and those whose parents do not, as functions of parental age at childbirth, *with income groups classified using quintiles.* The upper-left panel displays the probability differentials of being in the 1st-, 2nd-, 3rd-, 4th-, and 5th-quintiles of incomes within one single plot. The rest panels depict the five differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.

Figure 35: Probability Differentials (Multiple Treatments), as Functions of Parental Income



Notes: This figure presents the probability differentials of being in various income classes between children from non-Black families whose parents have a college degree and were aged 30 at childbirth, and those from Black families whose parents do not have a college degree and were aged 18 at childbirth, as functions of parental income, *with income groups classified using quintiles.* The upper-left panel displays the probability differentials of being in the 1st-, 2nd-, 3rd-, 4th-, and 5th-quintiles of incomes within one single plot. The rest panels depict the five differential curves individually, each accompanied by its 90% and 95% pointwise confidence bands delineated by dark and light gray areas, respectively.