WORKING PAPER · NO. 2023-36

A Trajectories-Based Approach to Measuring Intergenerational Mobility

Yoosoon Chang, Steven N. Durlauf, Seunghee Lee, and Joon Y. Park

MARCH 2023

Becker Friedman Institute
FOR ECONOMICS AT UCHICAGO

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

A TRAJECTORIES-BASED APPROACH TO MEASURING INTERGENERATIONAL MOBILITY

Yoosoon Chang
Steven N. Durlauf
Seunghee Lee
Joon Y. Park

March 2023

The authors are grateful for useful feedback from participants of seminars at Emory University, Indiana University, Kyiv School of Economics, University of Southern Illinois, Sciences Po, Seoul National University, Sungkyunkwan University, Korea Development Institute, Korea Labor Institute, 2019 Jordan River Conference, 2022 International Symposium on Econometric Theory and Applications, 2022 Meeting of the Midwest Econometrics Group, and 2023 American Economic Association meetings. We are also grateful to Neil Cholli, Kristina Butaeva, Youngjoo Cha, Aleksandra Lukina, Oscar Volpe, Ruli Xiao, and Runzhong Xu for helpful comments. Chang and Durlauf thank the James M. and Cathleen D. Stone Foundation for financial support.

A Trajectories-Based Approach to Measuring Intergenerational Mobility
Yoosoon Chang, Steven N. Durlauf, Seunghee Lee, and Joon Y. Park
March 2023
JEL No. C13,J62

ABSTRACT

This paper develops an approach to intergenerational mobility in which the trajectories of parental incomes during childhood and adolescence are the conditioning objects for characterizing dependence across generations. We use functional regression methods to produce an intergenerational elasticity curve that measures how marginal changes in income at each age affect expected offspring permanent income. Using the PSID, estimates of this curve exhibit near monotonicity with respect to age, so that parental incomes in middle childhood and adolescence have larger marginal effects than incomes in early childhood. When interactions are allowed to occur between incomes at different ages, we find a complex pattern of substitutability between incomes at ages that are close in time versus complementarity between parental incomes for ages early childhood and adolescence. Qualitatively similar results hold for offspring education while we do not find evidence of age-specific effects for occupation. We conclude that important information about the links between parental incomes and children exists beyond the scalar characterization of parental permanent income.

Yoosoon Chang
Department of Economics
Indiana University
100 S. Woodlawn Avenue
Bloomington, IN 47405-7104
yoosoon@indiana.edu

Steven N. Durlauf
University of Chicago
Harris School of Public Policy
1307 E. 60th Street
Chicago, IL 60637
and NBER
durlauf@gmail.com

Seunghee Lee
Korea Development Institute
263 Namsejong-ro
Sejong-si 30149
Korea
shlee@kdi.re.kr

Joon Y. Park
Department of Economics
Indiana University
100 S. Woodlawn Avenue
Bloomington, IN 47405-7104
joon@indiana.edu

1 Introduction

This paper proposes and implements a trajectories-based approach to the empirical analysis of intergenerational mobility. We measure how the complete trajectories of family incomes experienced through childhood and adolescence predict adult outcomes and demonstrate that there is important information in these trajectories that is missed when parental permanent income is used as a summary statistic for income influenced. We show how interactions between incomes at different ages can be incorporated into mobility analyses and find patterns of both complementarity and substitutability between parental incomes at different offspring ages.

The study of how income trajectories matter for mobility, as opposed to permanent income, may be justified on numerous grounds. First, credit constraints can imply that the timing of parental income matters for investments in children Lochner and Monge-Narano (2012) for a survey and Caucutt and Lochner (2020) for an important recent contribution. Second, the modern skills literature (cf., Cunha and Heckman (2008), Heckman and Mosso (2014)) has emphasized the dynamic influences that produce adult cognitive and noncognitive abilities, so that the investments at one age affect the marginal value of investments at others. Third, the nature of the investments made by parents qualitatively changes across the life course. Wodtke et al. (2011), for example, show the relative influences of family versus neighborhood characteristics change across the life course with neighborhoods becoming more influential in adolescence. Such dynamics imply a special role of adolescent incomes in terms of the ability of parents to locate their families in higher quality neighborhoods. While our work involves reduced form relationships and so does not directly speak to mechanisms, it provides empirical contours that claim about mechanisms, and more structural analyses should produce.

By characterizing intergenerational mobility and persistence as properties of trajectories, we provide an intergenerational elasticity of income (IGE) or mobility curve that links parental income at each offspring age to their future permanent income. The IGE curve is a generalization of the standard regression of offspring permanent income against parental permanent income, as the latter imposes that the IGE curve is a constant function. We further explore a quadratic trajectories model that interacts between income values of trajectories at different times that way of measuring dynamic complementarities and substitutes in income effects. Our analysis employed methods from Functional Data Analysis (FDA), The FDA approach we employ is based on functional principal components, which fulfills certain optimality properties relative to other methods of measuring age-specific income effects such as linear regressions of offspring permanent income on parental age-specific incomes. We will argue that conventional approaches to measuring age-specific parental income effects produce, in comparison, unstable and imprecise results.

We apply this approach to US data taken from the Panel Study of Income Dynamics. Our results fall into four main categories.

First, we find clear evidence a child's permanent income is sensitive to the timing of parental income. The standard permanent income to permanent income model thus loses information about the mobility process. This timing sensitivity follows a robust pattern: parental incomes in middle and late adolescence are associated with larger marginal effects on predicted offspring income than earlier parental income years. In fact, until 18, a standard milestone in parent child relations, the marginal effects of income appear to be monotonically increasing. Our upward sloping IGE curve has the important implication that children from families with rising income trajectories have higher expected permanent incomes than others whose parents have the same permanent incomes but whose trajectories exhibit lower growth. Our IGE curve estimates are robust to trend-cycle decompositions of parental income to isolate parental permanent income and to the use of instrumental variables methods to isolate shocks to parental income associated with unpredicted changes in employment.

Second, we identify some evidence that the mobility process has changed across cohorts defined as births in 1967-1970, 1971-1973, and 1974-1977. We find evidence that the effects of a permanent parental income starting at birth are larger for the earliest cohort compared to the two later ones. This is due to the fact that sensitivity of offspring to income in later childhood and adolescence seems to have declined relative to the first cohort. Each curve preserves the upward shape identified for the sample as a whole.

Third, when expanding the analysis to a quadratic trajectory model of mobility, we find evidence that interactions between incomes at different child ages matter. The interactions exhibit both substitutability and complementarity. If one focuses on early years (ages 0-5) or later adolescence (15-19) in isolation, substitutability between incomes at different ages occurs within these intervals. In contrast, complementarities exist between incomes across earlier and later years. This pattern suggests a need to think about complements and substitutes as functioning at different time scales. This result is surprising given claims of uniform complementarity between investments in the formation of skills during childhood and adolescence and demonstrates, at a minimum, that one cannot extrapolate from claims of dynamic skills complementarities to claims about the optimal timing of income transfers to the disadvantaged.

Fourth, we study the relationship between family income trajectories to two of the essential proximate causes of offspring income: education and occupation. Here we find very similar results for education to our income regressions. The evidence on occupations in contrast is weak at best; while there is some very slight upward slope in the coefficient curve linking incomes to the probability of a child having a skilled job versus an unskilled one, the confidence bands around the curve are quite wide. Further, the slope of the estimated IGE curve influencing probability of having a semi-skilled job versus unskilled job is negative, although with little precision to the estimates.

We are not the only ones to apply FDA to mobility analysis. Durlauf et al. (2021) apply functional data analysis to the PSID, independently from ours. They find similar results on the effects of later income on children.[1] Our work differs from theirs in significant respects. First, our focus on the joint income determinants of education and occupation provides a substantial generalization to income-based studies. Second, our expansion of FDA methods allows for interactions between incomes at different ages and produces new substantive findings. Third, our use of the functional principal component approach developed in Chang et al. (2021) means our estimates possess various optimality properties and that confidence bands may be generated via bootstrap methods. This all said, our paper should be regarded as complementary to these others, as the functional regression approach to mobility is in its infancy.

Important predecessors to our work exist in terms of the general substantive questions addressed. In terms of timing effects of parental income, Duncan et al. (2010) also use the PSID. This paper divides family income into three periods: prenatal to age 5, ages 6-10, and ages 11-15. Unlike us, the analysis concludes that early childhood income plays a dominant role in future offspring income. The analysis includes very different controls from ours, such as parental expectations, so their analysis is more about identifying differences in marginal effects of income at different ages than in the construction of intergenerational income mobility statistics per se. Carneiro et al. (2021) engage in a similar analysis using Norwegian data with discounted family incomes calculated for 0-5, 6-11, and 12-17. These authors find that early and later periods of income has stronger effects than middle years for both offspring income and offspring years of schooling. An important feature of this work is the use of nonparametric methods to estimate income effects, which allows for flexible interactions between age intervals, and concludes there is complementarity between them.[2] Like us, they also consider offspring education and have corroborative results. We note both these papers impose common sensitivities to incomes with the time intervals studied while our approach places no restrictions on the implied IGE curve that is estimated. The nonparametric approach in Carneiro et al. (2021) allows for interactions between incomes in

[1] Chaussé et al. (2021) also use FDA to relate parental and offspring incomes over five periods. However, they include the age profiles of parents and their children as controls and do not report age effects.

[2] Their approach may be more accurately described as being partially nonparametric since they assume linearity for the effects of control variables.

all three periods, while our quadratic model imposes interactions between pairs of incomes at different ages and does not impose homogeneity with respect to age intervals.[3]

Cheng and Song (2019) is closest in spirit to our work, using the PSID to link characteristics of parental income trajectories to characteristics of offspring income trajectories. This paper takes a fundamentally different approach to ours in that it treats offspring income as a quadratic function of time, with family-specific heterogeneity in coefficients. These coefficients are then allowed to measure intergenerational mobility through the dependence structure of the coefficients across generations. A unique feature of this approach is that they compare parental and offspring trajectories while we map parental trajectories to offspring permanent income.

The paper is organized as follows. Section 2 describes our approach to functional regression analysis and compares it with conventional econometric methods used in the intergenerational mobility literature. Section 3 describes the data. Section 4 presents a functional model of intergenerational mobility and provides estimates of a mobility curve that relates parental incomes at each age of childhood and adolescence to future adult permanent income, while Section 5 extends our basic model to study the interaction effect between parental income at different ages on offspring. Section 6 studies educational mobility and occupational mobility as mechanisms underlying intergenerational mobility. Section 7 concludes. The Appendix describes the econometric methods used for our empirical analyses and various robustness checks.

2 Econometric Models and Methodologies

This section presents the econometric models and methodologies used in the paper. Our econometric models involve functional variables, and here we provide some technical details of how to set up and estimate such models.

2.1 Functional Models

The basic functional regression model that underlies this paper is

$$y_i = \alpha + \int_p^q \beta(r) f_i(r) dr + \varepsilon_i, \tag{1}$$

[3]Our approach can, in principle, allow for fully general interactions at all different ages and as such would be far more flexible than Carneiro et al. (2021). However, for the data we study, any nonparametric approach to estimate multi-dimensional covariates would not produce reliable results due to the curse of dimensionality.

where (y_i) is the variable to be explained, (f_i) is a functional covariate, i.e., $f_i \equiv f_i(\cdot)$, and (ε_i) is the regression error for $i = 1, \ldots, n$. The functional parameter $\beta \equiv \beta(\cdot)$, where $\beta(r)$ signifies the marginal response of y_i on $f_i(r)$ at each $r \in [p, q)$ for $i = 1, \ldots, n$. In the integral, r denotes child age, and p and q are the beginning and end ages for which we measure parental influence. Throughout we assume that (ε_i) satisfies the standard conditions for regression errors. For clarity, β and (f_i) are assumed to be defined continuously for all $r \in [p, q)$ although this is not required for our methods.[4]

When (y_i) denotes offspring permanent income and $f_i(r)$ denotes family incomes across childhood and adolescence ages $[p, q)$, this formulation generalizes the standard intergenerational mobility regression as this specification allows for family income at different ages to have different effects on children: the effect of the functional covariate (f_i) around the age r, $[r - \delta/2, r + \delta/2)$, is approximately $\int_{r-\delta/2}^{r+\delta/2} \beta(u)du \approx \delta\beta(r)$ for small $\delta > 0$.

We assume that the functions $\beta(\cdot)$ and $f_i(\cdot)$, $i = 1, \ldots, n$, are all square integrable, and regard them as elements in some Hilbert space H with inner product defined as[5]

$$\langle \beta, f_i \rangle = \int_p^q \beta(r) f_i(r) dr.$$

The Hilbert space H possesses a countable basis. Our model represents a general linear relationship between a scalar variable (y_i) and a functional covariate (f_i). By the Riesz representation theorem, any continuous linear functional from H can be represented as a functional given by an inner product. This implies that, for any continuous linear functional $\ell : H \to R$ given, we can find $\beta \in H$ such that

$$\ell(f) = \langle \beta, f \rangle$$

for all $f \in H$. The regression function in (1) therefore accommodates any continuous linear effect of the functional covariate (f_i) on (y_i). We may therefore interpret our functional regression model (1) as representing a functional regression model $\mathbb{E}(y_i | f_i) = \ell(f_i)$ with the conditional mean of y_i given as a general continuous linear functional of f_i for $i = 1, \ldots, n$.

This linear functional regression model may be generalized to allow for nonlinear effects. In this paper, we work with a quadratic extension of our original model based on the inner

[4] They are applicable also for β and (f_i) defined at discrete values of r, in which case we may just interpret dr in all integrals (representing integrals with respect to the Lebesgue measure) appearing in the paper as representing integrals with respect to the counting measure assigning unit measure only to every integer point in $[p, q)$. None of our analysis qualitatively changes if β and (f_i) are only defined on discrete values of r and the counting measure is used to evaluate integrals with dr.

[5] More precisely, (f_i) are defined as random elements taking values in H. For brevity, however, we use the same notation (f_i) to denote our functional covariate given as either random functions or their realizations.

product between $f_i(\cdot)$ at a given age and linear combinations of $f_i(\cdot)$ at other ages, i.e.

$$\langle f_i, \Gamma f_i \rangle = \int_p^q \int_p^q \Gamma(r,s) f_i(r) f_i(s) ds dr,$$

where $\Gamma : H \to H$ is a linear operator defined by

$$(\Gamma f)(r) = \int_p^q \Gamma(r,s) f(s) ds \tag{2}$$

for all $f \in H$, where $\Gamma(\cdot,\cdot)$ is a bivariate function on $[p,q)^2$.[6] This allows us to extend our functional regression model (1) to

$$y_i = \alpha + \int_p^q \beta(r) f_i(r) dr + \int_p^q \int_p^q \Gamma(r,s) f_i(r) f_i(s) ds dr + \varepsilon_i, \tag{3}$$

so that interactions between parental incomes at different ages can affect future child income.[7] The effect on a child's income (y_i) of parental income defined as a functional covariate (f_i) over an interval $[r - \delta/2, r + \delta/2)$ around child age r is now given by the sum of the linear effect and quadratic effect, viz.,

$$\int_{r-\delta/2}^{r+\delta/2} \beta(u) du + \int_{r-\delta/2}^{r+\delta/2} \int_{r-\delta/2}^{r+\delta/2} \Gamma(u,v) dv du \approx \delta\beta(r) + \delta^2 \Gamma(r,r). \tag{4}$$

Further, the interactive effect on a child's income (y_i) of a functional covariate (f_i), parental income, over intervals $[r - \delta/2, r + \delta/2)$ and $[s - \delta/2, s + \delta/2)$ around child ages r and s for $s \neq r$ is given by

$$2 \int_{r-\delta/2}^{r+\delta/2} \int_{s-\delta/2}^{s+\delta/2} \Gamma(u,v) dv du \approx 2\delta^2 \Gamma(r,s). \tag{5}$$

Therefore, the total effect of a change in $f(r)$ consists of a linear effect, a quadratic effect, and an interaction effect.

The existing literature of intergenerational mobility mostly relies on the standard regressions when attention is focused on the timing of parental income. To compare standard regressions directly with our approach, we let

$$[p,q) = \bigcup_{j=1}^{m} [p_j, q_j)$$

[6]Here we use Γ to denote both an operator (on the left hand side) and a kernel (on the right hand side) to simply notation.

[7]For the identification of operator parameter $\Gamma = \Gamma(\cdot,\cdot)$, we let $\Gamma(r,s) = \Gamma(s,r)$ for all $r, s \in [p,q)$ in (3).

where $\big([p_j, q_j)\big)$ is a partition of $[p, q)$ for $j = 1, \ldots, m$, and define

$$f_{ij} = \frac{1}{q_j - p_j} \int_{p_j}^{q_j} f_i(r) dr \tag{6}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. From this vantage point, conventional approaches are a special case of the functional approach in which regressions (1), and (3) reduce to

$$y_i \approx \alpha + \sum_{j=1}^{m} \beta_j f_{ij} + \varepsilon_i, \tag{7}$$

and

$$y_i \approx \alpha + \sum_{j=1}^{m} \beta_j f_{ij} + \sum_{j=1}^{m} \gamma_j f_{ij}^2 + \sum_{k>j} \gamma_{jk} f_{ij} f_{ik} + \varepsilon_i, \tag{8}$$

where

$$\beta_j = \int_{p_j}^{q_j} \beta(r) dr, \ \gamma_j = \int_{p_j}^{q_j} \int_{p_j}^{q_j} \Gamma(r, s) ds dr, \ \gamma_{jk} = 2 \int_{p_j}^{q_j} \int_{p_k}^{q_k} \Gamma(r, s) ds dr$$

for $i = 1, \ldots, n$ and $j, k = 1, \ldots, m$.[8] Under these assumptions, (1) becomes the conventional regression to study the intergenerational mobility with age varying income effects.[9] Eq. (3) becomes a quadratic analog to our quadratic model below.

The regressions (7) and (8) might seem the natural extension of mobility analysis to trajectories, but this approach suffers from important weaknesses. The trajectories of parents' income reduce infinite-dimensional functions, which implies that any regression employing them as covariates implicitly relies on taking a stance on how one can represent the infinite-dimensional functions by finite-dimensional regressors. The regressions (7) and (8) associated with this approximation turn out to produce imprecise and unstable estimates, which we will demonstrate for the data under study. The recognition that regressions such as (7) and (8) do a poor job of revealing the underlying process (1) and (3) helps explain why results from functional regressions are qualitatively different from conventional regressions (7) and (8).

In the paper, we also use some functional qualitative response models, where the variable (y_i) to be explained is categorical and the corresponding latent variable (y_i^*) is assumed to be given by the functional regression model in (1) with (y_i) replaced by (y_i^*). Functional qualitative response models may be understood in direct analogy to the functional regression model, where a continuous latent variable that produces discrete behaviors.

Functional regression analysis has been an active area of research in statistics and econo-

[8] If (f_i) is observed discretely at $r = 0, \ldots, 19$, one sets dr to be the counting measure, and replaces $q_j - p_j$ by the number of integers in $[p_j, q_j)$ for $j = 1, \ldots, m$.

[9] An important exception is Carneiro et al. (2021), which essentially sets $m = 3$ and uses a nonparametric regression to allow for general nonlinear effects of parental incomes in early, middle, and late childhood.

metrics for over two decades. Tools to estimate (1) and (3) have been developed by many authors including Bosq (2000), Ramsay and Silverman (2005), Hall and Horowitz (2007), Yao and Müller (2010), Park and Qian (2012) and Chang et al. (2021), among others. We will use the approach developed in Chang et al. (2021) for several reasons. First, the framework allows one to compare functional regressions with conventional regression and demonstrates the value-added of the functional regression approach taken here. Second, Chang et al. (2021) provides a bootstrap method to obtain the confidence bands for estimates of various functional parameters, while other approaches only provide consistency and convergence rates. Third, the approach naturally extended to the functional regressions with quadratic terms and the functional qualitative response models which we employ here.[10] Finally, the approach adopted in the paper provides a framework for non-asymptotic analysis which makes it possible to evaluate differences in finite sample behavior for asymptotically equivalent methods. This will be discussed in more detail later.

2.2 Finite-Dimensional Representation

To estimate the functional regressions (1) and (3), we need to represent the functional covariate (f_i) by finite-dimensional vectors. This representation is implemented by choosing an orthonormal basis (v_j) for H and approximating (f_i) as

$$f_i \approx \sum_{j=1}^{m} \langle v_j, f_i \rangle v_j$$

for $i = 1, \ldots, n$ with an appropriately chosen m. Any $f \in H$ can be written as

$$f = \sum_{j=1}^{\infty} \langle v_j, f \rangle v_j \tag{9}$$

using any orthonormal basis (v_j). The choice of the basis (v_j) and the number m of terms in the approximation controls the efficiency of functional coefficient estimates.[11]

Once the orthonormal basis (v_j) and the number m of terms in the approximation are chosen, we may represent a function f in H as an m-dimensional vector in R^m. Formally,

[10] Although asymptotic theory for the functional qualitative response models has not been formally developed in the literature, we are sure that consistency and asymptotic normality established in Chang et al. (2021) hold also for general nonlinear models under suitable regularity conditions.

[11] Orthonormality of the basis (v_j) is not necessary. Our analysis also applies non-orthonormal basis functions such as a spline basis if one orthogonalizes and normalizes the basis sequentially by the Gram-Schmidt process.

we work with a mapping π on H defined as $\pi(f) = \llparenthesis f \rrparenthesis$, where

$$\llparenthesis f \rrparenthesis = \begin{pmatrix} \langle v_1, f \rangle \\ \vdots \\ \langle v_m, f \rangle \end{pmatrix}$$

for any $f \in H$. Denote H_m as the subspace of H spanned by the sub-basis $(v_j)_{j=1}^m$. For any $f \in H_m$, there exists a unique m-dimensional vector $\llparenthesis f \rrparenthesis$. Moreover, for any $f \in H_m$ written as $f = \sum_{j=1}^m c_j v_j$, we have $\llparenthesis f \rrparenthesis = (c_1, \ldots, c_m)'$ and

$$\|f\|^2 = \sum_{j=1}^m c_j^2 = \|\llparenthesis f \rrparenthesis\|^2,$$

where we use the notation $\| \cdot \|$ to denote the Hilbert space norm in H for f, $\|f\|^2 = \langle f, f \rangle$, on the left-hand side and the usual Euclidean norm in R^m for $\llparenthesis f \rrparenthesis$, $\|\llparenthesis f \rrparenthesis\|^2 = \|(c_1, \ldots, c_m)'\|^2$, on the right-hand side. Therefore, the mapping π defines an isometry between H_m and R^m. Since any $f \in H_m$ can be written as $f = \sum_{j=1}^m \langle v_j, f \rangle v_j$, we may easily find $f = \pi^{-1}\big(\llparenthesis f \rrparenthesis\big)$ in H_m for any $\llparenthesis f \rrparenthesis$ given in R^m.

Our representation of a function by a finite-dimensional vector is straightforward. For any $f \in H$, we approximate f by $\Pi_m f$, where Π_m is the projection on H_m, and represent $\Pi_m f$ as an m-dimensional vector $\llparenthesis f \rrparenthesis$ in R^m using an isometry π. The error incurred in approximating f by $\Pi_m f$ is given by $(1-\Pi_m)f$, where $1-\Pi_m$ is the projection on the orthogonal complement of H_m, i.e., the subspace of H spanned by $(v_j)_{j=m+1}^\infty$. See Chang et al. (2021) for formal development of this argument. Note that our approach makes explicit the approximation error incurred in representing infinite-dimensional functions by finite-dimensional vectors.

For our subsequent analysis, we assume (y_i) and (f_i) are already demeaned and redefine y_i and f_i as $y_i - (1/n)\sum_{i=1}^n y_i$ and $f_i - (1/n)\sum_{i=1}^n f_i$ for $i = 1, \ldots, n$, respectively, so that we may let $\alpha = 0$ without loss of generality in our functional regression (1). This is to focus on the estimation of the functional regression coefficient $\beta(\cdot)$. Under this convention, it follows that

$$y_i = \langle \beta, f_i \rangle + \varepsilon_i = \langle \beta, \Pi_m f_i \rangle + \big\langle \beta, (1 - \Pi_m) f_i \big\rangle + \varepsilon_i, \tag{10}$$

which can be approximated as

$$y_i \approx \langle \beta, \Pi_m f_i \rangle + \varepsilon_i. \tag{11}$$

However, we have

$$\langle \beta, \Pi_m f_i \rangle = \left\langle \beta, \sum_{j=1}^{m} \langle v_j, f_i \rangle v_j \right\rangle = \sum_{j=1}^{m} \langle v_j, \beta \rangle \langle v_j, f_i \rangle = \llparenthesis \beta \rrparenthesis' \llparenthesis f_i \rrparenthesis,$$

so the approximate functional regression (11) reduces to a standard regression

$$y_i \approx \llparenthesis \beta \rrparenthesis' \llparenthesis f_i \rrparenthesis + \varepsilon_i, \tag{12}$$

which has the m-dimensional regressor $(\llparenthesis f_i \rrparenthesis)$ with m-dimensional coefficient $\llparenthesis \beta \rrparenthesis$. Using the least squares estimator $\llparenthesis \hat{\beta} \rrparenthesis$ of $\llparenthesis \beta \rrparenthesis$, we immediately obtain an estimate for the functional coefficient $\beta(\cdot)$ by $\hat{\beta}(\cdot) = \pi^{-1}\big(\llparenthesis \hat{\beta} \rrparenthesis\big)$. Therefore, our estimation procedure is simple, once an orthonormal basis (v_j) and truncation number m are chosen.[12]

The same approach can be used to estimate the quadratic functional regression in (3). To see this, define

$$\llparenthesis \Gamma \rrparenthesis = \begin{pmatrix} \langle v_1, \Gamma v_1 \rangle & \cdots & \langle v_1, \Gamma v_m \rangle \\ \vdots & \vdots & \vdots \\ \langle v_m, \Gamma v_1 \rangle & \cdots & \langle v_m, \Gamma v_m \rangle \end{pmatrix}$$

for the linear operator Γ on H introduced in (2). Similarly as before, the mapping π from a linear operator Γ on H_m to an $m \times m$ matrix $\llparenthesis \Gamma \rrparenthesis$ is a one-to-one correspondence. Therefore, we may easily obtain $\Gamma = \pi^{-1}\big(\llparenthesis \Gamma \rrparenthesis\big)$ for any $m \times m$ matrix $\llparenthesis \Gamma \rrparenthesis$ given, since any linear operator Γ on H_m is uniquely defined once $\Gamma v_j = \sum_{i=1}^{m} \langle v_i, \Gamma v_j \rangle v_i$ is specified for all $j = 1, \ldots, m$.

We have

$$\langle f_i, \Gamma f_i \rangle \approx \big\langle \Pi_m f_i, \Gamma \Pi_m f_i \big\rangle,$$

ignoring the approximation error involved in finite dimensional representation of (f_i). It follows that

$$\big\langle \Pi_m f_i, \Gamma \Pi_m f_i \big\rangle = \left\langle \sum_{j=1}^{m} \langle v_j, f_i \rangle v_j, \Gamma \sum_{j=1}^{m} \langle v_j, f_i \rangle v_j \right\rangle = \llparenthesis f_i \rrparenthesis' \llparenthesis \Gamma \rrparenthesis \llparenthesis f_i \rrparenthesis,$$

and therefore, the functional regression with quadratic term (3) yields

$$y_i \approx \llparenthesis \beta \rrparenthesis' \llparenthesis f_i \rrparenthesis + \llparenthesis f_i \rrparenthesis' \llparenthesis \Gamma \rrparenthesis \llparenthesis f_i \rrparenthesis + \varepsilon_i \tag{13}$$

approximately if we set $\alpha = 0$ as before. The regression (13) is a standard regression on the

[12] Discrete choice models with functional covariates can be estimated by approximating the regression for the underlying latent variable (y_i^*) as $y_i^* \approx \llparenthesis \beta \rrparenthesis' \llparenthesis f_i \rrparenthesis + \varepsilon_i$.

covariate $(\llparenthesis f_i \rrparenthesis)$ with its additional quadratic terms. Note that

$$\llparenthesis f_i \rrparenthesis'\llparenthesis \Gamma \rrparenthesis \llparenthesis f_i \rrparenthesis = \mathrm{trace}\left(\llparenthesis \Gamma \rrparenthesis'(\llparenthesis f_i \rrparenthesis \llparenthesis f_i \rrparenthesis')\right) = \left(\mathrm{vec}\,\llparenthesis \Gamma \rrparenthesis\right)'\left(\mathrm{vec}\,(\llparenthesis f_i \rrparenthesis \llparenthesis f_i \rrparenthesis')\right).$$

Therefore, we may easily obtain the least squares estimator $\llparenthesis \hat{\Gamma} \rrparenthesis$ of $\llparenthesis \Gamma \rrparenthesis$ together with the least squares estimator $\llparenthesis \hat{\beta} \rrparenthesis$ of $\llparenthesis \beta \rrparenthesis$. We may recover the corresponding estimators $\hat{\beta}$ and $\hat{\Gamma}$ as $\pi^{-1}\big(\llparenthesis \hat{\beta} \rrparenthesis\big)$ and $\pi^{-1}\big(\llparenthesis \hat{\Gamma} \rrparenthesis\big)$, respectively.

What we have called the conventional regression implements our general approach with a particular choice of basis. To see this, consider the basis (v_j) defined as

$$v_j = \frac{1}{\sqrt{q_j - p_j}} 1\{p_j \leqslant r < q_j\}, \tag{14}$$

where $1\{\cdot\}$ denotes the indicator function. This basis produces

$$\langle v_j, f_i \rangle = \frac{1}{\sqrt{q_j - p_j}} \int_{p_j}^{q_j} f_i(r)dr,$$

and the regressors (f_{ij}) in the conventional regression (7) are therefore given by

$$f_{ij} = \frac{1}{\sqrt{q_j - p_j}} \langle v_j, f_i \rangle$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. We will show that this choice of basis is not optimal and can lead to uninterpretable results.[13]

We do not impose any continuity or smoothness condition for the parameters β and Γ, which are a function and an operator, respectively, in our functional regressions (1) and (3). The continuity or smoothness of our estimates for β and Γ is entirely determined by that of the basis (v_j) we use to represent our functional regressions in (1) and (3) as the standard regressions in (12) and (13), respectively. This is because our estimator $\hat{\beta} = \pi^{-1}\big(\llparenthesis \hat{\beta} \rrparenthesis\big)$ for β is obtained as a linear combination of $(v_j)_{j=1}^m$ with coefficients given by an m-dimensional vector $\llparenthesis \hat{\beta} \rrparenthesis$, and our estimator $\hat{\Gamma} = \pi^{-1}\big(\llparenthesis \hat{\Gamma} \rrparenthesis\big)$ for Γ is obtained similarly as a linear combination of $(v_j \otimes v_k)_{j,k=1}^m$ with coefficients given by an $m \times m$ matrix $\llparenthesis \hat{\Gamma} \rrparenthesis$. For instance, if the basis consisting of indicators (v_j) defined in (14) is used, our estimates for β and Γ are discontinuous with jumps at the boundaries of supports for those indicators. However, they

[13] Carneiro et al. (2021) use this basis with $m = 3$. As will be shown subsequently, however, the choice of this basis is clearly sub-optimal, and we may measure the effect of parental income trajectory (f_i) on intergenerational mobility more efficiently by more carefully choosing a basis. Although not explicitly shown, Chang et al. (2021) makes it clear that we may also use $\big(\llparenthesis f_i \rrparenthesis\big)$ to nonparametrically estimate the effect of (f_i) on (y_i).

become continuous and smooth if we use a basis with elements that are functions defined continuously or smoothly over their domain.[14]

2.3 Optimal Basis: Non-Asymptotic Analysis

For empirical analysis, we employ the functional principal component basis, following Bosq (2000), Ramsay and Silverman (2005), Hall and Horowitz (2007), Yao and Müller (2010), Park and Qian (2012) and Chang et al. (2021), among others.[15] Let

$$Q = \sum_{i=1}^{n}(f_i \otimes f_i), \tag{15}$$

where "$\otimes$" signifies the tensor product defined in the Hilbert space H which reduces to the outer product when H is finite-dimensional.[16] In empirical work, we may consider the functions f and g as long vectors consisting of their values in the ordinate corresponding to a sufficiently fine grid of values in the abscissa, and interpret the tensor product $f \otimes g$ of two functions f and g as their outer product fg', analogously as we may interpret the inner product $\langle f, g\rangle = \int f(r)g(r)dr$ as $f'g$ if we scale r so that $dr = 1$. We let $(\lambda_j^*, v_j^*)_{i=1}^n$ be the set of pairs of eigenvalue and eigenfunction of Q, which are ordered so that $\lambda_1^* > \cdots > \lambda_n^*$. By construction, (i) (v_j^*) are orthonormal, (ii) $\sum_{i=1}^n \langle v_j^*, f_i\rangle\langle v_k^*, f_i\rangle = 0$ for all $j \neq k$, and (iii) $\sum_{i=1}^n \langle v, f_i\rangle^2$ is maximized at $v = v_j^*$ under the constraints $\langle v_k^*, v\rangle = 0$ for all $k < j$ with $\sum_{i=1}^n \langle v_j^*, f_i\rangle^2 = \lambda_j^*$ for $j = 1, \ldots, n$.

The eigenfunctions (v_j^*) thus defined are indeed normalized functional principal components. The functional principal component analysis (FPCA) extends the principal component analysis (PCA) to analyze variations in functional observations. If (f_i) are finite-dimensional, Q reduces to $\sum_{i=1}^n f_i f_i'$. The leading principal component of (f_i), in this case, is given by the eigenvector of Q associated with the largest eigenvalue, and it represents the linear combination of (f_i), $\sum_{i=1}^n c_i f_i$ with $\sum_{i=1}^n c_i^2 = 1$, whose Euclidean norm is maximized. For functional (f_i), the leading functional principal component is defined analogously as the

[14] As discussed subsequently, we use the functional principal component basis obtained from the actual parents' income trajectories, which are given as functions defined continuously over $[p, q)$.

[15] This section introduces some technical non-asymptotics that allow a formal comparison of our approach with existing alternatives. Readers who are only interested in our empirical findings and their interpretations may skip this section.

[16] If $(f_i \otimes f_i)$ has a common expectation for $i = 1, 2, \ldots$, we would expect $n^{-1}Q \to_p \mathbb{E}(f_i \otimes f_i)$ under suitable regularity conditions, where the limit becomes the variance operator of (f_i) when (f_i) is of mean zero. Therefore, Q may be viewed as the (unnormalized) sample variance. If the variance operator of (f_i) is known, we may use as a basis its eigenfunctions (v_j) associated with eigenvalues (λ_j), $\lambda_1 \geqslant \lambda_2 \geqslant \cdots$, in which case our representation in (9) is commonly referred to as the Karhunen-Loève expansion used in the analysis of functional data observed from stochastic processes. Note that $\mathbb{E}\langle v_j, f_i\rangle\langle v_k, f_i\rangle = \big\langle v_j, [\mathbb{E}(f_i \otimes f_i)]v_k\big\rangle = 0$ for all $j \neq k$, and therefore, $\big(\langle v_j, f_i\rangle\big)$ is an uncorrelated sequence of random variables with variances (λ_j).

eigenfunction of Q in (15) associated with the largest eigenvalue, and it represents the linear combination of (f_i), $\sum_{i=1}^n c_i f_i$ with $\sum_{i=1}^n c_i^2 = 1$, whose Hilbert space norm is maximized. Although we use Hilbert space notions and languages to formally present the FPCA, it becomes identical to the PCA if we identify functions as long vectors as mentioned above. The PCA is used widely to identify and estimate factors in factor models. See, e.g., Bai and Ng (2002). Similarly, the FPCA may be used to identify and estimate functional factors in functional factor models, as shown in Chang et al. (2021).

Asymptotically, the choice of a basis is largely nonconsequential. It is, however, crucially important in finite samples as clearly demonstrated in Appendix A. Using the functional principal component basis (v_j^*) has some non-asymptotic optimal properties, which can be readily shown in our framework relying on the approach by Chang et al. (2021). Let Π_m^* be the projection on the subspace H_m^* spanned by the m-leading principal components $(v_j^*)_{j=1}^m$. Then, as shown in Chang et al. (2021), we have

$$\Pi_m^* = \underset{\Pi_m}{\operatorname{argmin}} \sum_{i=1}^n \left\|(1-\Pi_m)f_i\right\|^2 = \underset{\Pi_m}{\operatorname{argmin}} \operatorname{trace}(1-\Pi_m)Q(1-\Pi_m), \tag{16}$$

where Π_m signifies a projection on the m-dimensional subspace H_m of H spanned by any m-number of functions $(v_j)_{j=1}^m$. This means that the functional principal component basis best approximates (f_i) by $(\Pi_m^* f_i)$ in the squared error sense. The squared error is larger when any other m-number of functions are used. Second, when the functional principal component basis is used, we have

$$\sum_{i=1}^n \left(\Pi_m^* f_i \otimes (1-\Pi_m^*)f_i\right) = \Pi_m^* Q(1-\Pi_m^*) = 0. \tag{17}$$

Note that $\langle v_j^*, Q v_k^* \rangle = 0$ for all $j \neq k$. This implies that the two functional regressors $\left(\Pi_m^* f_i\right)$ and $\left((1-\Pi_m^*)f_i\right)$ in (10) with $\Pi_m = \Pi_m^*$ are orthogonal. Therefore, the estimation of β in regression (11) on functional covariate $\left(\Pi_m f_i\right)$ with $\Pi_m = \Pi_m^*$ in place of (f_i), ignoring the approximation error given by $\left((1-\Pi_m^*)f_i\right)$, does not create any omitted variable bias problem. No other basis provides this property, since $\sum_{i=1}^n \left(\Pi_m f_i \otimes (1-\Pi_m)f_i\right) \neq 0$ for the projection Π_m defined by any other basis.[17]

The properties we discussed above can be used to establish the optimality of the functional principal component basis (v_j^*) in estimating the functional parameter β in the functional regression (1). We assume that the standard assumptions on the classical regression hold for our functional regression with $\mathbb{E}(\varepsilon_i^2) = \sigma^2$. Here and in our subsequent discussions, we

[17]The property holds only for linear functional regression models and does not apply to more general nonlinear functional models in the paper.

denote by $\mathbb{E}$ the conditional expectation given (f_i). We let $\bar{\beta} = \Pi_m \beta$ and $\hat{\beta}$ be the estimator of $\bar{\beta}$ obtained by the approach we introduced in Section 2.2, and consider the integrated mean squared error (IMSE) defined by

$$\mathbb{E}\big\|\hat{\beta} - \bar{\beta}\big\|^2 = \mathbb{E}\int_p^q \big(\hat{\beta}(r) - \bar{\beta}(r)\big)^2 dr$$

for $\hat{\beta}$ as an estimator of $\bar{\beta}$, which relies on an arbitrary choice of m-number of functions $(v_j)_{j=1}^m$. Note that not β itself, but only $\bar{\beta}$, a finite-dimensional approximation of β, is estimable. For comparison, we denote by $\bar{\beta}^*$ and $\hat{\beta}^*$ the approximated β and its estimate based on the functional principal component basis $(v_j^*)_{j=1}^m$, which correspond to $\bar{\beta}$ and $\hat{\beta}$, respectively, and consider the IMSE $\mathbb{E}\big\|\hat{\beta}^* - \bar{\beta}^*\big\|^2$ for $\hat{\beta}^*$ as an estimator of $\bar{\beta}^*$.

For the estimator $\hat{\beta}^*$ based on the functional principal component basis (v_j^*), we may readily deduce that $\mathbb{E}\hat{\beta}^* = \bar{\beta}^*$ and

$$\mathbb{E}\big\|\hat{\beta}^* - \bar{\beta}^*\big\|^2 = \mathbb{E}\big\|\hat{\beta}^* - \mathbb{E}\hat{\beta}^*\big\|^2 = \sigma^2 \operatorname{trace}\big(\Pi_m^* Q \Pi_m^*\big)^+,$$

where $\big(\Pi_m^* Q \Pi_m^*\big)^+$ is the inverse of the operator $\Pi_m^* Q \Pi_m^*$ restricted on the subspace H_m^* of H spanned by $(v_j^*)_{j=1}^m$. For any other estimator $\hat{\beta}$, we have

$$\mathbb{E}\big\|\hat{\beta} - \bar{\beta}\big\|^2 = \mathbb{E}\big\|\hat{\beta} - \mathbb{E}\hat{\beta}\big\|^2 + \big\|\mathbb{E}\hat{\beta} - \bar{\beta}\big\|^2 = \sigma^2 \operatorname{trace}\big(\Pi_m Q \Pi_m\big)^+ + \big\|\operatorname{bias}(\hat{\beta})\big\|^2$$

with

$$\operatorname{bias}(\hat{\beta}) = (\Pi_m Q \Pi_m)^+ \Pi_m Q (1 - \Pi_m)\beta,$$

where $(\Pi_m Q \Pi_m)^+$ is the inverse of the operator $\Pi_m Q \Pi_m$ restricted on the subspace H_m of H spanned by $(v_j)_{j=1}^m$. The IMSE of $\hat{\beta}$ is therefore defined as the sum of the integrated variance (IVAR) and the integrated bias squared (IBS) of $\hat{\beta}$. The technical details of the econometric results here are provided in Appendix C.

Although it follows immediately from (16) that $\operatorname{trace}\big(\Pi_m^* Q \Pi_m^*\big) \geqslant \operatorname{trace}\big(\Pi_m Q \Pi_m\big)$, we cannot compare the IVAR $\mathbb{E}\big\|\hat{\beta}^* - \bar{\beta}^*\big\|^2$ of $\hat{\beta}^*$ with the IVAR $\mathbb{E}\big\|\hat{\beta} - \bar{\beta}\big\|^2$ of $\hat{\beta}$ unambiguously. The comparison depends roughly upon how evenly distributed their eigenvalues are. To see this, let the nonzero eigenvalues of $\Pi_m^* Q \Pi_m^*$ and $\Pi_m Q \Pi_m$ be given by $(\lambda_j^*)_{j=1}^m$ and $(\lambda_j)_{j=1}^m$, where $\sum_{j=1}^m \lambda_j^* \geqslant \sum_{j=1}^m \lambda_j$. The IVARs of $\hat{\beta}^*$ and $\hat{\beta}$ are given by $\sum_{j=1}^m (1/\lambda_j^*)$ and $\sum_{j=1}^m (1/\lambda_j)$, respectively. If both of $(\lambda_j^*)_{j=1}^m$ and $(\lambda_j)_{j=1}^m$ are evenly distributed, then the IVAR of $\hat{\beta}^*$ would obviously be smaller than that of $\hat{\beta}$. In all cases we consider in this paper and our other related work, we have $\lambda_j^* \geqslant \lambda_j$ for all $j = 1, \ldots, m$, if both of them are defined in descending order, and the IVAR of $\hat{\beta}^*$ is always smaller than that of $\hat{\beta}$.

Finally, note that we did not consider the approximation errors $\bar{\beta}^* - \beta$ and $\bar{\beta} - \beta$ in comparing the estimators $\hat{\beta}^*$ and $\hat{\beta}$. The magnitudes of these errors are not generally analyzable, since they are entirely dependent upon the unknown functional coefficient β. A specific functional basis (v_j) may be used if we know, although rarely, that it is particularly well fit to approximate β efficiently with a small number of functions. Even in this case, however, β cannot be reliably estimated unless (f_i) has enough variation over the space spanned by these functions. The use of the functional principal component basis (v_j^*) means that we approximate β by projecting it onto the space where (f_i) has the largest variation and β is most strongly identified for any given m.

Appendix B demonstrates the superior performance of functional principal components as a choice of basis over two standard alternatives, the B-spline basis and the flexible Fourier basis, when we estimate our FDA model on the Panel Study of Income Dynamics data used in this paper. These alternative bases produce positive levels of integrated bias (IBS) while the principal component basis has zero bias. The optimal choice of B-Spline components also produces a substantially higher integrated variance (IVAR). Hence the theoretical justifications for the principal component approach are verified for the data under study.

3 Data

We investigate the association between a parental income trajectory (f_i) and an approximation of child's permanent income (y_i) using the functional regression model (1). To be able to estimate the model, we need functional observations on (f_i) for each child i, which requires the parental income to be observed at each age over some period, say from birth to age 19, for each child i. Our empirical analysis, therefore, requires longitudinal data.

We use the Panel Study of Income Dynamics (PSID) for our analysis. The PSID data are available at an annual frequency from 1968 to 1997, but only biennially after 1997. We define (f_i) as the parental log income trajectory from their children's age 0 to 19.[18] The reason we choose these 20 years is mainly to cover parental income for all childhood years and when the children enter college.[19] We denote by (y_i) the offspring's permanent log income, and approximate it by the time average of a child's log income between ages 30 and 35 since most adults are economically active and stable in their early 30s.

To construct our data from PSID, we classify children by birth cohort. In our empirical

[18]Throughout, we use the Daubechies 3 tap wavelet to represent functional variables (f_i) observed discretely as functions defined continuously. As a robustness check, we also used (f_i) defined only at discrete points $r = 0, \ldots, 19$ and found that all our empirical results do not change in any significant manner, as mentioned in Footnote 4.

[19]Most common age range of college freshmen is between 17 and 19.

analysis, we consider 11 birth cohorts starting from 1967 to 1977.[20] Since PSID is available since 1968, it is possible to include more birth cohorts in the analysis. However, for the birth cohorts 1979 and later, it is not feasible to track parental incomes annually, since the PSID data is available only biennially after 1997. Therefore, to avoid potential problems from using mixed frequency data, we limit our analysis to the birth cohorts from 1967 to 1978 whose parental income is available at every age from birth to age 19. Specifically, we construct our data for each year from 1967 to 1977 as follows:

1. Construct a birth cohort: collect families who have a newborn baby in that year.
2. Define (f_i): track parental income for twenty years (from child's birth to age 19).
3. Define (y_i): track a child's adult income for six years (from child's age 30 to 35).[21] [22]

We define parental income as the sum of father and mother's income and offspring's income as the sum of the child's income and that of any partner.[23] We do not adjust family income by family size (or consider the number of siblings as a control variable). Parental income reflects factors such as parents' education, capability, quality of the neighborhood, etc. as noted by Heckman and Mosso (2014), which do not naturally scale with family size. Our final data set includes 11 birth cohorts, consisting of the birth cohorts from 1967 to 1977, and 817 families in total. All income is adjusted in 2010 dollars using the Consumer Price Index for all Urban Consumers (CPI-U).

Following the broader mobility literature, we include a range of control variables in our analysis. These are reported in Table 1.

4 Functional Regressions for Intergenerational Mobility: Linear Models

In this section, we present and analyze the functional regression characterization of intergenerational mobility. We follow the mobility literature by using a time average of child's log income (y_i) as our measure of permanent income and evaluate how the trajectory of parental

[20]PSID reports the income in the previous year, therefore, the time when the survey is taken. For this reason, it is possible to track parental income for 20 years from the birth cohort of 1967.

[21]This approach is taken by Haider and Solon (2006) for example.

[22]Note that the main PSID family and individual data sets are available annually until 1997 and biennially thereafter. Therefore, children's income in their early 30s is available only every two years. For example, for those in the 1967 birth cohort, we observe their adult incomes at ages 30, 32, and 34. We take the average of these incomes as a proxy for their permanent income, and use it as the regressand (y_i).

[23]Our definition of family income is total pre-tax income, which is defined as a sum of labor, asset, and transfer income.

Table 1: Summary Statistics of Control Variables

	Mean	Stdev
A. Age at Birth		
Average Age of Parents at Birth	27.43	6.00
B. Race of Head		
Caucasian	0.93	0.26
African American	0.05	0.23
Others	0.02	0.14
C. Region		
East	0.18	0.38
Midwest	0.35	0.48
South	0.30	0.46
West	0.17	0.37
D. Education Level		
College Educated Head	0.26	0.44
E. Family Structure		
Single-Parent	0.36	0.48
Working Mother	0.48	0.50
F. Occupation Type		
White Collar	0.45	0.50
Skilled/Semiskilled	0.49	0.50
Unskilled	0.05	0.22
Farmer	0.01	0.08

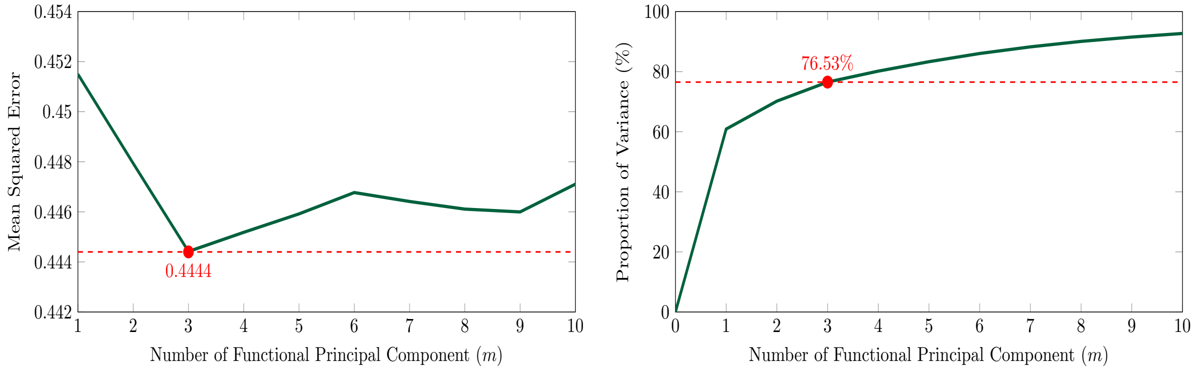
Notes: Panel A presents average age of parents at child's birth. In the case of a single-parent family, the age of single-parent is used in place of average age. Panels B, C, D, E, and F report summary statistics for the variables related to family characteristics at child's birth. All variables in these panels are dummy variables. For example, 'Caucasian' in Panel B is 1 if the race of the household head is Caucasian and 0 otherwise. Education dummy variables in Panel D are defined similarly.

log income over the first 20 years of child's life, from birth, age 0, to age 19 predict its value, i.e., we estimate the functional regression

$$y_i = \alpha + \int_0^{20} \beta(r) f_i(r) dr + \varepsilon_i,$$

which is (1) with $p = 0$ and $q = 20$, where y_i is average log income of child i and $f_i(r)$ is a trajectory of parental log income across child's age $r \in [0, 20)$, covering ages 0 to 19. The functional coefficient $\beta(r)$ is an intergenerational elasticity of income i.e mobility curve

Figure 1: Cross-Validation Errors and Cumulative Screeplot



Notes: The left panel presents the leave-one-out cross-validation (LOOCV) errors as a function of the number m of functional principal components used to approximate parental income trajectory. In the right panel is the cumulative scree plot, which presents the total percentage of variation explained by the leading m-functional principal components also as a function of m.

that measures the age-specific association between parental income trajectory and offspring's permanent income. This is the key objective of our analysis.

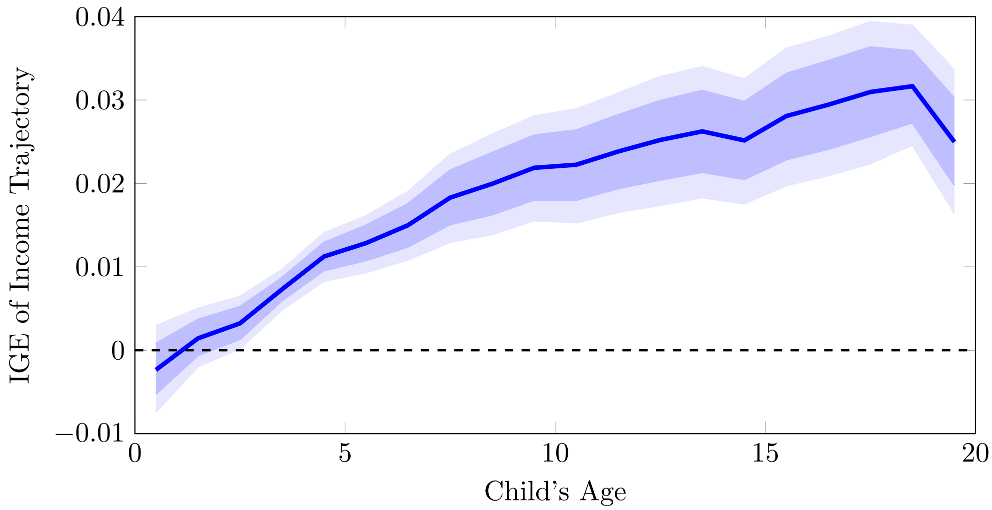
The choice of m, the number of functional principal components, is critical in estimating β. We choose $m = 3$ based on minimizing the leave-one-out cross-validation error of the approximating regression (12) and captures 76.5% of the total variation of parental income trajectories. Figure 1 illustrates the leave-one-out cross-validation error computed across m values for the data under analysis. Our estimates of the IGE curve are qualitatively robust with respect to the choice of m when set between 2 and 6. Appendix A and Appendix B discuss this in more detail and provide comparisons of our IGE curve estimates with estimates from conventional regression and functional regressions using other basis functions.

4.1 The Intergenerational Elasticity Curve

Figure 2 presents our estimated $\hat{\beta}$ along with a set of bootstrap confidence intervals. The overall message of the figure is that there is strong evidence that the intergenerational elasticity of income differs across childhood and adolescence. Interestingly, the curve is essentially monotonically increasing, with a dip after age 18. Thus, the marginal sensitivities of offspring permanent income to adolescent age parental incomes are higher than for earlier years.

The most striking feature of the shape of this estimated mobility curve is its nearly uniform monotonicity, which means that the children of parents with increasing income trajectories have higher expected permanent incomes than others. These differences are large. For example, the average sensitivity of income in ages 0 to 3 is 0.002 while the average sensitivity to income in ages 16 to 18 is an order of magnitude larger, 0.03. Age 18 is of course

Figure 2: Age-Varying IGE of Income



Notes: This figure plots the estimated functional coefficient curve in our benchmark functional regression (1). The estimated function $\hat{\beta}(r)$ measures the percentage change of a child's income in response to parental income at child's age r, covering child's ages 0 to 19. Darker (lighter) shaded areas indicate the 68% (90%) confidence intervals obtained by the wild bootstrap applied to the fitted errors in the approximating regression (12).

a benchmark makes intuitive sense for changes in parent child relationships due to such as college attendance or labor force entry. So the change in parental income sensitivity relative to other teenage years is presumably due to this. Overall, our estimated mobility curve leads to two broad conclusions. First, parental incomes in later childhood and adolescence have qualitatively much stronger marginal effects on predictive income than parental incomes in earlier years. Second, for families with equal permanent incomes, children whose parents exhibit increasing income trajectories have higher expected permanent incomes than those whose parents exhibit stable or declining trajectories.[24]

A pair of standard statistics help summarize the substantial age-specific heterogeneity in income effects. The first measure is $\hat{\beta}_{range}$, the studentized range of $\hat{\beta}(\cdot)$. This measure is defined as

$$\hat{\beta}_{range} = \frac{\max\limits_{0\leqslant r<20} \hat{\beta}(r) - \min\limits_{0\leqslant r<20} \hat{\beta}(r)}{\sqrt{\dfrac{1}{20}\displaystyle\int_0^{20} \left(\hat{\beta}(r) - \bar{\beta}\right)^2 dr}},$$

where $\bar{\beta} = (1/20)\int_0^{20} \hat{\beta}(r)dr = 0.0209$. The value of studentized range is given by 3.52, which implies that support of the age-specific heterogeneity in IGE's is more than three times larger

[24]These results are qualitatively similar to Cheng and Song (2019) who find, using parameter heterogeneity in quadratic functions of time for parental and offspring income trajectories, that the growth rate of parental income predicts offspring coefficients.

than the overall variation across ages, as measured by the standard deviation. The second measure is $\widehat{\beta}_{dev}$, the scale adjusted average area between two curves, which is defined as

$$\widehat{\beta}_{dev} = \frac{\frac{1}{20}\int_0^{20} \left|\hat{\beta}(r) - \bar{\beta}\right| dr}{\bar{\beta}}.$$

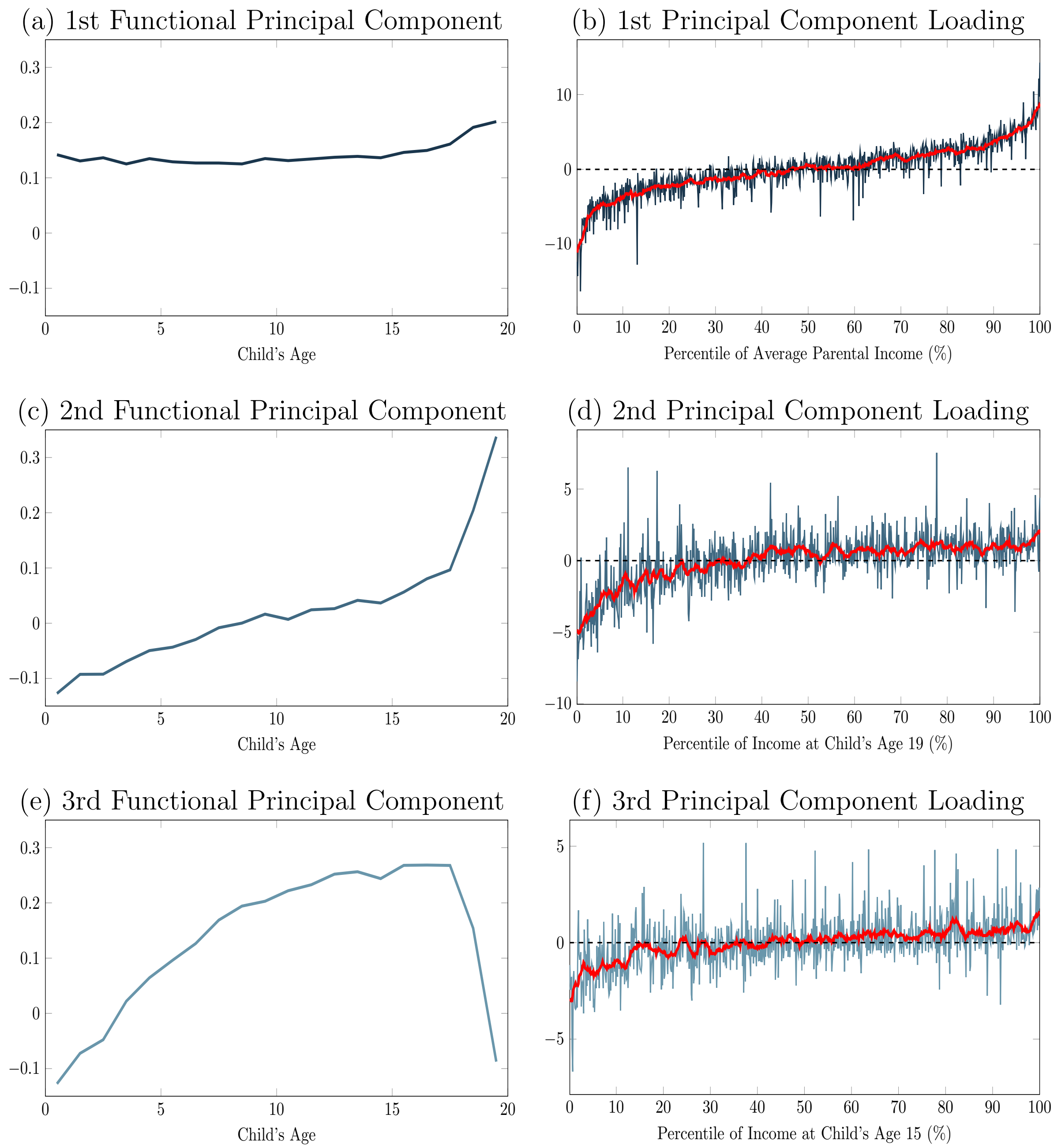
The value is given by 0.45, which means the estimated IGE at each child's age is deviated from its mean value by 45% of its mean value on average. Together, these measures reinforce the visual message of first order age heterogeneity in income effects.

How does our regression curve relate to the standard intergenerational elasticity of income coefficient produced by a linear regression? For the overall sample, if one were to estimate the conventional bivariate regression of permanent income onto permanent income, the estimated IGE is 0.57, similar to the estimate in Mazumder (2016). The natural comparison to our function model involves the calculation of the effect of a unit increase in $y(r)$ for all r. The effect of such a change on expected future income change is the integral of $\beta(\cdot)$ over the 20 years we analyze. Using our estimated $\hat{\beta}(\cdot)$ one has $\int_0^{20} \hat{\beta}(r)dr = 0.42$. This leads to the interesting result that, from the perspective of permanent income changes, the conventional linear regression overstates persistence across generations. Put differently, our results show how the conventional regression produces misleading results when the underlying mobility process relies on features of parental income trajectories beyond the mean.

The relationship between $\beta(r)$ when compared to the conventional β can be further understood by considering the functional principal components and associated loadings that are produced by the FDA analysis. Figure 3 presents the loadings for the first 3 principal components, the shapes of the components, and the effects of the components on the parental income trajectory. Following Figure 1, the first component, which exhibits little variation across childhood and adolescence, explains 60.9 percent while the second component, which increases dramatically in later adolescence, explains about 9.3 percent, and a third component that increases dramatically in the earlier years explains 6.3 percent. The first principal component, which exhibits little variation over time, is analogous to the standard object used in mobility analyses. Components 2 and 3, in contrast, demonstrate two separate types of variation, one that involves growth in earlier years, and the second in later years. These dimensions of trajectories are lost by the use of parental permanent income as a sufficient statistic to characterize how parental incomes matter.

How might one interpret these findings? Heterogeneity in the effects of parental incomes at different ages can be produced when parents face borrowing constraints of various types, so that permanent income is not a sufficient statistic for investments in children, but as

Figure 3: Functional Principal Components and Their Loadings



Notes: The top, middle, and bottom panel in the first column of the figure presents, respectively, the first, second, and third functional principal components which are extracted from cross-sectional observations on log parental income trajectories, while each row in the second column presents their loadings which vary across households. The loadings associated with the first, second, and third principal components are sorted by the average parental income, parental income at child's age 19, and parental income at child's age at 15, respectively. The red lines indicate the moving average, which is calculated by averaging loadings across 20 households.

far as we know, there is no model of borrowing constraints that produce this particular pattern; Grawe (2004) in fact indicates the difficulties in mapping credit constraints to clean predictions about income sensitivities. We conjecture that our findings speak to issues of the qualitative differences in the inputs to children versus those of adolescents; more on this below.

4.2 Robustness to Permanent/Transitory Income Distinctions

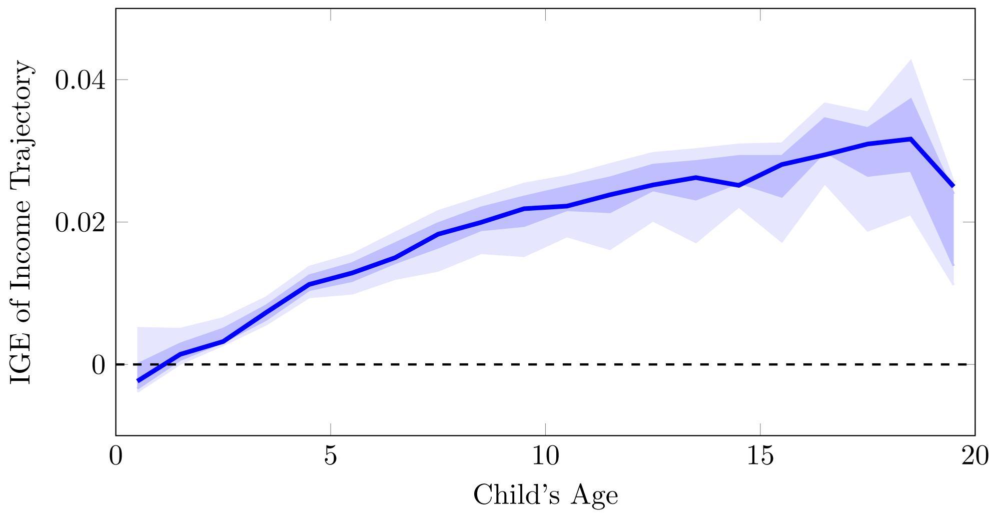
One natural concern with our findings on the IGE curve is that the curve does not distinguish between the effects of permanent versus transitory income on children, which is a primary idea in the empirical mobility literature. This distinction is a justification for using averages of parental income, as emphasized in Solon (1992). Further, when yearly family income is the sum of a time invariant permanent component and a time-varying transitory one, the more years over which averaging occurs, the smaller the effect of measurement error in creating downward bias in the IGE; the relationship between number of years used to construct parental permanent income and the magnitude of the IGE is well documented in Mazumder (2005). Hence a possible response to our findings is that the IGE curve variation reflects biases induced by the transitory components of parental incomes experienced by children and that the conventional IGE regression is in fact correct. One response to this objection is to observe that our first principal component, as discussed above, exhibits very little temporal variation across incomes at different childhood ages and so essentially captures the measure of permanent income in other studies. The second and third components, which are constructed as orthogonal to the first component, demonstrate the marginal information in the levels of incomes in predicting offspring permanent income.

To further buttress the evidence that year by year income fluctuations have informational content about offspring permanent income, we investigate the robustness of our estimate against different specifications of permanent income. First, extract permanent income estimates from family income series using time series filtering methods and use these for functional IGE estimation Second, we employ an instrumental variables strategy to extract In both cases, we find qualitatively similar mobility curves to those we estimate using the income levels series. Taken together, these two exercises lead us to conclude that the time-varying IGE curve is not an artifice of failing to distinguish permanent and transitory income.

4.2.1 Modelling Permanent Income as a Random Walk

Following a longstanding in the income dynamics literature, MaCurdy (1982), Blundell and Preston (1998), we construct estimates of family permanent income via the time series struc-

Figure 4: Re-estimated IGE Using Permanent Parental Income



Notes: The figure presents the functional coefficient β, intergenerational income elasticity curve, re-estimated with the permanent parental income trajectory. The permanent income/transitory income distinction is made by applying the Beveridge-Nelson (BN) decomposition, and the permanent parental income is obtained as the long term forecast, a trend of the series given its history. We redefine the log parental income trajectory (f_i) to be this trend for each i, and re-estimate the functional coefficient β using the redefined log parental income trajectory in our benchmark functional regression in (1). The darker (lighter) shaded areas indicate 68% (90%) confidence bands obtained by the usual wild bootstrap applied to the fitted residuals from the approximate standard regression (12).

ture of income trajectories and model permanent income as a random walk. Letting (w_{it}) be the observed parental income in log for family $i = 1, \ldots, n$ at time $t = 1, \ldots, \tau$,[25] and define it as

$$w_{it} = w_{it}^P + w_{it}^T,$$

where the superscripts P and T signify the permanent and transitory components of (w_{it}), respectively. This specification is different from Solon (1992) which assumes (w_{it}^P) is time invariant, but is the natural formulation of permanent income as a stochastic process. To operationalize this specification, we apply the Morley (2002) form of the Beveridge-Nelson (BN) decomposition and set

$$\Delta w_{it} - \mu_i = \phi_i(\Delta w_{i,t-1} - \mu_i) + \epsilon_{it},$$

[25] Here (w_{it}) denote the raw panel observations, from which we construct our functional observation $f_i(r)$ defined for each i using a Wavelet basis as a function taking values continuously in r.

from which it follows that

$$w_{it}^P = w_{it} + \frac{\phi_i}{1-\phi_i}(\Delta w_{it} - \mu_i)$$
$$w_{it}^T = w_{it} - w_{it}^P = -\frac{\phi_i}{1-\phi_i}(\Delta w_{it} - \mu_i).$$

The parameters μ_i and ϕ_i are estimated for each i using (w_{it}) for $t = 0, \ldots, 19$. To do this, we redefine the log parental income trajectory (f_i) to be the function given by (w_{it}^P) in place of the one given by (w_{it}) for each i, and re-estimate the functional coefficient β using the redefined log parental income trajectory.

Our estimates of family-specific permanent income allow us to estimate our functional regression using parental permanent income trajectories. The re-estimated IGE curve is presented in Figure 4 with the confidence band obtained, as before, by the wild bootstrap applied to the fitted errors in the approximate standard regression. The basic message of the figure is simple: accounting for the permanent/transitory income distinction using the BN filter has little effect on our results.

Observe that these time series decompositions of observed parental incomes also speak to the concern that an age-varying mobility curve reflects differences in the signal to noise ratio between permanent and transitory income, which is the key variable in the downward bias studied by Solon (1992) and others. There is an intuition that permanent income is more accurately measured after age 30 versus before. If so, then the variance of transitory income is higher than permanent income in early years. This can be evaluated by exploring how permanent and transitory incomes evolve across childhood and adolescence. Figure 17 presents the observed parental incomes, extracted permanent incomes and extracted transitory incomes across different ages for families at the 10th, 50th, and 90th percentiles.[26] These figures provide little evidence, based on these time series decompositions that the role of transitory versus permanent income in explaining income levels is higher in early childhood than later.[27]

[26]The 10th, 50th, and 90th percentiles reported for each age are obtained from the cross-sectional distributions of parental incomes at the corresponding age.

[27]Further, if the increasing IGE curve is an artifice of lower variances for transitory income in older ages than younger one, then parental income after age 18 should contain high information content on permanent income, rendering it as useful in proxying for permanent income in a similar way to incomes in the other late adolescent years; in contrast, we find a qualitative decline in the age-specific IGE after age 18.

4.2.2 IV Estimates and Interpretations of the Intergenerational Income Relationship

We next consider an instrumental variables estimation of the mobility curve where instruments are designed to capture shocks to the income process. Here we take a different strategy compared to the time series decompositions, one that extracts transitory income estimates by considering mechanisms that underlie permanent income. While a full analysis of the determinants of permanent income is beyond the scope of this paper, we employ a strategy due to Lefgren et al. (2012) that isolate income shocks. In this approach, shocks to employment status are modeled as the residuals to employment after controlling for years of schooling and educational attainment. For our model, this means constructing 20 year residual employment status trajectories and using these as a functional instrumental variable, where this vector is treated as a finite-dimensional approximation of an underlying function in the same way we treated the observed income vector.[28]

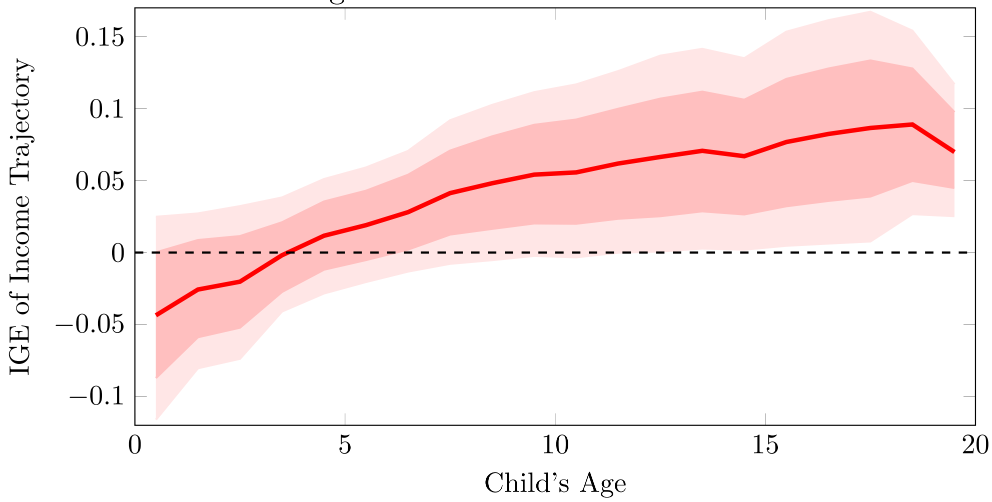
Figure 5 presents our IV estimates of the mobility curve. The IV curve has the same qualitative shape as the least squares curve: monotonic increases in the curve until age 18. Interestingly, the coefficient magnitudes are larger for the IV case than the original functional regression, which is the opposite of what is predicted by permanent income theories. Note that the IV estimation produces confidence bands that are significantly larger than the least squares estimation. This is expected since the IV estimator is generally less efficient than the least squares estimator.

4.3 Temporal Changes in Mobility Curves

We next consider changes in mobility across cohorts within our overall sample to explore changes in intergenerational mobility over time. Davis and Mazumder (2022) is the most important predecessor to our exercise. This paper argues, using the National Longitudinal Survey of Youth data that mobility was higher for individuals born in the 1950s versus the 1960s. Here we compare the β curve, in our benchmark functional regression (1) for three distinct birth cohorts in our sample; 1967-1970, 1971-1973, and 1974-1977 cohorts to uncover higher frequency differences in the mobility process. The use of 3 year cohorts is similar to Chetty et al. (2014). This complements Davis and Mazumder (2022) by comparing a late time period. The IGE curve approach also opens different perspectives on mobility changes by asking about changes in the shapes of the curves. The estimated IGE curves for these three cohorts are presented in Figure 6.

[28]For the functional IV, we represent it as a 12 dimensional vector, which explains about 90% of total variation. We use a larger value of m for the functional IV, since the ill-posed inverse problem appears to

Figure 5: IV Estimates of IGE



Notes: This figure plots the functional-IV estimate of the mobility curve. Following Lefgren et al. (2012), we consider, for each child's age, shocks to parental employment status, which are estimated as the residuals to the parental employment status trajectory after projecting out the effect of human capital such as years or schooling and educational attainment. We use the trajectory of these age-varying shocks to parental employment status as a functional instrumental variable to compute the functional-IV estimate reported in this figure. The darker (lighter) shaded areas indicate the 68% (90%) bootstrap confidence intervals.

Each of these cohort-specific curves is generally upward sloping, and so confirms the major conclusion from Section 4.1 that for the sample as a whole, incomes in later years have larger coefficients than earlier ones. At the same time, the curves reveal some interesting differences.

First, there is evidence that the effects of parental income on offspring permanent income are greater for the 1967-1970 cohort compared to later ones. The IGE curve for the 1967-1970 cohort exhibits larger values than the others. The sum of the coefficients for these three cohorts are 0.5369, 0.3426, and 0.3719, respectively.[29] In this sense, one can say the standard measure of mobility, how a permanent change in parental income starting at birth affects predictions of offspring's permanent income, has declined as have sensitivities to adolescent incomes.

What interpretations might be associated with changes across cohorts? One notable distinction between the cohorts is that those born in 1967-1970 were adolescents during the 1981-1982 recession and as such may be influenced by this experience in terms of labor market behavior, see Shigeoka (2019) for the impact of adolescence recession experiences on subsequent risk aversion.[30]

Second, there is a change in the dynamics of the IGE around age 18 across cohorts. The

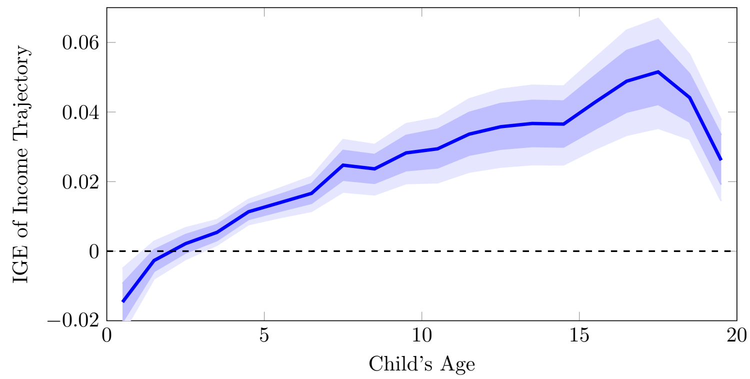
be less problematic in this context.

[29] The estimated constant term for the cohorts are 10.41, 10.22, and 12.33, respectively.

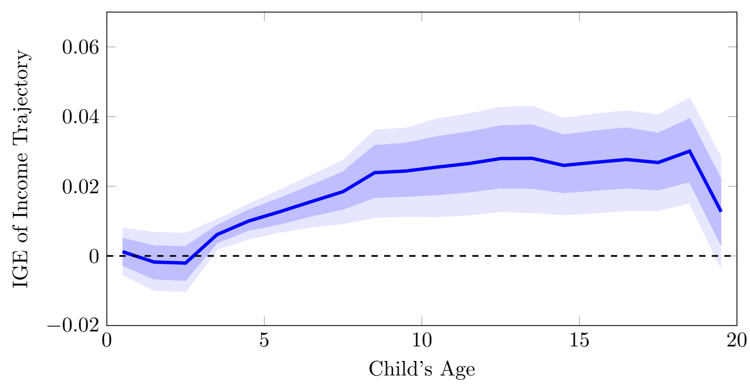
[30] We thank Neil Cholli for this observation.

Figure 6: Age-Varying IGE of Income - Cohorts Groups

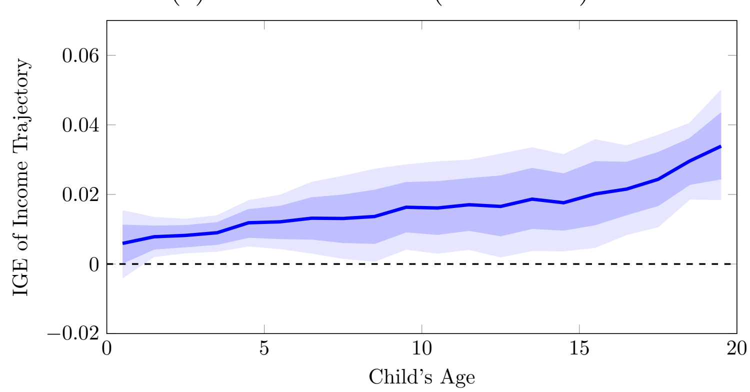
(a) Earliest Cohorts (1967-1970)



(b) Middle Cohorts (1971-1973)



(c) Latest Cohorts (1974-1977)



Notes: These figures compare the estimated functional coefficient curve in our benchmark functional regression (1) for each of the three birth cohorts groups; (a) earliest cohorts (1967-1970 cohorts), (b) middle cohorts (1971-1973 cohorts), and (c) latest cohorts (1974-1977 cohorts). Darker (lighter) shaded areas indicate the 68% (90%) confidence intervals obtained by the wild bootstrap applied to the fitted errors in the approximating regression (12).

IGE curve for the 1974-1977 cohort does not exhibit any decline in the mobility curve around age 18, as occurs for the other cohorts and for the sample as a whole. While this finding is obviously not dispositive, it hints at the possibility that norms with respect to parental support for children may have been evolving over our sample period.

We will generally work with the full sample in our subsequent analyses, but believe that these results illustrate that functional regressions methods can augment conventional evaluations of changes in mobility.

4.4 Relative Mobility

A distinct way to understand the implications of trajectories for mobility is to consider relative mobility. Figure 7 characterizes sets of parental trajectories that lead to children having expected permanent incomes in the top 10 percent and bottom 10 percent of the expected offspring income distribution. As before, we first divide our sample into three cohorts: children born between 1967 and 1970, those born between 1971 and 1973, and those born between 1974 and 1977. In this case, the cohorts are studied separately to respect the fact that relative locations of parents and children in the overall sample will be affected by trends in per capita income growth. For each cohort, the expected income level of each child was computed based on our estimated mobility curve, which revealed the sets of parental trajectories in our sample that produced children in each of the two deciles. The 68 percent and 90 percent subsets are those that capture these respective fractions of the trajectories that placed children in each decile. Properties of these trajectories are summarized in Table 2.

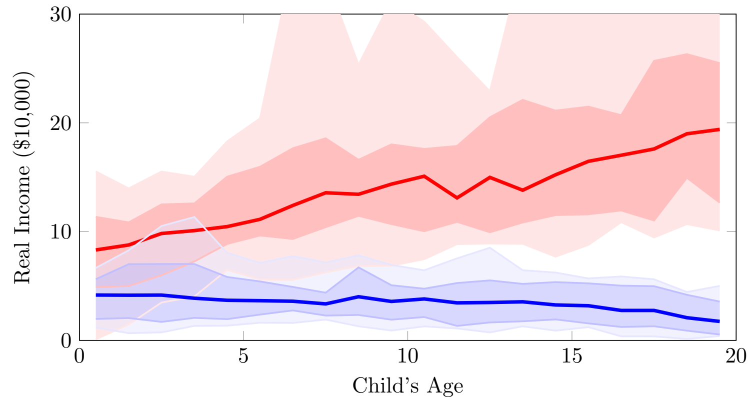
As illustrated by Figure 7, the set of trajectories that place children in the top 10 percent and the set of trajectories that placed children bottom 10 percent of expected permanent incomes exhibit quite different income levels; Table 2 indicates top 10 percent families exhibiting average incomes consistently around 4 times larger than bottom 10 percent families. This means that the average incomes of parents contain much information to distinguish the trajectory sets. However, the sets are not disjoint, especially for the 1967-1970 and 1974-1977 cohorts. The reason for this is the large difference in growth rates that occurs between the deciles after age 12, which interact with the increasing values of the mobility curves for the subsamples.

4.5 A Life Cycle Great Gatsby Curve

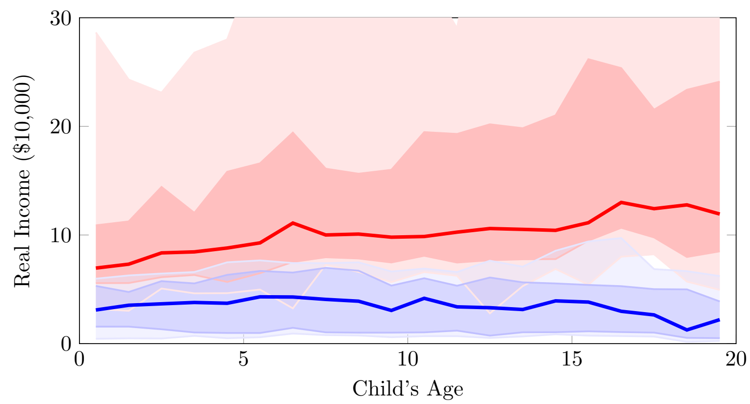
As a final exercise, we consider how our IGE curve speaks to analysis provides to the Great Gatsby Curve (Corak (2006), Corak (2013), Durlauf et al. (2022)), which refers to the findings

Figure 7: Parental Income Trajectories of Predicted Top and Bottom 10% of Expected Offspring Income - Cohorts Groups

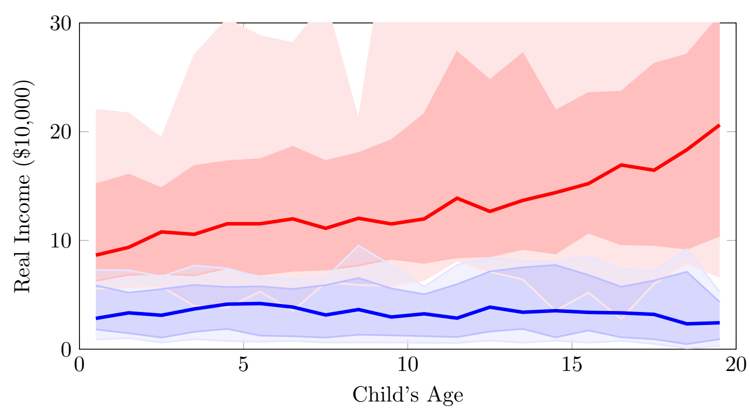
(a) Earliest Cohorts (1967-1970)



(b) Middle Cohorts (1971-1973)



(c) Latest Cohorts (1974-1977)



Notes: These figures compare the two sets of parental income trajectories of the families which are associated with the predicted top and bottom 10% of expected offspring income for each birth cohorts group; (a) earliest cohorts (1967-1970 cohorts), (b) middle cohorts (1971-1973 cohorts), and (c) latest cohorts (1974-1977 cohorts). For each family, the expected offspring income is obtained by fitting a functional regression model. The red (blue) line plots the median parental income associated with the top and bottom 10% of expected offspring income. The darker (lighter) shaded areas indicate the 68% (90%) interval.

Table 2: Statistics for Trajectories Producing Top and Bottom 10% of Expected Offspring Income - Cohorts Groups

(a) Earliest Cohorts (1967-1970)

	Parents Income Group	
	Top	Bottom
Average ($)	152,056	35,882
Variance ($\times 10^8$)	74.61	2.70
Growth Rate (%)	6.72	-3.97

(b) Middle Cohorts (1971-1973)

	Parents Income Group	
	Top	Bottom
Average ($)	147,951	35,233
Variance ($\times 10^8$)	252.53	2.20
Growth Rate (%)	3.10	-2.50

(c) Latest Cohorts (1974-1977)

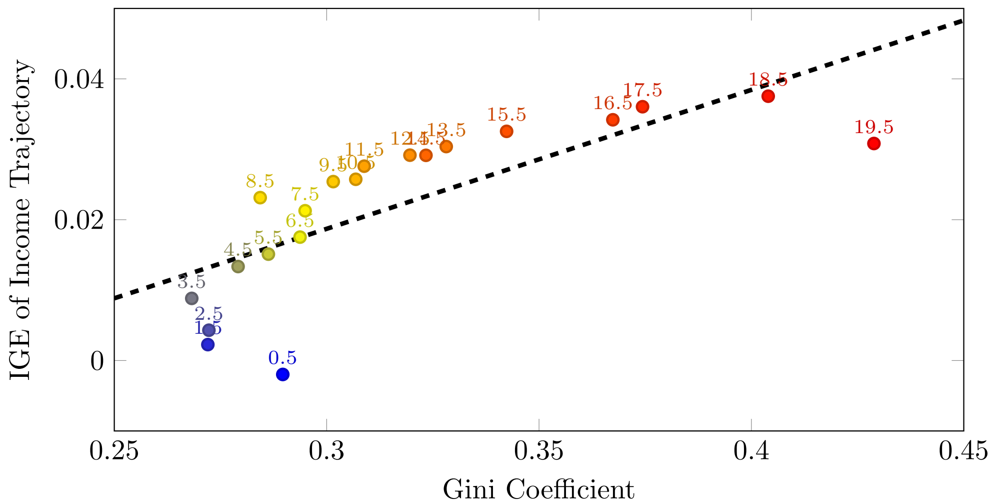
	Parents Income Group	
	Top	Bottom
Average ($)	158,503	36,696
Variance ($\times 10^8$)	89.57	2.86
Growth Rate (%)	3.51	-1.11

Notes: These tables compare key statistics – average, variance, and average growth rate – of the real incomes associated with each of the sets of parental income trajectories producing top and bottom 10% of expected offspring income for each birth cohorts group; (a) earliest cohorts (1967-1970 cohorts), (b) middle cohorts (1971-1973 cohorts), and (c) latest cohorts (1974-1977 cohorts) in Figure 7. To obtain these statistics, we first calculate them for each parent, and take cross-sectional average of these statistics. Average is computed as $\sum_{i=1}^{N}\left(\sum_{t=0}^{19} f_{i,t}/20\right)/N$, where $f_{i,t}$ is parents i's income at their child's age t, and N indicates the number of families in the set of interest. Similarly, we calculate variance as $\sum_{i=1}^{N}\left(\sum_{t=0}^{19}(f_{i,t}-\overline{f}_i)^2/19\right)/N$ where $\overline{f}_i$ is time-average parental income of parents i. Average growth rate is computed as $\sum_{i=1}^{N}\left(((f_{i,19}/f_{i,0})^{1/19}-1)\times 100\right)/N$.

in a range of contexts, that there is a positive association between measures of inequality and measures of intergenerational mobility. Our analysis permits the calculation of a variant of the Curve within the intergenerational transmission process by comparing age-specific values of the IGE against the cross-sectional differences in the incomes of parents at those ages, measured as age-specific IGE's with the Gini coefficients of parents in our sample, where incomes are measured at the associated ages. Figure 8 plots this relationship.

As revealed by the scatter plot and associated trend line, there is a clear Gatsby curve

Figure 8: The Great Gatsby Curve across Child's Ages



Notes: This scatter plot illustrates the relationship between inequality and intergenerational income mobility from child's birth to age 20. Each number on the point presents the child's age. A horizontal axis shows the Gini coefficient, which is calculated using parental income at each child's age. A vertical axis displays the IGE of income trajectory, which is the value of $\widehat{\beta}(r)$ from model (1), at each child's age, $r = 0.5, \cdots, 19.5$. The black dashed line presents the trend line of the scatter plot.

present. Periods of greater parental income volatility are associated with greater sensitivity of future offspring income to parental levels. We leave exploration of the interpretation of this relationship to future work. Here we conclude by noting that, as discussed in Durlauf et al. (2022), nonlinearities in the parent/offspring transmission process are needed to provide a pathway by which increased inequality produces lower mobility.

5 Quadratic Mobility Models

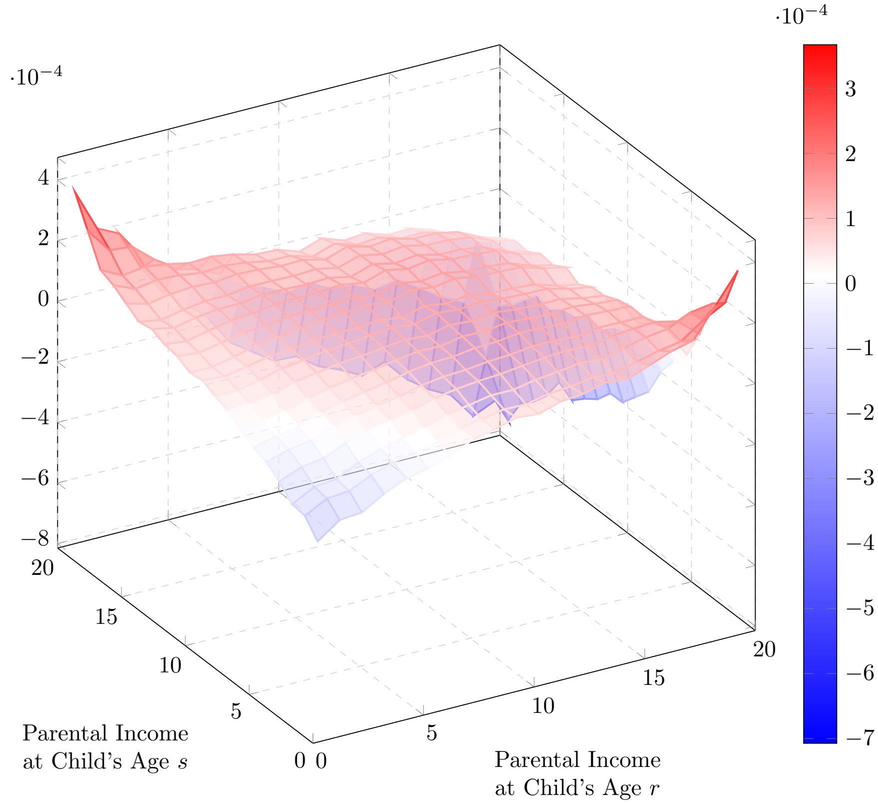
In this section, we extend our functional regression framework to allow for interactions between parental incomes at different ages based on the specification

$$y_i = \alpha + \int_0^{20} \beta(r) f_i(r) dr + \int_0^{20} \int_0^{20} \Gamma(r,s) f_i(r) f_i(s) ds dr + \varepsilon_i. \tag{18}$$

The quadratic coefficient function, $\Gamma(r,s)$ in (18), is a symmetric bivariate function. This model may be thought of as a second order Taylor series expansion of a general nonlinear functional regression model. A useful feature of the quadratic model is that the parameters of the interaction terms are interpretable as exhibiting complementarity or substitutability effects between incomes at different ages. That said, it should be observed that quadratic models do not well equipped to uncover phenomena such as affluence or poverty traps.

The introduction of the interactions between incomes at different ages means that the

Figure 9: Quadratic Effect on Offspring's Income - Surface



Notes: This figure depicts the estimated bivariate coefficient function $\hat{\Gamma}(r, s)$ for $r, s \in [0, 20)$. The red-colored area indicates the positive values of the estimated coefficient function (complementarity), while the blue-colored area denotes negative value (substitutability).

total effect of a change in income at one age on expected permanent income of a child is much more complicated than the linear functional case. The total effect of a change in income at a given age r may be decomposed as

$$\frac{\Delta y_i}{\Delta f_i(r)} \approx \beta(r) + 2\int_0^{20} \Gamma(r, s) f_i(s) ds \tag{19}$$

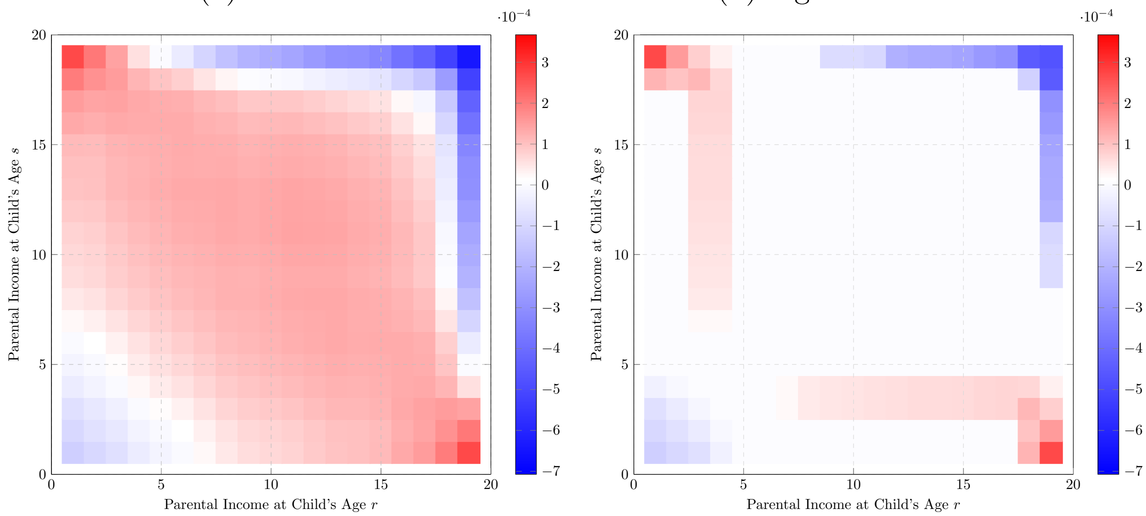
One can rewrite this expression as

$$\frac{\Delta y_i}{\Delta f_i(r)} \approx \beta(r) + 2\Gamma(r, r)\Delta f_i(r) + 2\int_{[0,20)\backslash[r-1/2,r+1/2)} \Gamma(r, s)\Delta f_i(s) ds, \tag{20}$$

which reveals that the total effect of a change in parental income at a given time involves three distinct channels: a linear component ($\beta(r)$) that corresponds to the age-specific IGE in the linear functional regression, a nonlinear own income component $(2\Gamma(r, r)\Delta f_i(r))]$ and an interaction ($2\int_{[0,20)\backslash[r-1/2,r+1/2)} \Gamma(r, s)\Delta f_i(s) ds$) which captures how a change at r affects

Figure 10: Quadratic Effect on Offspring's Income - Heatmap

(a) Estimate (b) Significant at 68%



Notes: The left panel of this figure illustrates the estimated bivariate coefficient function $\hat{\Gamma}(r,s)$ for $r,s \in [0,20)$. This heatmap presents estimated coefficients at each child's age r and s. The red-colored cells indicate the estimated coefficient is positive for specific r and s (complementarity), while the blue-colored ones are negative estimates (substitutability). The right panel of this figure presents the estimated effects that are statistically significant at the 68% confidence level.

the marginal product of incomes at other ages as well as the effect of incomes at other ages on the effects of income at r.

We again use functional principal component analysis for the model estimation. Using the same cross-validation procedure as before, we set the number of functional principal components m to 3.

Figure 9 presents the 3D surface for the bivariate $\hat{\Gamma}(r,s)$ for $r,s \in [0,20)$ function while Figure 10 presents heat maps of the function with and without a 68 percent significance level. Red denotes a positive value for $\hat{\Gamma}(r,s)$, i.e. complementarity, while blue denotes a negative value, i.e. substitutability.

Together, Figure 9 and Figure 10 illustrate two important qualitative regularities. First, there appears to be local substitutability between incomes within early childhood and within the adolescent years. Second, there is evidence of complementarities between parental income in early childhood and later adolescence.

These results are interesting from the perspective of the early childhood investment literature, cf. Cunha and Heckman (2008) and Cunha et al. (2010). This literature emphasizes the importance of complementarities in skills-producing investments in offspring across childhood and adolescence. At first glance, our results suggest a more subtle pattern of local substitutability and global complementarity. There is no logical contradiction, since the uniform complementarities assumed in the early childhood literature concern a property of the skills production function while our results involve complementarity and substitutability in the determinants of the resource constraints for those investments. Further, estimates of uniform complementarities, e.g. Cunha et al. (2010) only evaluate complementarities between t and $t+1$, which are extrapolated to make claims about t and $t+k$, while our work explicitly allows for much richer patterns.

That said, if one wishes to combine uniform complementarity in investment productivity with our local substitutes/global complementarity distinction for incomes, it is necessary to develop a more complex modeling of parental investments and their consequences than is conventional. One possibility is that, within early childhood and later adolescence, parents are able to smooth investments across adjacent years, producing local substitutability. Further, suppose that the primary investments in children during later childhood and adolescence are indirect in the sense that the investments are in memberships in a neighborhood and school district and that these types of investments are sensitive to income in those years in ways that differ from the direct investments made at earlier ages. Under these circumstances, the income and investment results are consistent. Wodtke et al. (2011) find that the exposure to disadvantaged neighborhoods in adolescence has particularly deleterious effects on the likelihood of high school graduation, which combined with our conjectures links our explanation with high marginal effects of parental incomes in later years.

The decomposition of the income effect given in (20) means that the marginal effect of a change in parental income at a given time involves both the effect from the linear term as well as the nonlinear effects at a single time-point ($r = s$) and the interaction effects at two-time points ($r \neq s$). Figure 12 illustrates how the overall age-varying IGE is changed by parental income profile by these additional effects. We divide the sample into 1967-1970, 1971-1973, and 1974-1977 cohorts in order to avoid mixing age 0 percentiles from families starting too far apart in time; secular economic growth, for example, would render such a comparison problematic. Since the total effect of a change in income at one age depends on the income trajectory of the family, we consider three examples of family trajectories. We consider sets of families distinguished by whether they are located when their children are age 0, in sets centered at the (a) 10th percentile, (b) 50th percentile, and (c) 90th percentile respectively among parents for all age 0 children in their cohort. We then average the trajectories in each of these sets to produce representative trajectories for each set. Figure 11 describes these three counterfactual family income trajectories and Figure 12 provides a visualization of levels of the IGE as determined by linear, nonlinear, and interaction factors.

Figure 12 reveals a similar story for each of the sets of cohorts. For families in the vicinity of the 10th percentile, accounting for nonlinearities and cross-age interactions leads to lower age-specific income sensitivities, relative to the direct income effects for all years up to age 17. In contrast, nonlinearities and interactions produce, relative to the direct effect a slightly larger age-specific income sensitivity at age 18 and a substantially larger income sensitivity for age 19. Families with incomes around the median exhibit negligible differences between the overall age-specific income effects and the direct income effects. For families around the 90th percentile in age 0 incomes, each cohort group exhibits the opposite effects of nonlinearities and interactions than occurred for 10th percentile families: age-specific income sensitivities of children are higher through age 17, slightly lower at 18 and much lower at 19. The summary of these patterns is simple: the quadratic generalization of functional mobility regressions reduces the sensitivity of the children of poorer families and increases the income sensitivities of children in high income families before age 18 and has the opposite effect for ages 18 and especially 19.

One can additionally calculate an approximate overall IGE measure as $\int_0^{20}(\Delta y_i/\Delta f_i(r))dr$ for families at different percentiles. The overall IGE measure for poorer families in the earliest, middle, and latest cohorts group equal 0.3776, 0.3792, and 0.3783, respectively. The same measure for families in the middle class of each cohort group is 0.3917, 0.3971, and 0.3984 while the one for richer families is 0.4104, 0.4092, and 0.4118, respectively. Families at different percentiles thus exhibit similar IGEs across cohorts, with slightly larger values for 10th percentile families than 90th percentile ones. This means that the augmentation of income effects for more affluent families in earlier years more than offsets the lower income effects at ages 17 and 18. Together, our results indicate how new information emerges from explicitly looking at trajectories.

6 Proximate Mechanisms

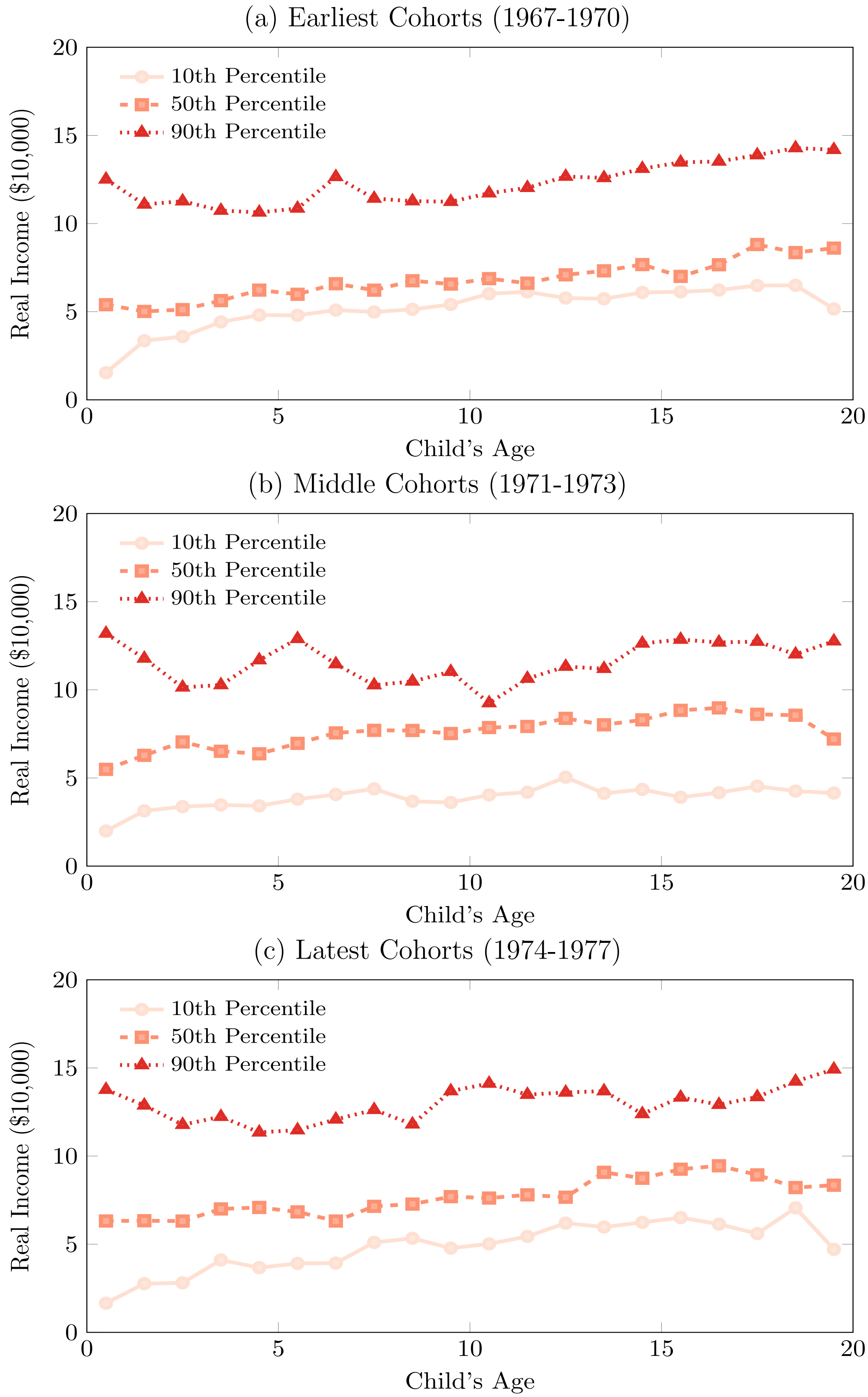
In this section, we explore how education and occupation are influenced by family income trajectories. Education and occupation are of course primary proximate determinants of income.

6.1 Education

We first evaluate the relationship between parental income trajectories and the probability of college attendance by the beginning of age 20.[31] To do this, we estimate a functional logit

[31] We leave to future work more systematic evaluation of the relationship between parental income and offspring higher education dynamics on future income, as doing so would require an assessment of incomes after reaching age 20, as well as a dynamic discrete choice of the type pioneered by Cameron and Heckman (2001) for education levels. Such an analysis would also need to address years of schooling versus graduation effects; Ashworth et al. (2021) is a recent sophisticated evaluation of their distinct roles.

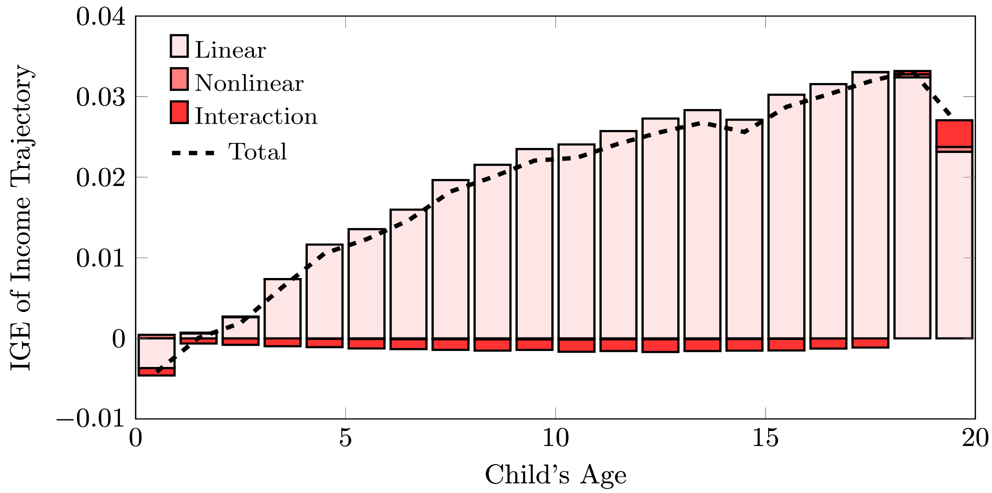
Figure 11: Counterfactural Family Income Trajectories



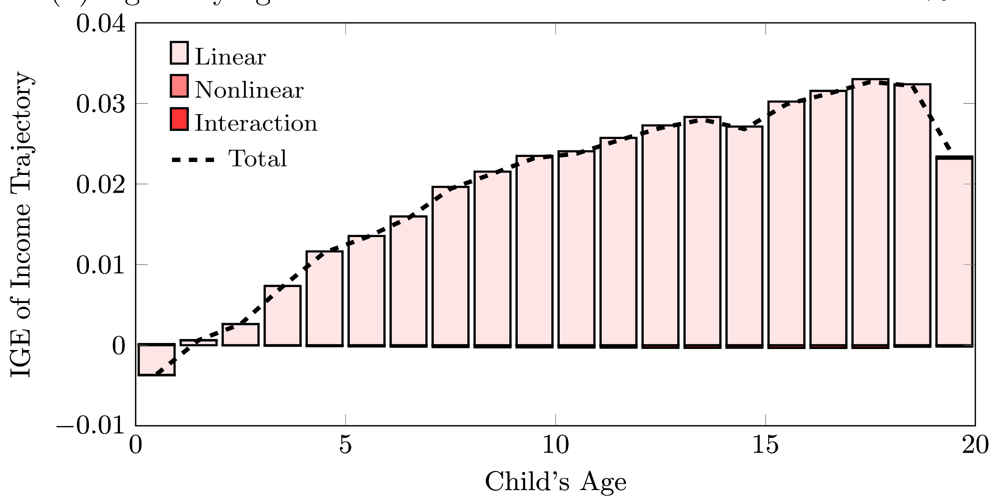
Notes: These figures present three counterfactual parental income trajectories for each birth cohorts group; (a) earliest cohorts (1967-1970 cohorts), (b) middle cohorts (1971-1973 cohorts), and (c) latest cohorts (1974-1977 cohorts). The line with pink-circles/orange-squares/red-triangles signifies average parental income trajectory of (a) bottom 10%, (b) middle 10%, and (c) top 10% at child's birth.

Figure 12: Age-Varying IGE Decomposition with Counterfactual Family Income Trajectories - Earliest Cohorts (1967-1970)

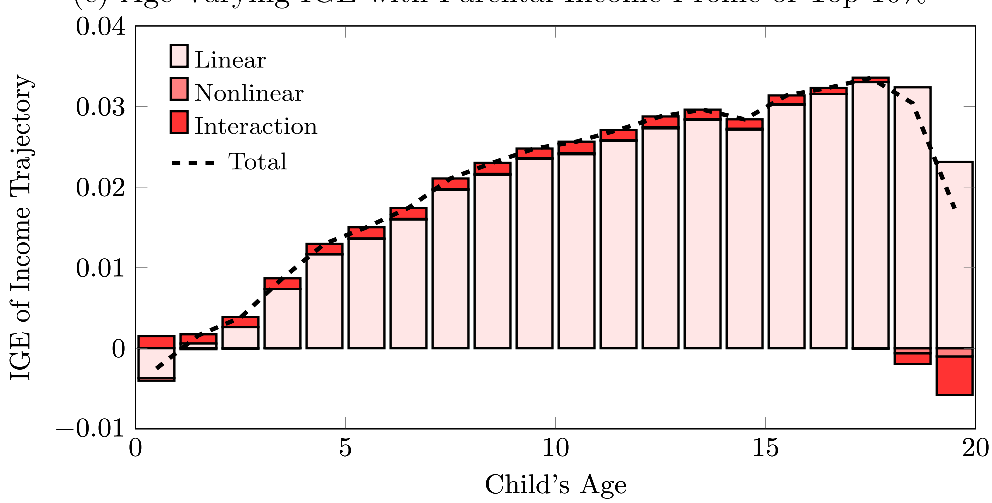
(a) Age-Varying IGE with Parental Income Profile of Bottom 10%



(b) Age-Varying IGE with Parental Income Profile of Middle 10%



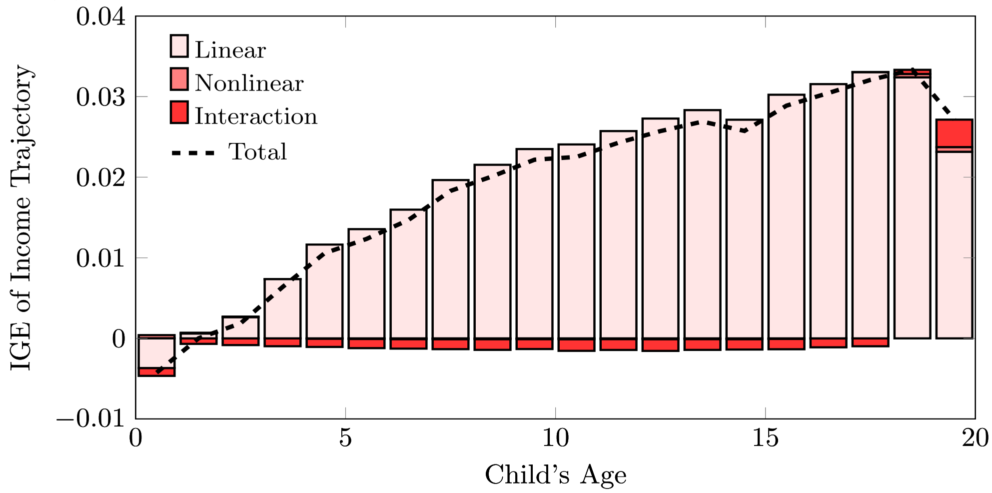
(c) Age-Varying IGE with Parental Income Profile of Top 10%



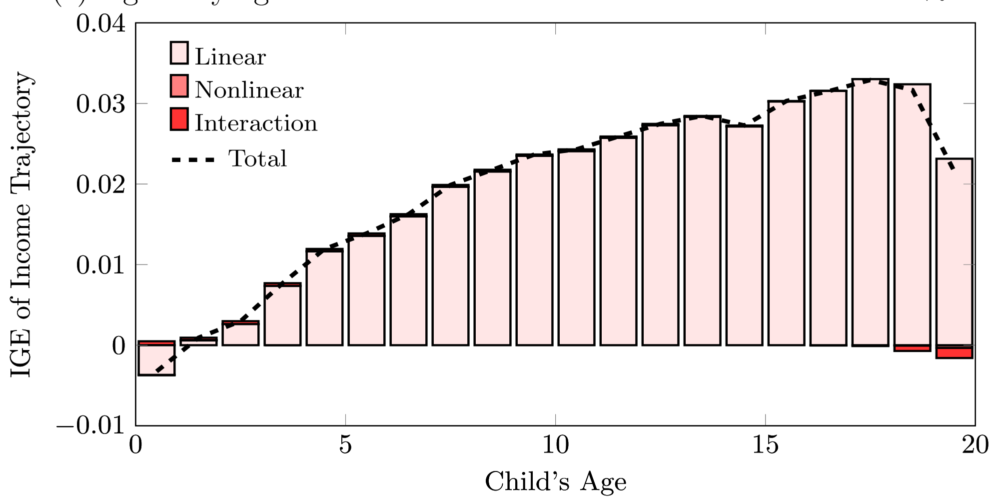
Notes: This figure presents age-varying IGE decomposition for the average parental income trajectory of the earliest cohorts (1967-1970) at (a) bottom 10%, (b) middle 10%, and (c) top 10% at child's birth. The line labeled 'total' indicates the time-varying IGE of income from the model (18). Specifically, the time-varying IGE at a child's age r is the sum of linear ($\beta(r)$), nonlinear ($2\Gamma(r,r)\Delta f_i(r)$), and interaction ($2\int_{[0,20)\setminus[r-1/2,r+1/2)}\Gamma(r,s)\Delta f_i(s)ds$) effects.

Figure 12: Age-Varying IGE Decomposition with Counterfactual Family Income Trajectories - Middle Cohorts (1971-1973)

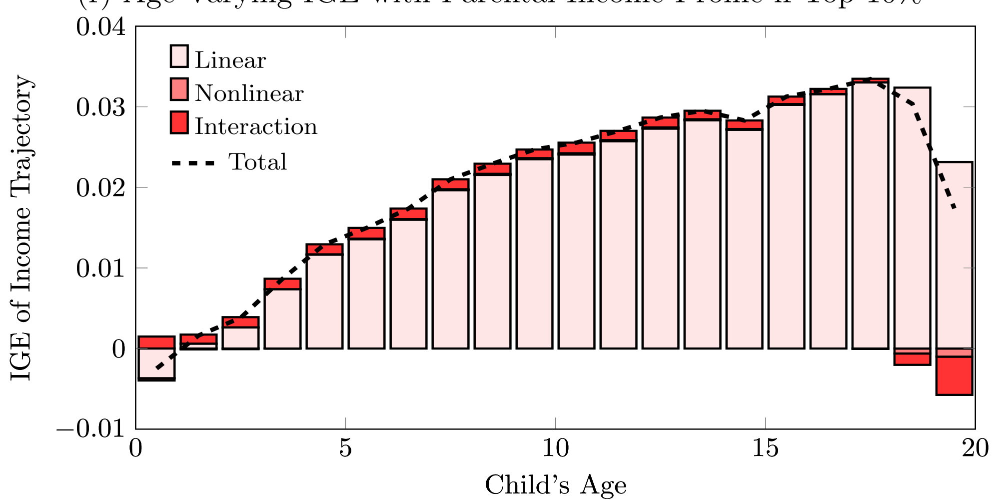
(d) Age-Varying IGE with Parental Income Profile of Bottom 10%



(e) Age-Varying IGE with Parental Income Profile of Middle 10%



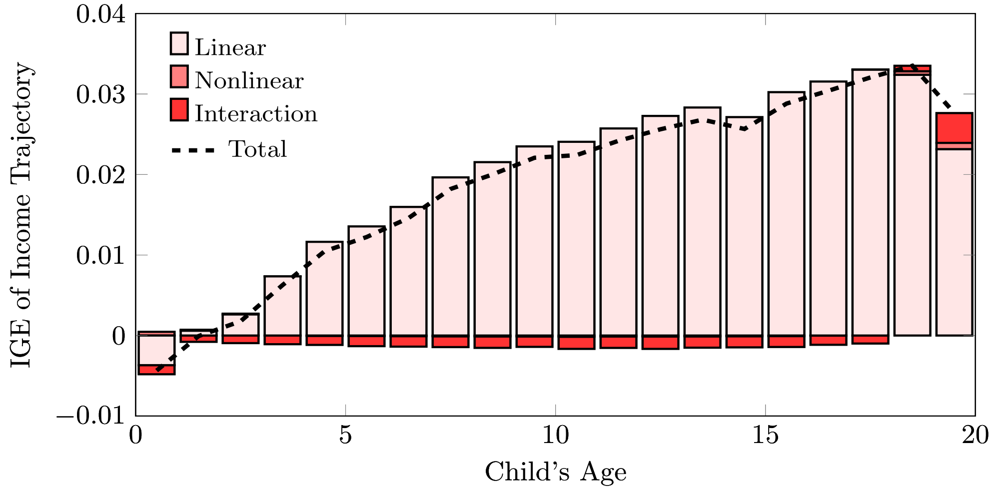
(f) Age-Varying IGE with Parental Income Profile if Top 10%



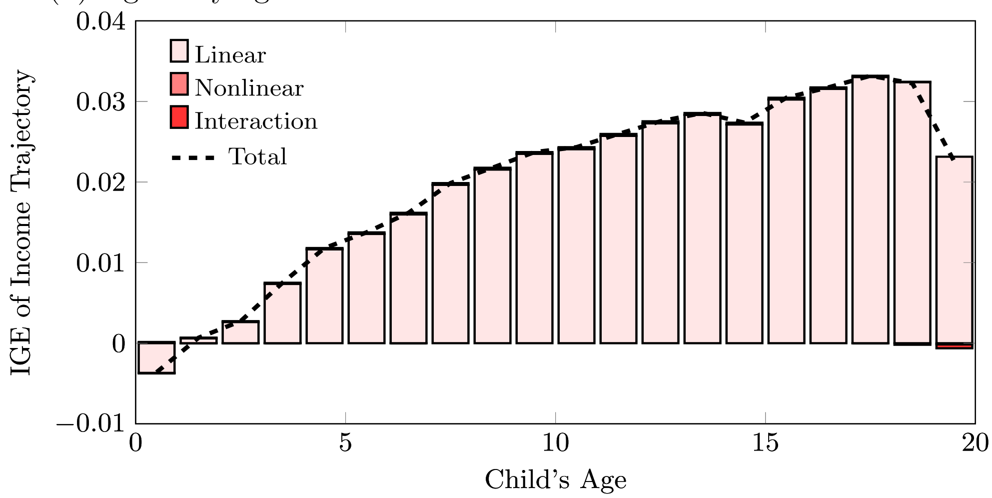
Notes: This figure presents age-varying IGE decomposition for the average parental income trajectory of the middle cohorts (1971-1973) at (d) bottom 10%, (e) middle 10%, and (f) top 10% at child's birth. The line labeled 'total' indicates the time-varying IGE of income from the model (18). Specifically, the time-varying IGE at a child's age r is the sum of linear ($\beta(r)$), nonlinear ($2\Gamma(r,r)\Delta f_i(r)$), and interaction ($2\int_{[0,20)\setminus[r-1/2,r+1/2)} \Gamma(r,s)\Delta f_i(s)ds$) effects.

Figure 12: Age-Varying IGE Decomposition with Counterfactual Family Income Trajectories - Latest Cohorts (1974-1977)

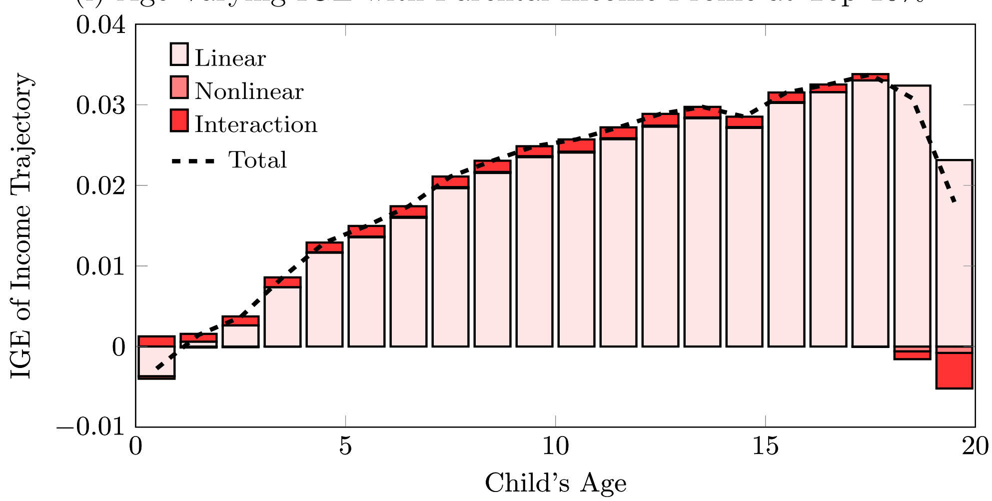
(g) Age-Varying IGE with Parental Income Profile at Bottom 10%



(h) Age-Varying IGE with Parental Income Profile at Middle 10%

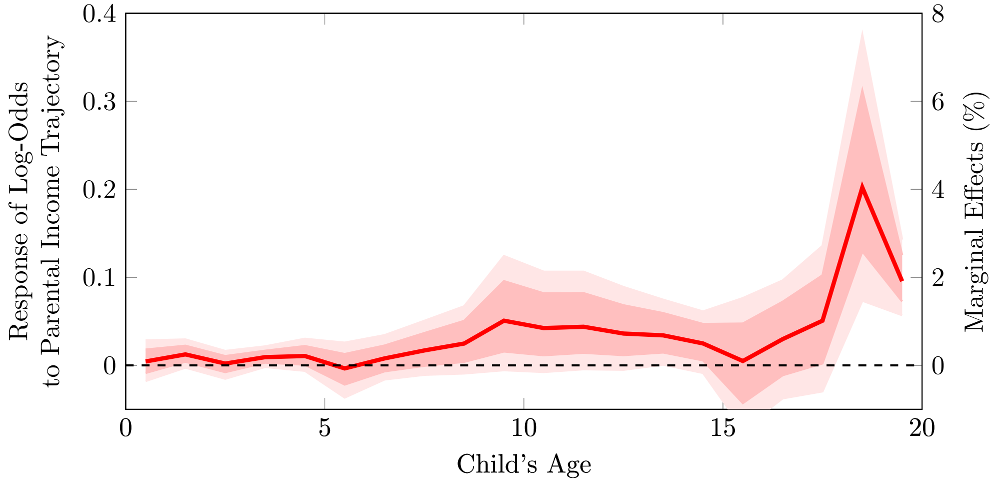


(i) Age-Varying IGE with Parental Income Profile at Top 10%



Notes: This figure presents age-varying IGE decomposition for the average parental income trajectory of the latest cohorts (1974-1977) at (g) bottom 10%, (h) middle 10%, and (i) top 10% at child's birth. The line labeled 'total' indicates the time-varying IGE of income from the model (18). Specifically, the time-varying IGE at a child's age r is the sum of linear ($\beta(r)$), nonlinear ($2\Gamma(r,r)\Delta f_i(r)$), and interaction ($2\int_{[0,20)\setminus[r-1/2,r+1/2)} \Gamma(r,s)\Delta f_i(s)ds$) effects.

Figure 13: Effect of Parental Income Trajectory on Offspring's College Attendance



Notes: This figure plots the estimate of functional logit model, $\hat{\mu}(\cdot)$ in (21) and the marginal effect of parental income at each child's age on the probability of offspring's college attendance in (23). This estimate measures the change of log-odds ratio in response to the change of parental income trajectory, and the marginal effect is a constant multiple of the estimate. The darker (lighter) shaded areas indicate 68% (90%) bootstrap confidence intervals.

generalization of our income model:

$$\log \frac{p_i}{1-p_i} = \omega + \int_0^{20} \mu(r) f_i(r) dr, \tag{21}$$

where $p_i = \mathbb{P}\{z_i = 1\}$, (f_i) is a trajectory of parental log income from birth to age 20 and

$$z_i = \begin{cases} 1 & \text{if the child } i \text{ attends college} \\ 0 & \text{otherwise} \end{cases}.$$

We estimate the response function $\mu(\cdot)$ in (21) using the same functional principal component approach developed in Section 2.2. For this model, we set the number of functional principal components m chosen at 5 as this number of components minimizes the leave-one-out cross-validation error.

We also compute the marginal effect of parental income at each child's age on the probability of offspring's college attendance. To do this, we first calculate p_i from (21),

$$p_i = \frac{\exp\left(\omega + \int_0^{20} \mu(r) f_i(r) dr\right)}{1 + \exp\left(\omega + \int_0^{20} \mu(r) f_i(r) dr\right)}. \tag{22}$$

Table 3: Occupation Types

White Collar (Type 1)	Professional, technical and kindred workers; Managers, officials and proprietors; Clerical and sales workers
Skilled/Semiskilled (Type 2)	Craftsmen, foremen, and kindred workers; Operatives and kindred workers
Unskilled (Type 3)	Laborers and service workers; Farm laborers

Notes: The definition of occupation types are based on Long and Ferrie (2013). Long and Ferrie (2013) considers four occupational types: White Collar, Skilled/Semiskilled, Unskilled, and Farmer. However, due to the lack of observations, the 'Farmer' category is excluded, and the other three occupation categories are considered in this paper.

Given (21), we first evaluate p_i at the mean level of parental income trajectory, and use this to obtain the marginal effect on p_i of the parental income at child's age r:

$$\frac{\Delta p_i}{\Delta f_i(r)} \approx p_i(1-p_i)\mu(r). \tag{23}$$

Note that the marginal effects are given as a function of r, more specifically by a constant multiple $p_i(1-p_i)$ of curve $\hat{\mu}(r)$.

Figure 13 plots the estimated response function $\hat{\mu}(\cdot)$ and the marginal effect of parental income on the probability of offspring's college attendance, with the left and right axes scale differentiating the two. The estimated coefficient $\hat{\mu}$ is relatively flat until age 15, which exhibits a sharp increase for the ages 15-18, with a decline at after 18. While overall qualitatively consistent with our income IGE curve, the differences between earlier and later ages are much stronger. The average value of the curve for ages 0-15 is 0.02 while the average for ages 16-18 is 0.09. The impact of these differences on college enrollment is indicated by the very high value of a 4 percent change in the probability of college attendance generated by a 1 percent change in age 18 income.

6.2 Occupation

We next consider how family trajectories predict offspring occupation status. Following a long tradition of occupational mobility studies in sociology, see Song et al. (2020) for an important recent example, we partition occupations into white collar, skilled/semiskilled, and unskilled.[32]

Table 3 lists the three occupation categories defined as in Long and Ferrie (2013). We

[32] Due to a lack of observations, we do not consider agricultural workers.

denote the occupational categories white collar, skilled/semiskilled, and unskilled as Type 1, 2, and 3 respectively. We limit our analysis to transitions between father's occupation to son's occupation. We do this as, in the sample, a majority of mothers were housewives (52%), but acknowledge that this is a significant limitation even though it is a common procedure for studying occupational mobility, cf, Guest et al. (1989), Long and Ferrie (2013), and Mazumder and Acosta (2015). For these transitions, we employ a functional multinomial logit model that generalizes the functional logit model used for college attendance. The number of functional principal components m is chosen at 3 by the leave-one-out cross-validation. Denoting 'k'for and son occupation type, we estimate

$$\log\left(\frac{\mathbb{P}\{z_i = k\}}{\mathbb{P}\{z_i = 3\}}\right) = \omega_k + \int_0^{20} \mu_k(r) f_i(r) dr \quad \text{for} \quad k = 1, 2, \tag{24}$$

where k is the occupation type for son and (z_i) is the indicator variable

$$z_i = \begin{cases} 1 & \text{if child } i\text{'s job belongs to Type 1 (White Collar)} \\ 2 & \text{if child } i\text{'s job belongs to Type 2 (Skilled/Semiskilled)} \\ 3 & \text{if child } i\text{'s job belongs to Type 3 (Unskilled)} \end{cases} .$$

The probabilities of transition between categories 1 and 2 are implicitly determined by (24) and the requirement that probabilities of transitions sum to 1, so (24) fully characterizes the occupational income trajectories between parents and offspring.

Eq. (24) may be used to compute marginal effects we compute of parental income at different ages on offspring job choice probabilities. For each offspring job type k,

$$\mathbb{P}\{z_i = k\} = \frac{\phi_k}{\phi_1 + \phi_2 + \phi_3} \quad \text{for} \quad k = 1, 2, 3,$$

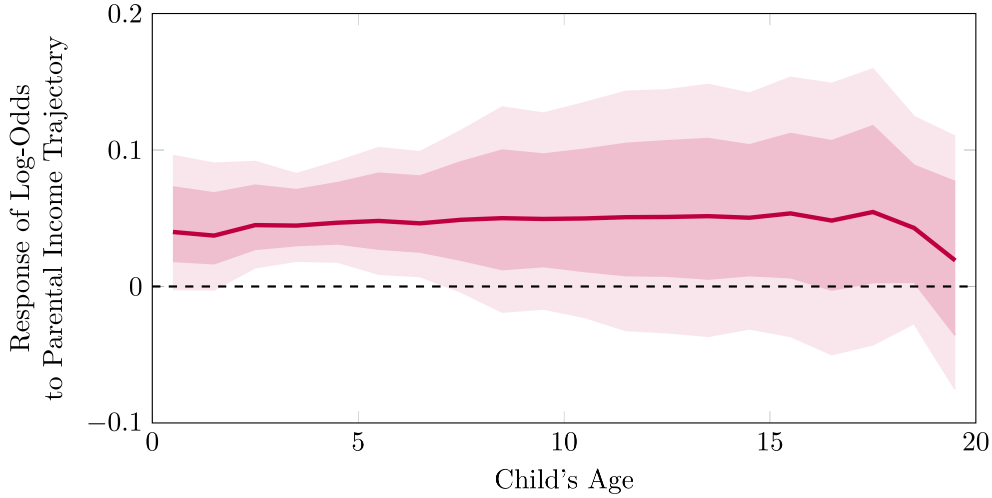
where

$$\phi_k = \begin{cases} \exp\left(\omega_k + \int_0^{20} \mu_k(r) f_i(r) dr\right) & \text{for } k = 1, 2 \\ 1 & \text{for } k = 3 \end{cases} .$$

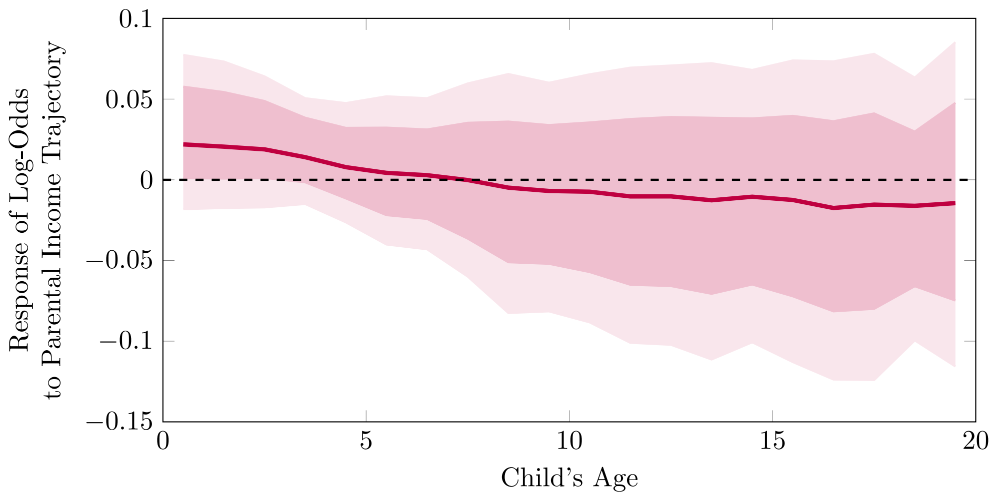
Therefore the marginal effect of the parental income at child's age s on the probability of their son having a Type k job is obtained by taking derivatives of $\mathbb{P}\{z_i = k\}$ with respect to

Figure 14: Effect of Parental Income Trajectory on Log-Odds of Occupational Choices

(a) Type 1 vs. Type 3



(b) Type 2 vs. Type 3



Notes: These figures plot the estimate of the multinomial functional logit model, $\hat{\mu}_1(\cdot)$, $\hat{\mu}_2(\cdot)$ in (24). These estimates measure the change of log-odds of the probability of offspring having (a) Type 1 job, (b) Type 2 job relative to having Type 3 job in response to the change in parental income trajectory. The darker (lighter) shaded areas indicate 68% (90%) bootstrap confidence intervals.

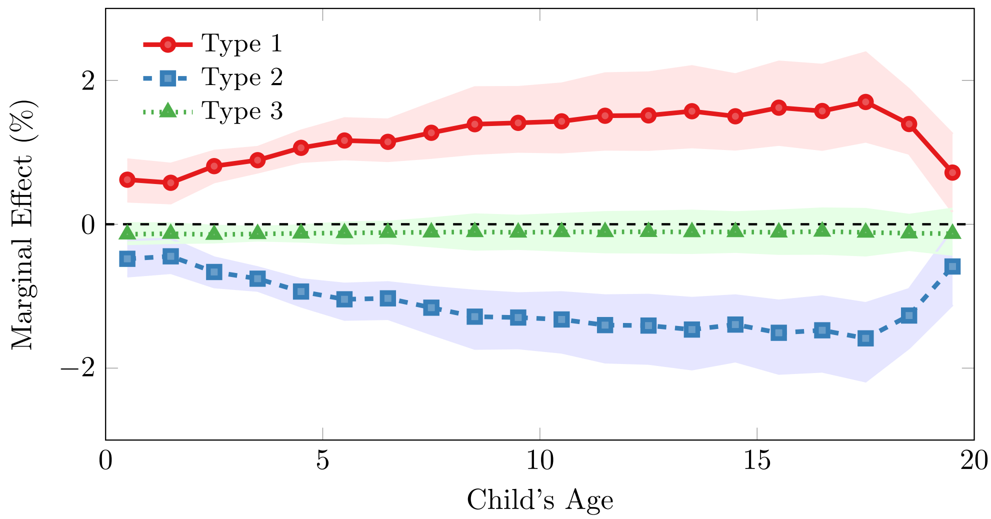
$f_i(r)$ as follows

$$\frac{\Delta \mathbb{P}\{z_i = k\}}{\Delta f_r(r)} \approx \mathbb{P}\{z_i = k\}\left(\mu_k(r) - \sum_{j=1}^{3} \mathbb{P}\{z_i = j\}\mu_j(r)\right) \quad \text{for} \quad k = 1, 2, 3, \tag{25}$$

where $\mu_3 = 0$.

Figure 14 presents the estimated response function $\hat{\mu}_1(\cdot)$ and $\hat{\mu}_2(\cdot)$, respectively. Fig-

Figure 15: Age-Varying Marginal Effect of Parental Income on Occupational Choices



Notes: This figure presents the marginal effect of parental income at each child's age on the probability of offspring's occupational choices (25). The red-circle line presents the marginal effect of parental income on the probability of offspring having a job in Type 1. The blue-square line and the green-triangle line indicate the marginal effect on the probability of offspring having a job in Type 2 and 3, respectively. The shades areas indicate 68% bootstrap confidence intervals.

ure 15 plots the marginal effects of parental income at each child's age on the probability of offspring's occupation choices.[33] Together, these figures tell a somewhat different story for our other IGE curves. While our point estimates are positive for the income effects at each age, we find little evidence of an upward shape to the occupation-specific IGE curve in Figure 14 for the likelihood of white collar versus unskilled job. The marginal effects on probabilities are fairly similar for all ages, although the marginal effect at the teens is very slightly larger than the effect during early childhood, until age 17 (which differs from other findings where downturns occur at 18). As for the probability of a skilled or semiskilled job versus unskilled job, we find little evidence of income effects as the confidence intervals are quite large. Further, the estimated mobility curve, is if anything, downward sloping, which differs from our findings elsewhere.

The marginal income effects in Figure 15 reflect the IGE curve findings. We find a gently increasing effect of parental income with age for white collar jobs, negative effects for skilled/semiskilled jobs, and no effect for unskilled jobs. This is not paradoxical in the sense that, for the income trajectory we employ as a baseline, higher probabilities of white collar jobs need to be offset and in this case are matched by reductions in skilled-semiskilled probabilities. These results are not obvious from the effects in Figure 14, and are not easily

[33]The marginal effects are evaluated at the mean of cross-sectional parental income distribution at each age from birth to 19. At the mean of the parental income distribution, the probabilities of offspring having a job in Type 1, 2, and 3 equal 61.74%, 29.49%, and 8.77%, respectively.

interpreted, although there is no logical contradiction.

7 Conclusion

This paper makes a sustained argument that measurement of intergenerational mobility can be enhanced by an explicit consideration of how the dynamics of family income over childhood and adolescence predict adult outcomes. Our approach develops mobility curves, that measure the marginal sensitivity of offspring outcomes to family income at each age. These constitute a strict generalization of the conventional mobility measure, the intergenerational elasticity of income. To implement this perspective, we proposed the use of Functional Data Analysis methods, which explicitly treat outcomes as functions of trajectories. We argue that a functional principal component approach to FDA provides a way to accurately estimate the effects of family income period by period, despite the slow moving nature of such series. We further show how this approach can be generalized to a quadratic trajectories model that allows for interactions between income trajectory values at different ages of children. As such, the quadratic model provides a way of measuring how values of family incomes exhibit complementarity or substitutability between different ages.

Applying this functional data analysis strategy to map parental trajectories to child permanent income, using the Panel Study of Income Dynamics, we draw four broad conclusions about the intergenerational mobility process.

First, income at different stages of childhood and adolescence produces different marginal sensitivities of offspring's permanent income to parental income at different ages. The magnitudes of these sensitivities are nearly monotonically increasing with age, with a decrease after age 18. Therefore, incomes in adolescence have larger marginal predictive effects on children's income than do incomes in early childhood. This upward slope also implies that the likelihood of offspring's upward mobility depends on the growth rate of parental income, not just its level.

Second, we find confirmation of these basic patterns when our sample is divided into 1967-1970, 1971-1973, and 1974-1977 cohorts. However, we find that the mobility curves exhibit some differences. The integral of the mobility curve is larger for the first cohort than later ones, implying the sensitivity of children to a permanent increase in parental income starting at birth is smaller. We also find less growth in the later mobility curves in later childhood and adolescence.

Third, using a quadratic income trajectory model, we are able to uncover interactions between incomes at different ages as a distinct determinant of offspring's permanent income on parental income trajectories. We find evidence of interactions between parental incomes at different ages in terms of their effects on children. These interactions exhibit substitutability between parental incomes within the early childhood and within parental incomes during adolescence. On the other hand, complementarities are present for incomes between these two periods. Hence we say that there the intergenerational transmission process exhibits local substitutability and global complementarity with respect to parental incomes at different ages.

Fourth, we apply these same techniques to study the effects of family income trajectories on two basic proximate causes of offspring's permanent income: education and occupational status. We find that family income trajectories exhibit similar influences on education as they do for income. This represents an important confirmation of our income results. Our results for occupations are mixed and imprecise, and suggest that a more complicated model for income effects on occupation is needed.

Relative to other studies that have tried to map the timing of parental income to offspring outcomes, perhaps the most striking major difference between our results and previous work is the finding of a monotonic income mobility curve, one which suggests an especially important role for incomes in adolescence. In our judgment, these results suggest the need to consider the different roles parental income plays across the life course in affecting children. By this, we mean that there is an important difference between direct parental investments and investments that determine a child's neighborhoods and schools that are elided when one studies income to income mobility. Our evidence of increasing income sensitivity may be due to the increasing the importance of social influences on adolescents. This conjecture is consistent with work such as Wodtke et al. (2011) and Wodtke et al. (2016), who show that the relative importance of family versus neighborhood factors decreases between childhood and adolescence. A natural sequel to this paper is the exploration of the role of social trajectories, specifically characteristics of residential neighborhoods and schools which children attend, as well as family income trajectories as is done here.

Appendix

A Functional vs. Conventional Regression Estimates

In this section, we will use our main model for intergenerational mobility, which defines the children's income as (y_i) and parents' income profiles as (x_i), and empirically evaluate and compare our novel functional regression (1) and the conventional regression (7). We consider various choices of m, which denotes the number of functional principal components used to approximate parents' income profiles for our functional regression, and also the number of age sub-intervals made in the partition of the total range of ages in the parents' income profiles for the conventional regression. Note that m denotes the number of regressors in both of our functional regression and the conventional regression, although they are generally set to be different in our subsequent discussions. In our data set, we have 20 years of parents' income profiles, and therefore, we consider the conventional regression with $m = 1, 2, 4, 5, 10$ and 20, which correspond to the partitions of the interval covering entire 20 years by sub-intervals of lengths $20, 10, 5, 4, 2$ and 1, respectively. For our functional regression, we consider $m = 1, \ldots, 20$.

Throughout this section, we let $(v_j)_{j=1}^m$ be the functional basis defined in (14) used in the conventional regression for intergenerational mobility, in contrast with the functional principal component basis $(v_j^*)_{j=1}^m$ used in our novel functional approach. The conventional regression and our functional regression approximate the functional coefficient β by $\bar{\beta} = \Pi_m \beta$ and $\bar{\beta}^* = \Pi_m^* \beta$, which are estimated by $\hat{\beta}$ and $\hat{\beta}^*$ from the regression in (12) with covariate $\big((f_i)\big)$ obtained from $\big(\Pi_m f_i\big)$ and $\big(\Pi_m^* f_i\big)$, respectively. The estimator $\hat{\beta}$ may also be defined equivalently as

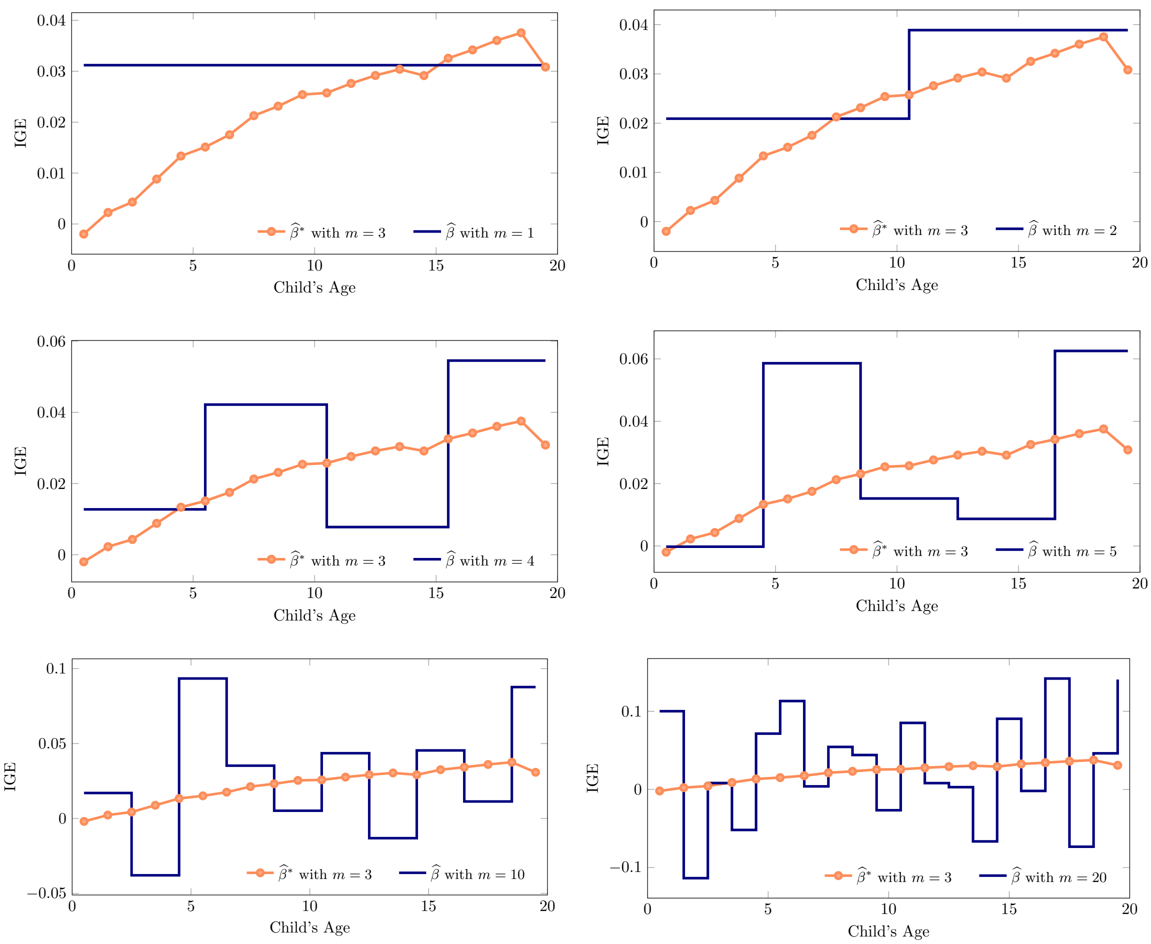
$$\hat{\beta}(r) = \sum_{j=1}^{m} \frac{\hat{\beta}_j}{q_j - p_j} 1\{p_j \leqslant r < q_j\}, \tag{26}$$

where $(\hat{\beta}_j)$ are the least squares estimators of the regression coefficients (β_j) in the conventional regression. Subsequently, we compare $\hat{\beta}$ and $\hat{\beta}^*$ as the estimators of $\bar{\beta}$ and $\bar{\beta}^*$ in terms of their precision and stability.[34]

This would be very informative, since their actual estimates are hugely different and not comparable in any meaningful sense. The actual estimates of the functional coefficients $\hat{\beta}$ and $\hat{\beta}^*$ are presented in Figure 16, where we present $\hat{\beta}^*$ from our functional regression with $m = 3$, which was set by a cross-validation method as explained in the next section,

[34] In our subsequent comparisons, we let (f_i) be discretely observed at $r = 0, \ldots, 19$ and dr be the counting measure. This is to avoid the effect of errors involved in approximating discrete observations by continuous functions for a more direct comparison.

Figure 16: Estimates of $\hat{\beta}^*(\cdot)$ and $\hat{\beta}(\cdot)$



Notes: This figure compares functional regression coefficient $\hat{\beta}^*(\cdot)$ (orange-circle line) and $\hat{\beta}(\cdot)$ (navy solid line) in (26). $\hat{\beta}^*(\cdot)$ is estimated using three leading functional principal components ($m = 3$), while $\hat{\beta}(\cdot)$ is obtained with various choice of m. Each panel of this figure compares $\hat{\beta}^*$ with $m = 3$ and $\hat{\beta}$ with a choice of $m = 1$ (upper-left), $2, 4, 5, 10$ and 20 (lower-right).

and $\hat{\beta}$ from the conventional regression with various choices of m. We also compute the value of $\delta(\hat{\beta}, \hat{\beta}^*) = \|\hat{\beta} - \hat{\beta}^*\|/\|\hat{\beta}^*\|$ to measure how distinct $\hat{\beta}$ is from $\hat{\beta}^*$, which is given by $\delta(\hat{\beta}, \hat{\beta}^*) = 0.56, 0.39, 0.75, 1.00, 1.54$ and 2.72 for $m = 1, 2, 4, 5, 10$ and 20, respectively. The distance between $\hat{\beta}$ and $\hat{\beta}^*$ is already more than half of the magnitude of $\hat{\beta}^*$ for $m = 1$, and increases more than one-and-half and two times of the magnitude of $\hat{\beta}^*$ respectively for $m = 10$ and 20. In fact, $\hat{\beta}$ looks nonsensical for $m = 10$ and 20.[35] The diagnostic results for our functional regression and the conventional regression are summarized and compared in Table 4.

The comparisons between $\hat{\beta}$ and $\hat{\beta}^*$ are made explicitly in exploring three different aspects

[35] Our functional regression can yield unstable and uninterpretable estimates as m becomes large, although these estimates are not as erratic as as those from the conventional regression.

Table 4: Comparisons of $\hat{\beta}^*$ and $\hat{\beta}$

	m	1	2	3	4	5	10	20
$\hat{\beta}^*$	FR2	0.61	0.70	0.77	0.80	0.83	0.93	1
	IVAR ($\times 10^{-4}$)	1.06	7.59	17.21	33.82	53.25	229.53	1133.88
$\hat{\beta}$	FR2	0.59	0.67		0.76	0.79	0.88	1
	IVAR ($\times 10^{-4}$)	1.01	9.73		49.60	75.70	337.49	1133.88
	IBS ($\times 10^{-5}$)	7.29	10.42		2.00	0.82	0.31	0

Notes: Presented are the FR2 and IVAR of $\hat{\beta}^*$ based on our functional regression, and the FR2, IVR and IBS of $\hat{\beta}$ obtained from the conventional regression, both with various choices of the number m of regressors.

of the two estimators. First, we define

$$\rho^2 = \frac{\sum_{i=1}^n \left\|\Pi_m f_i\right\|^2}{\sum_{i=1}^n \|f_i\|^2} = \frac{\operatorname{trace}\left(\Pi_m Q \Pi_m\right)}{\operatorname{trace} Q}, \tag{27}$$

and denote by ρ_*^2 the corresponding value obtained with Π_m^* in place of Π_m. For each m, ρ^2 and ρ_*^2 represent the proportions of the total variation of (f_i) explained by $\left(\Pi_m f_i\right)$ and $\left(\Pi_m^* f_i\right)$ in the conventional regression and our functional regression, respectively, which will be referred to as the functional R-square (FR2) of $\hat{\beta}$ and $\hat{\beta}^*$. Note that $\operatorname{trace}\left(\Pi_m^* Q \Pi_m^*\right) = \sum_{j=1}^m \lambda_j^*$ and $\operatorname{trace}(Q) = \sum_{j=1}^n \lambda_j^*$, where $(\lambda_j^*)_{j=1}^n$ are the eigenvalues of Q defined in descending order. As a result, ρ_*^2 in our functional regression increases monotonically and approaches unity as m gets large. This is not necessarily true for ρ^2 in the conventional regression, although we expect that it also normally increases with m and approaches one.[36]

As shown in Table 4, the FR2 of $\hat{\beta}$, as well as the FR2 of $\hat{\beta}^*$, increases up to unity as m approaches 20, although the latter has a slightly faster rate of convergence. It is worth noting that the leading basis, which yields the temporal means of parental incomes over the entire range of children's age not only in the conventional regression but also in our functional regression, alone has already contributed around 60% to FR2. The marginal increase in FR2 is relatively much smaller for $m \geqslant 2$. This means that we need to increase m as much as we can to get information on the differing effects of parental income at different ages of their children. Of course, we may increase the FR2s of both $\hat{\beta}$ and $\hat{\beta}^*$ simply by setting m large both in our functional regression and in the conventional regression. However, this comes at a cost. As m gets large, the IVAR of an estimator for the functional regression coefficient β increases, due to the so-called ill-posed inverse problem in functional regressions. The

[36] The basis $(v_j)_{j=1}^m$ for the conventional regression are defined differently for different m's and should be denoted as $(v_{mj})_{j=1}^m$ to be more precise, although throughout we have suppressed their dependence on m for notational brevity.

problem can be fatal if the total variation of (f_i) is explained by only a few dominant factors and we set m larger than the number of those dominant factors.

For the data we analyze, this problem can be seen. The IVARs of $\hat{\beta}$ and $\hat{\beta}^*$ increase each sharply as m gets large. The marginal cost of increasing FR2 by setting m large is therefore high. This is the reason why the choice of basis is critical. As mentioned earlier, we set $m = 3$ as selected by the cross-validation in our empirical study reported in the next section, and with this choice of $m = 3$, the FR2 of our functional regression is 77%. To achieve the same level of FR2 in the conventional regression, we need to set $m = 4$ or larger. Our functional regression with $m = 3$ is therefore largely comparable with the conventional regression with $m = 4$ in the FR2 sense. However, as shown in Table 4, $\hat{\beta}$ in the conventional regression has IVAR which is about three times larger than that of $\hat{\beta}^*$ in our functional regression. Our functional regression is therefore by far more efficient than the conventional regression.

Last, we also consider the integrated bias squared (IBS) of $\hat{\beta}$ to see how large the bias could be. As noted, $\hat{\beta}^*$ is unbiased and it has no bias term. The IBS of $\hat{\beta}$ depends on the unknown functional parameter β and we use as a proxy our functional regression estimator $\hat{\beta}^*$ with $m = 3$. The magnitude of the IBS of $\hat{\beta}$ appears to be substantial when m is set to be too small. The IBS is substantially bigger than the IVAR for $m = 1$, and the IBS is still much larger than the half of the IVAR for $m = 2$. It becomes, however, unimportant and negligible for m bigger than 4, at least relative to the IVAR.

B Comparisons with Approaches Using Trigonometric or B-Spline Basis Functions

In our empirical analyses, we either regard functions observed only at discrete set of points simply as functions defined only at those discrete points or use Daubechies 3 tap wavelets to convert them as functions defined continuously, and then use the functional principal component basis to approximate and represent them as finite-dimensional vectors. This is to minimize the potential distortions made by arbitrarily choosing a basis to define, approximate and represent functional variables, which has a critical impact on the finite sample performance of the estimators of parameters in functional models.

In this section, we consider using trigonometric and B-spline functions as basis to define, approximate, and represent functional variables, which are most commonly used in the functional data analysis. We mainly consider the use of these functions as a basis for the representation of our functional covariate as a finite-dimensional vector. In particular, we compare their effectiveness in representing cross-sectional variations of the functional covariate, and MSEs of the resulting estimators of the functional coefficient. These basis functions are also used to define functions continuously from observations made at discrete points, for which we use Daubechies 3 tap wavelets. However, the choice of functions used in this smoothing context is not important and yields only marginal effects. Their comparison is therefore not reported in the paper. See Ramsay and Silverman (2005).

For the trigonometric basis, we add constant and linear functions to more effectively represent our functional data, some of which exhibit strong trends over age. The basis will be referred to as the flexible Fourier basis. The first and second basis functions are given by constant and linear functions, respectively, which are followed by cosine and sine functions. The B-spline basis is defined specifically by an order k and a knot sequence $(r_0, \ldots, r_\ell)$. We set $k = m$ and choose no interior point for $1 \leqslant m \leqslant 4$, and set $k = 4$ and choose $m - k$ interior points for $m \geqslant 5$. For our comparison, we consider the estimators of functional coefficient β in our benchmark functional regression model (1) relying on the flexible Fourier basis and B-spline basis, and compare them with our estimate relying in particular on the functional principal component basis.

Our results are summarized in Table 5, where we present the FR2, IVAR, IBS, and LOOCV errors for each choice of basis, as well as the number m of basis functions used. Here we report the results obtained from the functional regression with observed parental income trajectories. For each m, all the results are largely comparable, and there are no meaningful differences among three approaches although the approach used in the paper yields slightly better results generally in all aspects. What makes our approach distinguishable from the

Table 5: Comparisons of Functional Coefficient Estimators with Observed Parental Income Trajectories

FPCA	1 basis	2 basis	**3 basis**	4 basis	5 basis	6 basis	7 basis	8 basis	9 basis	10 basis
FR2	0.61	0.70	**0.77**	0.80	0.83	0.86	0.88	0.90	0.92	0.93
IVAR ($\times 10^{-3}$)	0.10	0.76	**1.72**	3.38	5.32	7.56	10.33	13.64	17.87	22.95
LOOCV	0.4515	0.4479	**0.4444**	0.4452	0.4459	0.4468	0.4464	0.4461	0.4460	0.4471
B-Spline	1 basis	2 basis	3 basis	4 basis	5 basis	6 basis	**7 basis**	8 basis	9 basis	10 basis
FR2	0.59	0.70	0.76	0.79	0.82	0.85	**0.87**	0.88	0.90	0.91
IVAR ($\times 10^{-3}$)	0.10	0.80	1.82	3.73	5.90	8.78	**11.64**	15.72	20.03	25.58
IBS ($\times 10^{-6}$)	53.90	1.64	0.13	0.26	0.71	0.79	**0.19**	0.12	0.11	0.13
LOOCV	0.4445	0.4449	0.4459	0.4461	0.4465	0.4469	**0.4438**	0.4449	0.4460	0.4468
Flexible Fourier	**1 basis**	2 basis	3 basis	4 basis	5 basis	6 basis	7 basis	8 basis	9 basis	10 basis
FR2	**0.59**	0.70	0.75	0.79	0.82	0.84	0.86	0.88	0.89	0.90
IVAR ($\times 10^{-3}$)	**0.10**	0.80	1.89	3.79	6.12	8.93	12.51	16.97	21.88	27.19
IBS ($\times 10^{-6}$)	**53.90**	1.64	0.89	2.11	1.31	3.92	3.28	3.69	3.14	2.44
LOOCV	**0.4445**	0.4449	0.4460	0.4460	0.4472	0.4483	0.4471	0.4482	0.4484	0.4492

Notes: Presented are the FR2, IVAR and IBS of the functional coefficient estimators and the LOOCV errors obtained from employing the flexible Fourier basis and B-spline basis, which are compared with those of our estimator, for various choices of m. The observed parental income trajectories are used.

other two is the leave-one-out cross-validation result it provides. As discussed, our approach yields a very well defined LOOCV error function, which is minimized uniquely at $m = 3$. On the other hand, the use of the flexible Fourier or B-spline basis does not produce such a nicely defined LOOCV error function. The former finds $m = 1$, which completely ignores age heterogeneity with unacceptably large bias. The latter finds $m = 7$, which yields a functional coefficient estimator having about seven times larger variance than our estimator with no meaningful age heterogeneity pattern.

The same comparison is made for the functional regression run by the extracted permanent parental income trajectories, in place of the observed parental income trajectories, and the results are summarized in Table 6. As in the functional regression estimated using the observed parental income trajectories, the differences in the FR2, IVAR and IBS of the functional coefficient estimators are small and do not appear to be important for each m, although the use of the functional principal component basis generally provides a better estimator. However, the leave-one-out cross-validation procedure finds quite different values of m if the flexible Fourier or B-spline basis is used in this case, while our approach selects the same value of m, i.e., $m = 3$ as in the previous case. Our approach seems rather robust in this regard. The flexible Fourier basis and B-spline basis yield $m = 5$ and $m = 4$, respectively, which seem more appropriate than before and the estimated functional coefficients with these values of m produce more sensible functional coefficient estimators. However, the estimators are neither stable nor reliable, given their large variances.

Table 6: Comparisons of Functional Coefficient Estimators with Extracted Permanent Parental Income Trajectories

FPCA	1 basis	2 basis	**3 basis**	4 basis	5 basis	6 basis	7 basis	8 basis	9 basis	10 basis
FR2	0.65	0.76	**0.83**	0.87	0.90	0.92	0.93	0.95	0.96	0.97
IVAR ($\times 10^{-3}$)	0.10	0.68	**1.71**	3.37	5.79	8.75	12.94	18.20	24.51	31.18
LOOCV	0.4492	0.4487	**0.4432**	0.4438	0.4433	0.4435	0.4446	0.4438	0.4451	0.4460
B-Spline	1 basis	2 basis	3 basis	**4 basis**	5 basis	6 basis	7 basis	8 basis	9 basis	10 basis
FR2	0.63	0.74	0.81	**0.85**	0.88	0.91	0.93	0.94	0.95	0.96
IVAR ($\times 10^{-3}$)	0.11	0.82	1.96	**4.22**	6.94	10.34	14.51	20.17	27.13	35.39
IBS ($\times 10^{-6}$)	59.80	27.22	6.90	**0.97**	4.82	0.25	0.13	0.22	0.17	0.10
LOOCV	0.4449	0.4443	0.4454	**0.4419**	0.4431	0.4442	0.4428	0.4442	0.4454	0.4469
Flexible Fourier	1 basis	2 basis	3 basis	4 basis	**5 basis**	6 basis	7 basis	8 basis	9 basis	10 basis
FR2	0.63	0.74	0.80	0.84	**0.87**	0.90	0.91	0.93	0.94	0.96
IVAR ($\times 10^{-3}$)	0.11	0.82	2.21	4.10	**7.19**	9.80	15.55	20.15	28.12	34.60
IBS ($\times 10^{-6}$)	59.80	27.22	23.56	14.13	**12.99**	7.50	7.63	1.92	1.78	0.15
LOOCV	0.4449	0.4443	0.4456	0.4423	**0.4420**	0.4431	0.4445	0.4451	0.4465	0.4463

Notes: Presented are the FR2, IVAR and IBS of the functional coefficient estimators and the LOOCV errors obtained from employing the flexible Fourier basis and B-spline basis, which are compared with those of our estimator, for various choices of m. The permanent parental income trajectories are extracted and used.

C Derivations of Econometric Results in Section 2.3

Let $\llparenthesis\cdot\rrparenthesis$ and $\llparenthesis\cdot\rrparenthesis^*$ be defined as in Section 2.2 by the isometries $\pi : H_m \to R^m$ and $\pi^* : H_m^* \to R^m$, respectively, where H_m is the subspace of H spanned by $(v_j)_{j=1}^m$ and H_m^* is the subspace of H spanned by $(v_j^*)_{j=1}^m$. It follows that $\pi(f) = \llparenthesis f\rrparenthesis$ and $\pi^*(f) = \llparenthesis f\rrparenthesis^*$ for $f \in H_m$ and $f \in H_m^*$, respectively, and conversely, for $c = (c_1, \ldots, c_m)' \in R^m$, we have $\pi^{-1}(c) = \sum_{j=1}^m c_j v_j$ and $\pi^{*-1}(c) = \sum_{j=1}^m c_j v_j^*$. Moreover, we also use π and π^* to denote one-to-one correspondences in the vector space of linear operators on H_m and H_m^* with the vector space of $m \times m$ matrices, respectively. They also become isometries if we introduce suitable norms in these spaces. For linear operators A on H_m and A^* on H_m^*, we define $\pi(A) = \llparenthesis A\rrparenthesis$ and $\pi^*(A^*) = \llparenthesis A\rrparenthesis^*$ to be $m \times m$ matrices such that

$$\llparenthesis Af\rrparenthesis = \llparenthesis A\rrparenthesis\llparenthesis f\rrparenthesis \quad \text{and} \quad \llparenthesis Af\rrparenthesis^* = \llparenthesis A\rrparenthesis^*\llparenthesis f\rrparenthesis^*$$

for all $f \in H_m$ or $f \in H_m^*$, respectively. It is easy to show that, for an m-dimensional square matrix $C = (c_{jk})$, we have

$$\pi^{-1}(C) = \sum_{j,k=1}^m c_{jk}(v_j \otimes v_k) \quad \text{and} \quad \pi^{*-1}(C) = \sum_{j,k=1}^m c_{jk}(v_j^* \otimes v_k^*).$$

The reader is referred to Chang et al. (2021) for more details.

Note that

$$
\begin{aligned}
\llparenthesis\hat{\beta}\rrparenthesis &= \left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\llparenthesis f_i\rrparenthesis'\right)^{-1}\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis y_i\right)\\
&= \llparenthesis\beta\rrparenthesis + \left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\llparenthesis f_i\rrparenthesis'\right)^{-1}\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\langle\beta,(1-\Pi_m)f_i\rangle\right) + \left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\llparenthesis f_i\rrparenthesis'\right)^{-1}\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\varepsilon_i\right)
\end{aligned}
$$

and

$$
\llparenthesis\hat{\beta}\rrparenthesis^* = \left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis^*\llparenthesis f_i\rrparenthesis^{*\prime}\right)^{-1}\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis^* y_i\right) = \llparenthesis\beta\rrparenthesis^* + \left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis^*\llparenthesis f_i\rrparenthesis^{*\prime}\right)^{-1}\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis^*\varepsilon_i\right),
$$

and that

$$
\hat{\beta} = \pi^{-1}\big(\llparenthesis\hat{\beta}\rrparenthesis\big) \quad \text{and} \quad \hat{\beta}^* = \pi^{*-1}\big(\llparenthesis\hat{\beta}\rrparenthesis^*\big).
$$

Therefore, we have

$$
\begin{aligned}
\|\hat{\beta}-\mathbb{E}\hat{\beta}\| = \|\llparenthesis\hat{\beta}\rrparenthesis-\mathbb{E}\llparenthesis\hat{\beta}\rrparenthesis\| &= \left\|\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\llparenthesis f_i\rrparenthesis'\right)^{-1}\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\varepsilon_i\right)\right\|\\
\|\mathbb{E}\hat{\beta}-\bar{\beta}\| = \|\mathbb{E}\llparenthesis\hat{\beta}\rrparenthesis-\llparenthesis\beta\rrparenthesis\| &= \left\|\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\llparenthesis f_i\rrparenthesis'\right)^{-1}\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\langle\beta,(1-\Pi_m)f_i\rangle\right)\right\|,
\end{aligned}
$$

and

$$
\|\hat{\beta}^*-\bar{\beta}^*\| = \|\llparenthesis\hat{\beta}\rrparenthesis^*-\llparenthesis\beta\rrparenthesis^*\| = \left\|\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis^*\llparenthesis f_i\rrparenthesis^{*\prime}\right)^{-1}\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis^*\varepsilon_i\right)\right\|,
$$

since π is an isometry.

However, we have

$$
\begin{aligned}
\mathbb{E}\left\|\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\llparenthesis f_i\rrparenthesis'\right)^{-1}\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\varepsilon_i\right)\right\|^2 &= \sigma^2\,\mathrm{trace}\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\llparenthesis f_i\rrparenthesis'\right)^{-1}\\
\mathbb{E}\left\|\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis^*\llparenthesis f_i\rrparenthesis^{*\prime}\right)^{-1}\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis^*\varepsilon_i\right)\right\|^2 &= \sigma^2\,\mathrm{trace}\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis^*\llparenthesis f_i\rrparenthesis^{*\prime}\right)^{-1}
\end{aligned}
$$

and

$$\pi^{-1}\left(\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\llparenthesis f_i\rrparenthesis'\right)^{-1}\right)=\left(\sum_{i=1}^{n}\big(\Pi_m f_i\otimes\Pi_m f_i\big)\right)^{+}=(\Pi_m Q\Pi_m)^{+}$$

$$\pi^{*-1}\left(\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis^{*}\llparenthesis f_i\rrparenthesis^{*\prime}\right)^{-1}\right)=\left(\sum_{i=1}^{n}\big(\Pi_m^{*} f_i\otimes\Pi_m^{*} f_i\big)\right)^{+}=(\Pi_m^{*} Q\Pi_m^{*})^{+}.$$

Moreover, we have

$$\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\llparenthesis f_i\rrparenthesis'\right)^{-1}\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\langle\beta,(1-\Pi_m)f_i\rangle\right)=\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\llparenthesis f_i\rrparenthesis'\right)^{-1}\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\langle\beta,f_i\rangle\right)-\llparenthesis\beta\rrparenthesis$$

and

$$\pi^{1}\left(\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\llparenthesis f_i\rrparenthesis'\right)^{-1}\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\langle\beta,f_i\rangle\right)\right)=\pi^{-1}\left(\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\llparenthesis f_i\rrparenthesis'\right)^{-1}\right)\pi^{-1}\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\langle\beta,f_i\rangle\right)$$

with

$$\pi^{-1}\left(\sum_{i=1}^{n}\llparenthesis f_i\rrparenthesis\langle\beta,f_i\rangle\right)=\sum_{j=1}^{m}\left(\sum_{i=1}^{n}\langle v_j,f_i\rangle\langle\beta,f_i\rangle\right)v_j=\sum_{j=1}^{m}\big\langle v_j,Q\beta\big\rangle v_j=\Pi_m Q\beta$$

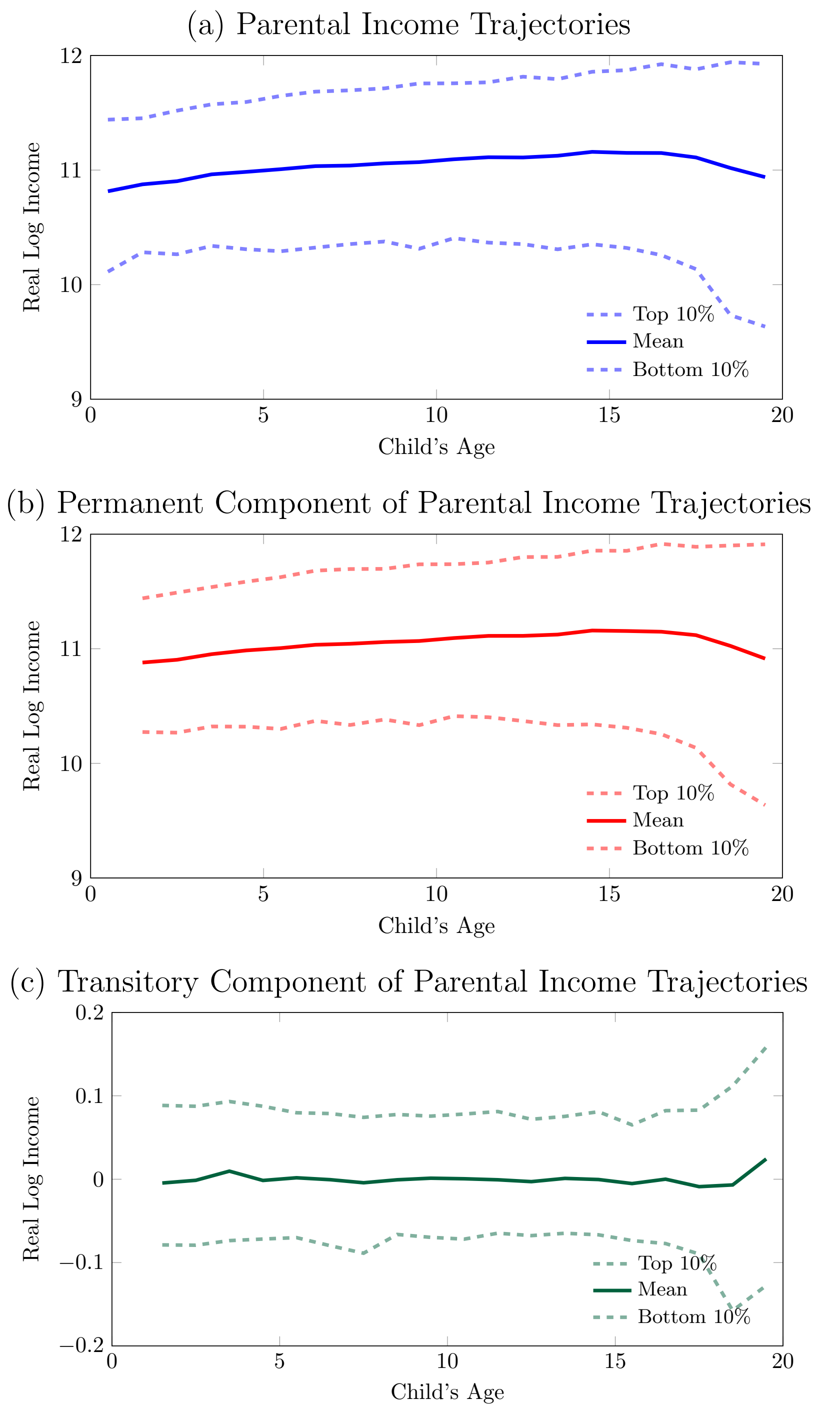
and $\pi^{-1}\big(\llparenthesis\beta\rrparenthesis\big)=\Pi_m\beta$, from which it follows that

$$\mathbb{E}\hat{\beta}-\bar{\beta}=(\Pi_m Q\Pi_m)^{+}\Pi_m Q(1-\Pi_m)\beta,$$

as was to be shown.

D Additional Figure

Figure 17: Parental Income Trajectories and Permanent/Transitory Component



Notes: This figure presents parental income trajectories (top panel) and the extracted permanent component (middle panel) and transitory component (bottom) of trajectories. The solid line in each panel presents the mean of parental income (or its component) at each child's age. The dashed lines above and below the solid line indicate the 90th and the 10th percentiles, respectively.

References

[1] Ashworth, J., V. J. Hotz, A. Maurel, and T. Ransom (2021). Changes across cohorts in wage returns to schooling and early work experiences. Journal of Labor Economics 39(4), 931–964.

[2] Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor model. Econometria 70, 191–221.

[3] Blundell, R. and I. Preston (1998). Consumption inequality and income uncertainty. Quarterly Journal of Economics, 603 – 640.

[4] Bosq, D. (2000). Linear Processes in Function Spaces. Springer-Verlag.

[5] Cameron, S. V. and J. J. Heckman (2001). The dynamics of educational attainment for black, hispanic, and white males. Journal of Political Economy 109(3), 455–499.

[6] Carneiro, P., I. L. García, K. G. Salvanes, and E. Tominey (2021). Intergenerational mobility and the timing of parental income. Journal of Political Economy 129(3), 757–788.

[7] Caucutt, E. and L. Lochner (2020). Early and late human capital investments, borrowing constraints, and the family. Journal of Political Economy 128, 1065–1147.

[8] Chang, Y., M. Choi, and J. Y. Park (2021). A factor model for functional time series. Working Paper, Indiana University.

[9] Chang, Y., J. Y. Park, and D. Pyun (2021). From functional autoregressions to vector autoregressions. Working Paper, Indiana University.

[10] Chaussé, P., T. Chen, and K. A. Couch (2021). Assessing inter-generational mobility: A functional data regression approach. Working Paper, University of Waterloo.

[11] Cheng, S. and X. Song (2019). Linked lives, linked trajectories: Intergenerational association of intragenerational income mobility. American Sociological Review 84(6), 1037–1068.

[12] Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. Quarterly Journal of Economics 129(4), 1553–1623.

[13] Corak, M. (2006). Do poor children become poor adults? Lessons for public policy from a cross country comparison of generational earnings mobility. Research on Economic Inequality 13: Dynamics of Inequality and Poverty. J. Creedy and G. Kalb, eds.

[14] Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. Journal of Economic Perspectives 27, 79–102.

[15] Cunha, F. and J. J. Heckman (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. Journal of Human Resources 43(4), 738–782.

[16] Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. Econometrica 78(3), 883–931.

[17] Davis, J. and B. Mazumder (2022). The decline in intergenerational mobility after 1980. Working Paper, WP 2017-05 (revised), Federal Reserve Bank of Chicago.

[18] Duncan, G. J., K. M. Ziol-Guest, and A. Kalil (2010). Early-childhood poverty and adult attainment, behavior, and health. Child Development 81(1), 306–325.

[19] Durlauf, S. N., A. Konstantinidi, and A. Kourtellos (2021). Intergenerational trajectories. Working Paper, University of Chicago.

[20] Durlauf, S. N., A. Kourtellos, and C. M. Tan (2022). The great gatsby curve. Annual Review of Economics 14, 571–605.

[21] Grawe, N. D. (2004). Reconsidering the use of nonlinearities in intergenerational earnings mobility as a test for credit constraints. Journal of Human Resources 39(3), 813–827.

[22] Guest, A. M., N. S. Landale, and J. C. McCann (1989). Intergenerational occupational mobility in the late 19th century united states. Social Forces 68(2), 351–378.

[23] Haider, S. and G. Solon (2006). Life-cycle variation in the association between current and lifetime earnings. American Economic Review 96(4), 1308–1320.

[24] Hall, P. and J. L. Horowitz (2007). Methodology and convergence rates for functional linear regression. Annals of Statistics 35, 70–91.

[25] Heckman, J. J. and S. Mosso (2014). The economics of human development and social mobility. Annual Review of Economics 6(1), 689–733.

[26] Lefgren, L., D. Sims, and M. J. Lindquist (2012). Rich dad, smart dad: Decomposing the intergenerational transmission of income. Journal of Political Economy 120(2), 268–303.

[27] Lochner, L. and A. Monge-Narano (2012). Credit constraints in education. Annual Review of Economics 4, 225–256.

[28] Long, J. and J. Ferrie (2013). Intergenerational occupational mobility in great britain and the united states since 1850. American Economic Review 103(4), 1109–37.

[29] MaCurdy, T. E. (1982). The use of time series processes to model the error structure of earnings in a longitudinal data analysis. Journal of Econometrics 18(1), 83–114.

[30] Mazumder, B. (2005). Fortunate sons: New estimates of intergenerational mobility in the united states using social security earnings data. Review of Economics and Statistics 87(2), 235–255.

[31] Mazumder, B. (2016). Estimating the intergenerational elasticity and rank association in the united states: Overcoming the current limitations of tax data. In Inequality: Causes and consequences, Volume 43, pp. 83–129. Emerald Group Publishing Limited.

[32] Mazumder, B. and M. Acosta (2015). Using occupation to measure intergenerational mobility. The Annals of the American Academy of Political and Social Science 657(1), 174–193.

[33] Morley, J. C. (2002). A state–space approach to calculating the beveridge–nelson decomposition. Economics Letters 75(1), 123–127.

[34] Park, J. Y. and J. Qian (2012). Functional regression of continuous state distributions. Journal of Econometrics 167, 397–412.

[35] Ramsay, J. O. and B. W. Silverman (2005). Functional Data Analysis. Springer.

[36] Shigeoka, H. (2019). Long-term effect of growing up in a recession on risk preferences. National Bureau of Economic Research Working Paper.

[37] Solon, G. (1992). Intergenerational income mobility in the united states. American Economic Review, 393–408.

[38] Song, X., C. G. Massey, K. A. Rolf, and Y. Xie (2020). Long-term decline in intergenerational mobility in the united states since the 1850s. Proceedings of the National Academy of Sciences 117, 251–258.

[39] Wodtke, G. T., F. Elwert, and D. J. Harding (2016). Neighborhood effect heterogeneity by family income and developmental period. American Journal of Sociology 121(4), 1168–1222.

[40] Wodtke, G. T., D. J. Harding, and F. Elwert (2011). Neighborhood effects in temporal perspective: The impact of long-term exposure to concentrated disadvantage on high school graduation. American Sociological Review 76(5), 713–736.

[41] Yao, F. and H.-G. Müller (2010). Functional quadratic regression. Biometrika 97(1), 49–64.