ISSN 2279-9362

Parenting with Patience: Parental Incentives and Child Development

Daniela Del Boca, Christopher Flinn, Ewout Verriest and Matthew Wiswall

No. 698
September 2023

Carlo Alberto Notebooks

www.carloalberto.org/research/working-papers

Parenting with Patience: Parental Incentives and Child Development *

Daniela Del Boca
University of Torino
Collegio Carlo Alberto

Christopher Flinn
New York University
Collegio Carlo Alberto

Ewout Verriest
Penn State University

Matthew Wiswall
University of Wisconsin-Madison
NBER

August 2023
JEL Classification J13, D1
Keywords: Time allocation, Child development, Parenting styles

Abstract

We construct a dynamic model of child development where forward-looking parents and children jointly take actions to increase the child's cognitive and non-cognitive skills within a Markov Perfect Equilibrium framework. In addition to time and money investments in their child, parents also choose whether to use explicit incentives to increase the child's self-investment, which may reduce the child's future intrinsic motivation to invest by reducing the child's discount factor. We use the estimated model parameters to show that the use of extrinsic motivation has large costs in terms of the child's future incentives to invest in themselves.

*This research is partially supported by NSF Grant 1357686, the C.V. Starr Center for Applied Economics, and the Institute of Human Development and Social Change at N.Y.U. For useful comments and suggestions we are grateful to three anonymous referees, Corina Boar, Jim Heckman, John Eric Humphries, John List, and Fabrizio Zilibotti, as well as to participants in seminars at Syracuse, Minnesota, the New York Fed, the 2019 PennCorn conference, the 2020 ASSA meetings and the CEHD at Chicago. This research utilizes the micro data from the Preference Parameters Study of Osaka University's 21st Century COE Program "Behavioral Macro-Dynamics Based on Surveys and Experiments," its Global COE project "Human Behavior and Socioeconomic Dynamics," and JSPS KAKENHI 15H05728 "Behavioral-Economic Analysis of Long-Run Stagnation." We are also indebted to Professor Laurence Steinberg of Temple University for permission to use the field data collected in connection with the research presented in Steinberg et al. (2009). We are responsible for all errors, interpretations, and omissions.

1 Introduction

In most models of human capital investment, the subject of the investments is either a young and passive agent or an adult making investments to increase their own productivity.[1] Within the child development literature, the vast majority of theoretical models and empirical studies have considered children to be passive agents, especially when they are very young, and view parents as the only active decision-makers.[2] Empirical evidence suggests that the effect of parental investments on child development declines during adolescence (Carneiro et al. (2003), Del Boca et al. (2014)), whereas the effect of time spent in the child's self-investment (e.g., homework) increases (Cooper et al. (2006)). It is during adolescence, in fact, when most teenagers begin taking responsibility for their own actions, and when they begin to increase their investments in themselves (Dauphin et al. (2011), Lundberg et al. (2009), Kooreman (2007), Del Boca et al. (2017), Fiorini and Keane (2014)). A natural and potentially important question to address is the evolution of the individual from a passive receiver of investment inputs into one whose actions at least partially determine their own rate of cognitive development.

We develop a model of child and parent interactions in which both are active agents in the child's cognitive and non-cognitive development processes. In our model not only do children choose how much time to devote to their own cognitive development, but parents can influence their children's development by providing goods and time expenditures directly but also in two other important ways. In our model, parents can incentivize their children to increase their self-investment time, which then increases the child's cognitive ability. In addition, parents have the ability to influence the development of their child's non-cognitive characteristics through their choice of a "parenting style." The non-cognitive characteristic upon which we focus is the child's time preference, as measured by their discount factor. An increase in the child's discount factor fosters future self-investments and cognitive development. Our model then expresses a fundamental trade-off: incentives for child self-investment foster short-term gains in cognitive skills but can reduce long-term skills by negatively affecting non-cognitive development and the child's propensity to invest in their own development.

We interpret our model and results in terms of the relationship between intrinsic and extrinsic motivation in the child development process, which contributes to a lengthy research literature in developmental psychology that dates back at least to Deci (1971). The model

[1]For example, Becker's model of fertility and investment in children views parents as choosing a quantity of offspring and an average quality (see, e.g., Becker and Tomes, 1976)). The canonical models of educational choice and on-the-job investment view the person in whom investments are being made as the decision-maker (e.g., Becker (1964), Ben-Porath (1967)). In some models of on-the-job investment, it is assumed that workers and firms jointly invest in the general human capital of the worker and/or match-specific capital (Wasmer (2006), Bagger et al. (2014), Lentz and Roys (2015), Flinn et al. (2017)). In this paper we limit attention to the development of cognitive human capital prior to labor market entry, so that the joint investments are limited to parents and children. We also discuss how schooling and other aspects of the child's environment affect this process.

[2]This literature is surveyed in Currie and Almond (2011).

provides a tractable micro-foundation for the complex ways in which parents influence both the cognitive and non-cognitive development of their children, and how skill development in one domain affects development in the other.[3] These features of our model are an important extension of existing empirical research in child development, including our earlier work, Del Boca, Flinn and Wiswall (2014), henceforth referred to as DFW, and more recent contributions such as Mullins (2022), Verriest (2022), and Daruich (2019).

Estimation of this model requires rich data with repeated measures of various child investments and skills, along with data on parental labor supply and resources. We combine three main sources of data. The Panel Study of Income Dynamics (PSID) and its Child Development Supplement (PSID-CDS) includes detailed time diary information on the time parents and children spend together or alone. Using detailed descriptions, certain types of these interactions we classify as investment activities. These sources also contain measures of child cognitive ability at various ages and information on parental use of self-investment incentives. One focus of our analysis is the time preferences of parents and children and the manner in which they affect the child development process. The PSID does not include systematic information on subjective discount factors. For information on the discount factors of parents we utilize relatively large sample data from the Osaka Preference Parameters Study (PPS). For information on the subjective discount factors of children we use rich data generously provided to us by Laurence Steinberg of Temple University. We develop a Method of Simulated Moments estimator that enables us to utilize information from all three data sources in estimating the parameters that characterize the model.

One motivation for our analysis is the observation that throughout childhood and adolescence the amount of time children spend "investing" alone (without their parents) increases markedly. In Figure 1a, we show that the average number of hours (per week) parents spend with their children declines with the age of their children, while at the same time the number of hours of self-investment by the children increases. The child's self-investment time as a fraction of total investment time (sum of parent's and child's time) rises from about 5 percent at age 6 to over 30 percent by the teenage years.

We find that children's self-investment time also varies with parental education and income, as well as other household characteristics that are a focus of our model. Figure 1b shows that self-investment time is greater for high SES families, as measured by parental education. By the teenage years, children in high-educated households spend about twice as many hours in self-investment as children with less-educated parents. This pattern also holds if we partition the data by household income, or if we look at child self-investment time as a fraction of total investment time. A measure that we use below that we interpret as indicating the use of extrinsic incentives by the parents is whether a child's regular allowance

[3] Cunha et al. (2010) estimate skill production functions including both cognitive and non-cognitive skills, but theirs is not a micro-founded household model of investment.

is linked to the study time of the child and their school performance more generally. We find that allowances tied to investment time have a positive impact on the contemporaneous self-investment time of the child. Moreover, the empirical patterns show that extrinsic incentives are more likely to be used by parents with younger children and by parents with lower educational attainment, which may be linked to differences in both children's and parents' intrinsic time preferences.

Our model of parent and child interaction considers a case in which the parents operate as a single decision-making unit, and we assume that they are able to choose their period t actions (labor supply, time with children, goods expenditures on children) prior to the child choosing theirs (self-investment time). The model is dynamic with two rational and forward-looking agents, parents and children. Although the parents have "private" egoistic preferences, their actual period payoff function is a weighted average of their private utility and that of the the child, with the weight on the child's utility interpreted as an altruism parameter. With respect to "private" preferences, both sets of agents value the child's cognitive ability, making this a pure public good.[4]

By explicitly incorporating the child's actions into their development process, a number of previously unexplored factors explaining the dispersion in cognitive outcomes at the end of adolescence can be investigated. When modeling the human capital investment decisions of adults (in themselves or in their children), it is typically assumed that a common discount factor is applied to an additively separable lifetime welfare function. Studies by developmental psychologists (e.g., Steinberg et al. (2009)) have demonstrated that the capacity to delay gratification changes markedly over adolescence, particularly around puberty. Because the motivation to invest depends critically on how forward-looking an agent is, it is important to allow for changes in this characteristic over the development period, and we use existing evidence to inform our parameterization of the child's age-varying discount factor sequence.

As in previous work, we also allow the skill development production function parameters to change across development periods, reflecting the possibility that time with parents may be declining in productivity as children age. In our model, we allow for the possibility that the productivity of investment by children themselves may be increasing as they age. Our model of the skill development process also allows for persistence in skill development for both time and goods (monetary) investments, and for the productivity of parental time to vary by the parent's level of human capital (as measured by their schooling attainment).

In our baseline specification, parents make their investment and consumption choices, as well as the level of expenditure on the private consumption good of the child. Given these choices, the child then chooses the amount of time to spend in study (self-investment), with the remainder of time outside of school consumed as leisure. In addition to allowing

[4]The parents' final utility function is a weighted average of the child's utility and their own private utility. Therefore all of the arguments of the child's utility function are public goods with respect to the final utility function of the parents by definition.

heterogeneity in household resources (wage offers for the mother and father, and non-labor income) and investment time productivity, the model allows for heterogeneity in household preferences, in particular by allowing the child's initial discount factor as well as the parents' time-invariant discount factor to be correlated with their educational attainment level.

We generalize the model by allowing parents to choose to incentivize their children to provide different amounts of self-investment time than they do in the Stackelberg equilibrium without such incentives. Incentives take the form of providing a level of private consumption to the child in period t that is a function of the amount of the time the child spends investing in themselves in that period. We assume that monitoring associated with implementing this incentive system carries a within-period utility cost that varies across households. The use of incentive schemes, or extrinsic motivation, allows even greater scope for parent and child interactions. We show that by using this incentive scheme an efficient intrahousehold outcome is attained, where efficiency is defined with respect to the parents' altruistic preferences. Due to the monitoring cost and the differences across households in the welfare gains associated with the use of such an incentive scheme, not all households choose to use them. We think of the choice of whether to use extrinsic rewards or not, loosely speaking, as a "parenting style." Similar to the recent models of Doepke and Zilibotti (2017) and Doepke et al. (2019), forward-looking and partially altruistic parents in our model can choose, at some cost, to influence their child's behavior during childhood, which impacts the child's human capital and welfare as an adult.

An additional crucial component of preferences, in particular with regard to decision-making in a dynamic environment, are the agents' discount factors. In the spirit of the Becker and Tomes (1979) model of inter-generational mobility, we consider "patience" to be an important component of family culture, which can be passed on from parents to their children, and which in turn affects both agents' propensity to invest in the child's human capital. We introduce novel features relating to agents' time preferences, allowing us to incorporate the following findings and stylized facts from the literature: (1) there is vast heterogeneity in discount factors across individuals, both among adults and children, as well as substantial correlation between parental and child discount factors within households; (2) the capacity of children and adolescents to delay gratification changes systematically with age; (3) parents can pass on certain preferences to their children both directly (e.g., through genetic endowments) and indirectly (e.g., through parenting styles); (4) extrinsic rewards offered to children, while productive in the short run, might reduce future intrinsic motivation or interest in the child's cognitive development; and (5) measures of forward-lookingness in adolescents are highly predictive of future human capital accumulation and economic success (e.g., high school graduation or disciplinary referrals).[5]

[5] See for example, the experimental evidence on heterogeneity in children's and adults' time preferences (Steinberg et al., 2009, and Falk et al., 2018), the evidence on intergenerational transmission of time preferences (Webley and Neyhus, 2006, Reynolds et al., 2009, Kosse and Pfeiffer, 2012, Arrondel, 2013, Brown

Our empirical results demonstrate that the use of an extrinsic reward mechanism such as the one we study does increase short-run time investments of the child, but at the cost of lowering their intrinsic motivation to invest in their own human capital in the future. In fact, we estimate that the main cost to the parents of using an incentive mechanism is not the short-run utility (monitoring) cost, but rather the degradation in the child's preference for future rewards, which results in lower future self-investment in their own human capital. We believe that this evidence is an important contribution to the extrinsic versus intrinsic motivation debate, and can help explain the wide dispersion in child skills and related outcomes observed by late adolescence.

We also use our estimated model to consider the role that parental characteristics play in accounting for intergenerational inequality. More highly educated parents clearly have higher wages and levels of non-labor income, and these provide their children with better life chances, but these factors are not the most important in accounting for differences in outcomes by parental education. We find that the impact of schooling on the productive value of a parent's investment is very important, as is the impact of parental characteristics on initial and final values of the child's discount factor. Thus this non-cognitive characteristic of parents and their children plays a large role in accounting for the persistence of cognitive outcomes across generations.

The outline of the paper is as follows. In Section 2, we describe the model and characterize the solution in the situations in which the parents offer incentives to their children for self-investment or do not. Section 3 contains a description of the data used to estimate the model and presents descriptive statistics. In Section 4, we discuss the estimation method that we employ and the identification of the primitive parameters describing the model. Section 5 presents the model estimates and discusses within-sample fit. In Section 6, we consider the role that incentives, or "parenting styles," play in the child development process using estimates from the model. We also examine the channels through which parental background impacts cognitive and non-cognitive development outcomes. We offer a brief conclusion in Section 7.

and van der Pol, 2015), the propensity to save (Knowles and Postlewaite, 2005), risk and giving (Cesarini et al., 2009), or risk and trust attitudes (Dohmen et al., 2012), the experimental evidence of the detrimental effects of extrinsic rewards on intrinsic motivation (see e.g. Deci, 1973, and Greene and Lepper, 1974, for early studies, and Deci, Koestner and Ryan, 1999, for an extensive meta-analysis), and the relation between children's time preferences, graduation rates and economic outcomes (Castillo et al. 2011, 2019).

2 Model

2.1 Model Primitives

The model consists of two (period/flow) utility functions, one for the parents (jointly) and one for the child.[6] The production technology for the cognitive ability of the child has a Cobb-Douglas form, as in DFW, with intertemporal linkages captured by the dependence of period $t+1$ cognitive ability on period t cognitive ability, in addition to inputs chosen by the parents and the child in period t. We allow for heterogeneity in (and correlation between) the intertemporal discount factors ("patience") of both agents. The parents' discount factor is stable (i.e., time-invariant) and possibly correlated with education levels, but the child's discount factor, which we interpret as an important measure of non-cognitive "ability," is allowed to evolve stochastically with age and to interact with parenting style choices of the mother and father. There is no lending or borrowing in the model, so that expenditures on investment goods for the child and household consumption in any period t are equal to the sum of all income sources for the household in period t, including the labor income of each parent and the non-labor income of the household. We now describe the components of the model more formally.

2.1.1 Environment

There are two sets of agents who inhabit the household, the parents and the child. Although we consider two-parent households and allow for different time allocation decisions for each parent, we assume the parents act as a single decision-maker when making choices. Both the parents and the child make decisions during each of the M (annual) child development periods, starting with the birth of the child at period $t = 1$ and ending at period $t = M$, when the phase of the child development process that we model concludes. We can think of period $M + 1$ as the beginning of the next phase of the young adult's cognitive development process, which starts with an initial cognitive ability level of k_{M+1} and an initial discount factor of $\beta_{c,M+1}$, which will jointly determine the child's welfare from that point on. We might imagine the next phase as being one in which the child is the sole decision-maker in their investment decisions, where the child's future choices and outcomes (e.g., educational attainment, income, savings and wealth) will be determined in part by the final outcomes of their childhood development process as summarized by the pair $(k_{M+1}, \beta_{c,M+1})$.[7]

[6] If the model considered only the actions of parents, it would appropriately be thought of as a *unitary* model of the household. However, since another decision-making agent occupies the household, the child, ours is not a unitary model.

[7] For further evidence on how discount factor heterogeneity can be an important driver of income and wealth inequality, see e.g. Krusell and Smith (1998), Hendricks (2007) and De Nardi and Fella (2017).

2.1.2 Instantaneous Utility Functions

Parents are allowed to be altruistic, although we preclude the child from acting altruistically toward the parents. All utility functions are assumed to be Cobb-Douglas, which greatly simplifies the solution of the model in the presence of so many choice variables. We assume the instantaneous utility functions of both parents and children to be homogeneous across households (as opposed to time preferences, which will be allowed to vary both across as well as within households).

The child's period t private utility is given by

$$\begin{aligned} u_{c,t} &= u_c(l_{c,t}, x_t, k_t) \\ &= \lambda_1 \ln l_{c,t} + \lambda_2 \ln x_t + \lambda_3 \ln k_t, \end{aligned}$$

where $l_{c,t}$ is the child's leisure (or play) time in period t, x_t is their consumption of a private good purchased in the market by the parents, and k_t is the child's cognitive ability at the beginning of period t. The preference weights are all strictly positive and are assumed to sum to one, $\sum_{i=1}^{3} \lambda_i = 1$, an inconsequential normalization given our assumptions.[8]

Parents are altruistic toward their child, but they also have a "private" utility function given by

$$\begin{aligned} u_{p,t} &= u_p(l_{1,t}, l_{2,t}, c_t, k_t) \\ &= \alpha_1 \ln l_{1,t} + \alpha_2 \ln l_{2,t} + \alpha_3 \ln c_t + \alpha_4 \ln k_t, \end{aligned}$$

where $l_{i,t}$ is the leisure of parent $i \in \{1, 2\}$ in period t, c_t is the level of consumption of a private good valued only by the parents, each α_j is strictly positive, and $\sum_{j=1}^{4} \alpha_i = 1$. We define this as the parents' private utility function because it does not include the altruism component. The *total* period t utility of the parents is given by

$$\tilde{u}_{p,t} = (1 - \varphi)u_{p,t} + \varphi u_{c,t},$$

where $\varphi \in [0, 1]$ indicates the extent of the parents' altruism, with $\varphi = 1$ indicating "pure" altruism on the part of the parents, and with $\varphi = 0$ indicating that they exhibit no altruistic behavior toward the child. We note that this specification resembles that of a Benthamite social welfare function for the household, with $(1-\varphi)$ being the weight given the parents and φ being the weight given to the child. In this sense, we can think of the parents as having preferences consistent with those of a social planner.

[8]Under the Cobb-Douglas assumptions only the ratio of preference parameters matter in determining choices, so that their scale is indeterminate. Since all parents and all children have the same preference parameters across households, any scale normalization applied to one household will be the same in any other. We make the usual assumption that the sum of Cobb-Douglas preference parameters for the parents and the sum of the preference parameters for the parents are both equal to 1.

Substituting the expression for the child's utility into the parents' utility yields the parents' period t utility function

$$\tilde{u}_{p,t} = \tilde{\alpha}_1 \ln l_{1,t} + \tilde{\alpha}_2 \ln l_{2,t} + \tilde{\alpha}_3 \ln c_t + \tilde{\alpha}_4 \ln k_t + \tilde{\alpha}_5 \ln l_{c,t} + \tilde{\alpha}_6 \ln x_t, \tag{1}$$

where

$$\begin{array}{lll} \tilde{\alpha}_1 = (1-\varphi)\alpha_1, & \tilde{\alpha}_2 = (1-\varphi)\alpha_2, & \tilde{\alpha}_3 = (1-\varphi)\alpha_3, \\ \tilde{\alpha}_4 = (1-\varphi)\alpha_4 + \varphi\lambda_3, & \tilde{\alpha}_5 = \varphi\lambda_1, & \tilde{\alpha}_6 = \varphi\lambda_2. \end{array}$$

This expression makes clear that, due to parental altruism, the child's consumption, leisure, and human capital are "public" goods that both sets of agents, the parents and the child, enjoy. In the case that $\varphi = 0$, the parent's total utility is simply equal to their private utility and the parents place no value on the leisure or private consumption of the child. For $\varphi = 1$, the parents' utility function is the same as the child's, which implies that all household choices would be made so as to maximize the child's welfare, narrowly defined. In such a case, both sets of players have the same objective, and the problem reduces to a single agent problem. However, this case produces several counterfactual implications, such as that the parents would consume no leisure and would set $c_t = 0$ in every period.

2.1.3 Terminal Period Utility

We assume the child's terminal payoff function in the final period of the development process $t = M + 1$ is given by the perpetuity value of their terminal flow utility, which is defined as their final stock of (log) cognitive skills, $\ln k_{M+1}$, weighted by the child's taste for human capital, λ_3, discounted by their final discount factor of this development stage $\beta_{c,M+1}$:

$$\begin{aligned} V_{c,M+1}(k_{M+1}, \beta_{c,M+1}) &= \lambda_3 \ln k_{M+1}(1 + \beta_{c,M+1} + \beta_{c,M+1}^2 + ...) \\ &= \frac{\lambda_3 \ln k_{M+1}}{1 - \beta_{c,M+1}} \end{aligned} \tag{2}$$

where k_{M+1} and $\beta_{c,M+1}$ are the final stocks of cognitive and non-cognitive skills, respectively.

For the parents, we assume that the terminal welfare function is given by the weighted sum of a "selfish" parental welfare component and a child welfare component, where the weight on the parents' welfare is defined as parents' altruism parameter, φ, and is discounted by each agent's respective discount factor:

$$\begin{aligned} V_{p,M+1}(k_{M+1}, \beta_{c,M+1}) &= (1-\varphi)\alpha_4 \ln k_{M+1}(1 + \beta_p + \beta_p^2 + ...) \\ &\quad + \varphi\lambda_3 \ln k_{M+1}(1 + \beta_{c,M+1} + \beta_{c,M+1}^2 + ...) \\ &= \Big(\frac{(1-\varphi)\alpha_4}{1-\beta_p} + \frac{\varphi\lambda_3}{1-\beta_{c,M+1}}\Big) \ln k_{M+1} \end{aligned} \tag{3}$$

Under this specification, which is consistent with how the parents' instantaneous utility was defined in equation (1), the parents directly value their child's final stock of cognitive human capital, as measured by k_{M+1}, as well as their final discount factor, $\beta_{c,M+1}$. Note that the term within the brackets, which denotes the parents' marginal value of cognitive skills, is strictly increasing in $\beta_{c,M+1}$ whenever $\varphi > 0$. This can be interpreted as a *complementarity* between cognitive and non-cognitive skills, which has often been documented in the literature.[9] From the perspective of the model, the parents internalize the fact that after their child becomes independent, their final "stock" of patience will co-determine many of their future choices, thus acting as a complement to the child's stock of cognitive skills.[10]

2.1.4 Technology of Cognitive Skill Formation

Following DFW,[11] the production technology of cognitive skills, or "child ability", is given by the law of motion:

$$\ln k_{t+1} = \ln R_t + \delta_{1,t} \ln \tau_{1,t} + \delta_{2,t} \ln \tau_{2,t} + \delta_{3,t} \ln \tau_{c,t} + \delta_{4,t} \ln e_t + \delta_{5,t} \ln k_t, \tag{4}$$

where $\tau_{i,t}$ is the amount of time parent $i \in \{1, 2\}$ spends in child investment, $\tau_{c,t}$ is the time the child spends in self-investment (i.e., without either of the parents actively present), and e_t is the amount of child investment goods purchased in period t. Note that the production function parameters are indexed by the child's age, t. Descriptive evidence from the CDS leads us to believe that the value of the child's time in self-investment is increasing in the age of the child, while the opposite may be true for the value of parental time investments, which is what was found in DFW. We also allow the productivity of the parental time inputs to depend on household characteristics, such as parental educational attainment, but for notational simplicity we have not explicitly noted this dependence. We interpret R_t as total factor productivity (TFP). We provide more details on the econometric specification in Section D.1.2.

2.1.5 Time and Budget Constraints

Each parent has a weekly amount of time T to allocate to leisure, investment in the child, and market work. The time spent in the labor market by parent $i = 1, 2$ in period t is given by $h_{i,t}$, where we allow either or both parents to not work at all. The wage offer available

[9]See e.g., Cunha, Heckman and Schennach (2010).

[10]For the extreme case in which $\varphi = 1$, the parents' terminal value function is identical to the child's terminal value function as defined in (2). However, when altruism is imperfect (i.e., $0 < \varphi < 1$), the parents' welfare function generically differs from the child's, except in the case in which preferences are perfectly aligned, i.e. $\alpha_4 = \lambda_3$ and $\beta_p = \beta_{c,M+1}$.

[11]DFW used a similar Cobb-Douglas growth specification, although the arguments differed. In particular, they further disaggregated parental time investments into passive and active time components, and they did not include child self-investment time.

to parent i in period t is given by $w_{i,t}$, and the household's non-labor income in period t is given by I_t.[12] Then the total income of the household in period t is

$$Y_t = w_{1,t}h_{1,t} + w_{2,t}h_{2,t} + I_t.$$

There are no capital markets available for transferring consumption between periods, so that the household spends all of its period t income on productive investment goods for the child, e_t, the private consumption of the child, x_t, and the private consumption of the parents, c_t:[13]

$$Y_t = e_t + x_t + c_t,$$

where we have assumed that the prices of all investment and consumption goods are equal to 1.

Each parent has a total amount of time T to be allocated, where time is denoted in terms of weekly hours. We assume that $T = 112$, so that 16 hours of time per day are assumed to be available for each parent. The time resource constraint for parent i is

$$T = l_{i,t} + h_{i,t} + \tau_{i,t}, \; i = 1, 2.$$

The child supplies no labor to the market, but has to spend the times chosen by the parents with them. In addition, the child spends time at school and being transported to school. This amount of time varies by age and is denoted s_t.[14] The child's time constraint is given by

$$T = l_{c,t} + \tau_{p,t} + \tau_{c,t} + s_t,$$

where $\tau_{p,t} = \tau_{1,t} + \tau_{2,t}$, is the total time spent with the parents in investment activities. In what follows we will define $\tilde{T}_t \equiv T - s_t$, which is the child's discretionary time outside of formal schooling. The child's only choice variable is their self-investment time, $\tau_{c,t}$.

2.1.6 Parental Incentives

In any given period, apart from choosing how to allocate their time and money, the parents also select what we refer to as a *parenting style*. This binary choice is between:

1. Providing an incentive scheme that maps the amount of time that the child spends in

[12]In the empirical estimation, we assume that wages and non-labor income are stochastic and are functions of parental observable characteristics (i.e., age and educational attainment), and that households have perfect foresight with respect to future shocks. See Section 4 for more details.

[13]The private consumption goods of the parents may include goods that are best thought of as public, in the sense that the child also profits from them. This includes housing, heat, transportation services, etc. To simplify the exposition of the model, we have assumed that the child does not perceive these expenditures as contributing to their welfare. They then are treated as private consumption expenditures of the parents.

[14]We assume that this time is set exogenously. Any gains to child ability generated by the time spent in school are assumed to be captured by the TFP measures $R_1, ..., R_M$.

self-investment in period t into the child's consumption of their private good in that period. We will refer to this as a Conditional Cash Transfer, or CCT. We will assume that the incentive function is given by

$$\ln x_t(\tau_{c,t}; r_t, b_t) = b_t + r_t \ln \tau_{c,t}, \tag{5}$$

where (b_t, r_t) are optimally chosen by the parents. Under our model specification, this particular choice of incentive function is without loss of generality.[15]

2. Providing no consumption incentives to the child in order to affect their self-investment time. In terms of Equation (5), we can think of the parents setting $r_t = 0$ in this case.

If parents choose to reward the child for their self-investment time in period t, we set $CCT_t = 1$ $(\Leftrightarrow r_t \neq 0)$, if they do not then $CCT_t = 0$ $(\Leftrightarrow r_t = 0)$. The parents need not choose between incentivizing their child in every period versus never incentivizing them. It is possible for parents to vary their use of monetary incentives over the development period. Thus the sequence $(CCT_1, ..., CCT_M)$ is unrestricted in principle.

As a parenting style, when parents choose $CCT_t = 1$ they partially rely on *extrinsic factors* to motivate the child in period t. Of course, the exact form of consumption schedule offered to the child depends on the child's own characteristics, including their preferences and their discount factor, as well as the parents' characteristics. Absent the explicit connection between study time and consumption in a period, we will characterize the child's motivation as being *intrinsic* in nature.

The use of a CCT allows parents to achieve their first-best in terms of welfare, so unless there are costs to using a CCT, all parents should use one in every development period. We allow two potential costs to using a CCT. The first is a current period utility loss, which can include psychic costs associated with setting up and negotiating an implicit or explicit contract with the child and the cost of having to verify the child's self-investment time. We assume that the utility cost of using a CCT in period t for a household is given by $\zeta \geq 0$,

[15] The parents' first-best outcome is produced when the parents are allowed to make all of the choices in period t, including the choice of the time the child spends in self-investment. Denote the child's consumption and study time in this case by $(\hat{x}_t, \hat{\tau}_{c,t}) \in \mathbb{R}^2_+$. As we show in Section 2.2.2 below, the child's decision regarding their study time in period t under this rule is a strictly increasing function of r_t and is independent of b_t. Given the parents' first-best choices of time investment with the child, the sum of which is $\hat{\tau}_{p,t} = \hat{\tau}_{1,t} + \hat{\tau}_{2,t}$, the discretionary time left to the child is $\tilde{T}_t - \hat{\tau}_{p,t}$. Then $\lim_{r_t \to \infty} \tau_{c,t}(r_t) = \tilde{T}_t - \hat{\tau}_{p,t}$, and $\exists r^{min} < 0$ such that $\lim_{r_t \to r^{min}} \tau_{c,t}(r_t) = 0$. Then we can find a unique $\hat{r}_t \geq r^{min}$ and $\hat{b}_t \in \mathbb{R}$ that solve

$$\hat{\tau}_{c,t} = \tau_{c,t}(\hat{r}_t)$$

and

$$\hat{x}_t = \hat{b}_t + \hat{r}_t \ln \hat{\tau}_{c,t}$$
$$\Rightarrow \hat{b}_t = \hat{x}_t - \hat{r}_t \ln \hat{\tau}_{c,t}$$

Thus the parents can implement their first-best choices using this rule.

where ζ is a random variable that varies across households.[16] We can rewrite the parents' "total" instantaneous utility function as a function of their period t choice vector $\mathbf{a}_{p,t}$:

$$\tilde{u}_{p,t}(\mathbf{a}_{p,t}) = \tilde{\alpha}_1 \ln l_{1,t} + \tilde{\alpha}_2 \ln l_{2,t} + \tilde{\alpha}_3 \ln c_t + \tilde{\alpha}_4 \ln k_t + \tilde{\alpha}_5 \ln l_{c,t} + \tilde{\alpha}_6 \ln x_t - \zeta \cdot \mathbb{1}[CCT_t = 1] \quad (6)$$

2.1.7 Time Preferences

Since we consider decision-making in a dynamic environment, another key component of preferences pertains to the agents' discount factors. In the spirit of the intergenerational mobility model of Becker and Tomes (1979), we consider "patience" to be a crucial component of family culture, which is transferable from parents to their children, and which in turn drives their propensity to invest in the child's human capital. In our model, we allow both parents' and children's intertemporal discount factors to be heterogeneous across households, as well as potentially correlated within households. Moreover, in the case of the child, we allow time preferences to be time-varying and possibly affected by parental choices, such as the decision of whether to use an extrinsic motivation scheme (i.e., a CCT).

The parents' discount factor is denoted by $\beta_p \in (0, 1)$, and is assumed to be time-invariant and potentially correlated with observable parental characteristics. In particular, we assume that for every household $h = 1, ..., N$, $\beta_{p,h}$ is randomly drawn from a discrete distribution with Z points of support, denoted by $\{\beta_p^j\}_{j=1,...,Z}$, and a conditional probability distribution that depends on the father's educational attainment level, $s_{h,2}$:

$$\beta_{p,h} = \beta_p^j \quad \text{w.p. } p^j(s_{h,2}) \quad \forall j = 1, ..., Z$$

where $\sum_{j=1}^{Z} p^j(s_{i,2}) = 1$ for all schooling levels $s_{i,2}$. The support set $\{\beta_p^1, \beta_p^2, ..., \beta_p^Z\}$ and conditional probabilities $\{p^j(s)\}_{j=1,...,Z}$ are fixed exogenously.[17]

For children, we denote the age t discount factor by $\beta_{c,t} \in (0, 1)$. As was the case for parents, we assume $\beta_{c,t,h}$ is a discrete random variable with Z points of support:

$$\beta_{c,t,h} \in \{\beta_c^1, \beta_c^2, ..., \beta_c^Z\}$$

for all households $h = 1, ..., N$ and all ages $t = 1, ..., M$.[18] Differently from parents, we allow the child's discount factor to time-vary in a potentially endogenous manner. In particular, we assume the transitions between $\beta_{c,t}$ and $\beta_{c,t+1}$ are governed by a first-order Markov process

[16] It is straightforward to allow this utility cost to also vary over time for each household, but this slight generalization entails increased computational costs.

[17] In the empirical application, we determine the support set and conditional distributions using data on the annual discount factors of a representative sample of U.S. adults between ages 25 and 45 in the 2010 wave of the Osaka Preference Parameters Study. See Section D.1.1 and Appendix C.3 for more information.

[18] In the estimation of the model we use the experimental survey data from Steinberg et al. (2009) in order to select the support set $\{\beta_c^j\}_{j=1,...,Z}$ and to compute empirical moments related to the conditional distribution of children's and adolescents' annual discount factors.

that depends on the child's current age and the current value of their discount factor, as well as on the parents' binary CCT choice in period t:

$$\Pr(\beta_{c,t+1,h} = \beta_c^{j'} | \beta_{c,t,h} = \beta_c^{j}, t, CCT_{t,h}) \quad \forall (j, j') = 1, ..., Z, \tag{7}$$

for all households $h = 1, ..., N$ and for all development periods $t = 1, ..., M$. We assume the child's initial discount factor is drawn randomly from an exogenous distribution that is allowed to depend on the parents' educational attainment level. More details on the empirical implementation and econometric specification are provided in Section D.1.1.

The introduction of an endogenous discount factor adds a non-trivial and novel element to the model by allowing us to capture richer interactions between the child's preferences and actions on the one hand, and the parents' own preferences, investments and parenting choices on the other. As we demonstrate in the empirical work reported below, changes in how forward-looking the child is as well as increases in the productivity of self-investment as the child ages are important factors in accounting for increases in self-investment during adolescence. Our econometric specification puts no restriction on the relationship between CCT use and the likelihood of increases or decreases in the child's discount factor. However, we estimate that CCT use in a period does (stochastically) decrease the discount factor in the following period. This constitutes another important cost to the parental use of CCTs. In Section 6.1 we demonstrate that this cost is the most consequential in explaining the limited use of CCTs by parents.

We note that our assumption regarding the child's discount factor does not imply any type of time-inconsistent behavior, such as is generated in models with hyperbolic discounting (e.g., Fang and Silverman (2009)). Given our assumption of rational decision-making on the part of both parents and children, a young child of age t makes decisions using the age t discount factor $\beta_{c,t}$, however their decisions reflect their knowledge that the future values of discount factors they use are (in a stochastic sense) increasing with age, and may depend on their parents' future choices of using CCTs. Although this assumption imposes a large degree of rationality on young children, its practical significance may not be great if the current discount factor is low and larger discount factors only emerge gradually. By making this assumption we avoid problems of time inconsistency, which can be particularly problematic when there are strategic interactions between agents.

2.2 Parent-Child Interactions

We assume that the actions taken by the parents and children are generated within a *Markov Perfect Equilibrium.* The sequential structure within a period has a Stackelberg form in which parents are the first-movers. The parents make all choices in the household in period t with the exception of how much time the child spends in self-investment, $\tau_{c,t}$. It is assumed that the child must abide by the choices of the parents in terms of their own time allocation.[19] That is, the parents choose $\tau_{1,t}$ and $\tau_{2,t}$, and the child must spend that amount of time with them. We assume that the only household members receiving earnings or non-labor income are the parents, so that it is natural to assume that they make all decisions in terms of household money expenditures.

In all periods the child is the Stackelberg follower who makes their choice of $\tau_{c,t}$ after the parents have announced all other household choices (including whether they will use a CCT in period t). The child's decision rule (reaction function) will therefore have the form $\tau^*_{c,t}(\mathbf{a}_{p,t})$, where $\mathbf{a}_{p,t}$ are all parental choices in period t. Given the function $\tau^*_{c,t}$ the parents choose their actions $\mathbf{a}_{p,t}$, including their labor supplies in the period, $h_{1,t}$ and $h_{2,t}$; their allocation of household income across their own private consumption, c_t, the private consumption of the child, x_t, and expenditures on child investment goods, e_t; their time investments with the child, $\tau_{1,t}$ and $\tau_{2,t}$; and finally their incentives offered to the child through the binary choice $CCT_t \in \{0, 1\}$, and the reward function parameters b_t and r_t (with $r_t = 0$ if $CCT = 0$, and x_t determined by equation (5) if $CCT_t = 1$).[20] Both agents are forward-looking and fully account for the impacts of current actions on future state valuations. As discussed above, children make decisions fully understanding that their valuation of future events (i.e., their discount factor) will be evolving as they age.

At the beginning of period t the household's vector of state variables is given by

$$\Gamma_t = (w_{1,t}, w_{2,t}, I_t, k_t, \beta_{c,t}),$$

which includes each parent's hourly wage offer, the household's non-labor income, and the child's cognitive skill and patience level (a non-cognitive characteristic). In order to simplify the problem, the wage and non-labor income processes are assumed to be exogenous with respect to household actions, although the wage draw of parent i in period t will only be observed if parent i supplies a positive amount of time to the labor market in period t. Thus the only endogenous dynamic processes are for the child's cognitive and non-cognitive "ability," as reflected in k and β_c, respectively.

The value function for the child in periods $t = 1, \ldots, M$ is given by

$$V_{c,t}(\Gamma_t|\mathbf{a}_{p,t}) = \max_{\tau_{c,t}|\mathbf{a}_{p,t},\mathbb{C}_{c,t}} u_c(l_{c,t}, x_t, k_t) + \beta_{c,t}\mathbb{E}_t V_{c,t+1}(\Gamma_{t+1}|\tau_{c,t}, \mathbf{a}_{p,t}, \Gamma_t),$$

where $\mathbf{a}_{p,t}$ denotes the actions of the parents in period t, which occur prior to the child's

[19] Any parent of adolescents could easily take issue with this assumption, but practically speaking, the assumption will not be that objectionable. This is due to the fact that in the data, time with parents decreases dramatically over the development period, beginning when the child enters formal schooling.

[20] If the parents are not altruistic, i.e., $\varphi = 0$, the parents' contribution to the child's private consumption is not well-defined. That is one technical motivation for at least allowing a minimal level of altruism on the part of the parents, although we believe most parents exhibit considerable altruism with respect to their children.

selection of the value $\tau_{c,t}$, and $\mathbb{C}_{c,t}$ denotes the choice set of the child in period t, which was defined above. The parents' problem is similarly structured. Being the leaders in terms of action choices in period $t = 1, \ldots, M$, the parents' problem is

$$V_{p,t}(\Gamma_t) = \max_{\mathbf{a}_{p,t}|\tau^*_{c,t}(\mathbf{a}_{p,t}),\mathbb{C}_{p,t}} \tilde{u}_p(\mathbf{a}_{p,t}) + \beta_p \mathbb{E}_t V_{p,t+1}(\Gamma_{t+1}|\mathbf{a}_{p,t}, \Gamma_t),$$

where the parents' actions are chosen given the child's period t reaction function $\tau^*_{c,t}(\mathbf{a}_{p,t})$ and their choice set $\mathbb{C}_{p,t}$. The functional forms we have specified produce solutions that have attractive properties from the point of view of solving for the decision rules of the agents and for the estimation of model parameters.[21] We provide details regarding the solution of the model in Appendix B. Here we briefly focus on some key aspects of the interaction between the parents and the child. To better understand the impact of the parents' incentive (CCT) choices on the child's actions and the agents' resulting welfare levels, we consider the two separate cases where parents either do or do not use a CCT.

2.2.1 Case 1: Parents do not use a CCT

When parents do not use a CCT in period t then they are constrained to set the reward elasticity $r_t = 0$. We denote the resulting choices with the superscript 0, and let $a^0_{p,t}$ represent all (optimal) choices made by the parents in period t under the restriction on r_t. In this case the child's reaction function is

$$\tau^0_{c,t}(a^0_{p,t}; \Gamma_t) = \gamma^0_t(\Gamma_t)(\tilde{T}_t - \tau^0_{p,t}) \tag{8}$$

where $\tilde{T}_t = T - s_t$ is the time available to the child outside of formal schooling in period t, s_t, and $\tau^0_{p,t}$ is total time spent with the parents in period t in the absence of a CCT, with $\tau_{p,t} \equiv \tau_{1,t} + \tau_{2,t}$. The proportion of uncommitted time in period t that is devoted to self-investment is given by

$$\gamma^0_t(\Gamma_t) = \frac{\Delta^0_{c,t}(\Gamma_t)}{\lambda_1 + \Delta^0_{c,t}(\Gamma_t)} \in (0, 1),$$

where $\Delta^0_{c,t}(\cdot)$ denotes the marginal "return" to self-investment from the perspective of the child in period t:

$$\Delta^0_{c,t}(\Gamma_t) = \beta_{c,t}\delta_{3,t}\psi^0_{c,t+1}(\Gamma_t),$$

This return is the product of the child's discount factor, $\beta_{c,t}$, the marginal productivity of the child's self-investment time, $\delta_{3,t}$, and the child's future marginal value of their own human

[21] However, because of the strategic aspects of the game played between parents and the child, the solutions to the parents' problem no longer have a closed form as they did in DFW.

capital in the absence of a CCT, defined as

$$\psi^0_{c,t+1}(\Gamma_t) = \frac{\partial \mathbb{E}_t\big[V_{c,t+1}(\Gamma_{t+1})|CCT_t = 0\big]}{\partial \ln k_{t+1}}. \tag{9}$$

Thus, $\psi^0_{c,t+1}(\cdot)$ represents future expected utility flows from the child's cognitive human capital as perceived by the child, which depend on the child's preference for own human capital, λ_3, as well as future discount factors and technology parameters. Appendix B provides the explicit solutions for $\psi^0_{c,t+1}(\Gamma_t)$.[22] The child's reaction function (8) can be used to determine the rate at which parental time "crowds-out" child self-investment time, with

$$\frac{\partial \tau^0_{c,t}}{\partial \tau_{p,t}} = -\gamma^0_t(\Gamma_t).$$

This expression indicates that in the absence of any extrinsic incentives to self-invest, every hour of parental time "crowds out" the child's self-investment time by $\gamma^0_t \in (0, 1)$ hours. Impatient or myopic children (low $\beta_{c,t}$) have a low γ^0_t and invest little of their own time in skill development. In this case there is a low degree of crowd-out. As children become more patient and as they become more productive in increasing their own cognitive ability (i.e., $\beta_{c,t}$ and/or $\delta_{3,t}$ increase), children willingly spend more time in self-investment for any given amount of parental time and the degree of crowd-out by parents' time investments increases.[23] Of course, the parents' choice of whether or not to use a CCT to further increase the child's self-investment time is in itself endogenous, and will depend on the relative preferences (i.e., instantaneous utilities and discount factors) and characteristics of the parents and the child. We analyze this choice below in Section 2.2.3.

2.2.2 Case 2: Parents use a CCT

When the parents choose to use a CCT in period t, they select both b_t and r_t optimally in addition to all of their other choices, with a choice vector given by $a^1_{p,t}$. In the presence of a CCT, the child's reaction function at time t is given by

$$\tau^1_{c,t}(a^1_{p,t}; \Gamma_t) \;\; = \;\; \gamma^1_t(r_t, \Gamma_t)(\tilde{T}_t - \tau^1_{p,t}) \tag{10}$$

[22] Given the recursive model structure and the stochastic Markov process governing the evolution of the child's discount factor over time, we can derive $\psi_{c,t+1}(\cdot)$ in closed form only for period $t = M$. For the remaining periods $t = 1, ..., M-1$, we approximate $\psi_{c,t+1}(\cdot)$ analytically based on a response surface approach where, conditional on a sequence of exogenous wage and non-labor income draws, we (closely) approximate the child's indirect value function $V_{c,t}(\cdot)$ through a polynomial in the endogenous state variables, $\ln k_t$ and $\beta_{c,t}$.

[23] The choices made by the parents and child in each period are unique and well-defined under our functional form assumptions. The reaction function of the child in terms of their self-investment time is a linear function of the amount of time that the parents choose to spend with her. Substituting this reaction function into the parents' choice problem makes this a single-agent optimization problem that produces unique solutions for the parents' choices in the period, $\mathbf{a}_{p,t}$.

The proportion of uncommitted time in period t that is devoted to self-investment (as opposed to leisure) is now given by

$$\gamma_t^1(r_t;\Gamma_t) = \frac{\lambda_2 r_t + \Delta_{c,t}^1(\Gamma_t)}{\lambda_1 + \lambda_2 r_t + \Delta_{c,t}^1(\Gamma_t)} \in (0,1),$$

where $\Delta_{c,t}^1(\Gamma_t)$ denotes the child's marginal return to self-investment conditional on the parents using a CCT:

$$\Delta_{c,t}^1(\Gamma_t) = \beta_{c,t}\delta_{3,t}\psi_{c,t+1}^1(\Gamma_t),$$

Similar to the no-CCT case, the scalar $\psi_{c,t+1}^1(\Gamma_t)$ denotes the child's expected marginal utility from having more cognitive human capital in the future if the parents use a CCT today:

$$\psi_{c,t+1}^1(\Gamma_t) = \frac{\partial \mathbb{E}_t\big[V_{c,t+1}(\Gamma_{t+1})|CCT_t = 1\big]}{\partial \ln k_{t+1}}, \tag{11}$$

The effects of the parents using a CCT on the child's reaction function is twofold. First of all, under the Markov process governing the child's discount factor given by (7), the parents' use of a CCT at time t may affect the child's future discount factor, which enters into their future marginal utility from human capital, $\psi_{c,t+1}^j(\cdot)$ for $j \in \{0,1\}$. If the use of an extrinsic reward scheme by the parents makes the child more myopic in the future in a stochastic sense (i.e., if $\psi_{c,t+1}^1 < \psi_{c,t+1}^0$), then their marginal return to self-investment today will be lower if the parents use a CCT, i.e. $\Delta_{c,t}^1(\Gamma_t) < \Delta_{c,t}^0(\Gamma_t)$. In other words, while using a CCT can induce the child to self-invest more today, it may discourage the child from self-investing in the future by reinforcing myopic behavior.

The parents freely choose the reward elasticity, r_t, which directly enters into the child's reaction function. It is easy to see that $\gamma_t^1(\cdot)$ is monotonically increasing in r_t, which implies that the parents can implement any desired level of child self-investment time by choosing r_t arbitrarily high or low (or even negatively) as we have shown in footnote 15.[24] Therefore, the additional degree of freedom allows the parents, as the Stackelberg leader, to choose r_t such that the child's resulting best-response corresponds to the parents' desired level of self-investment time. As we show in Appendix B, conditional on the child's current discount factor, we can solve for the optimal reward elasticity in closed form:

$$\begin{aligned} r_t^* &= \frac{\Delta_{p,t}^1(\Gamma_t) - \varphi\Delta_{c,t}^1(\Gamma_t)}{\varphi\lambda_2} \\ \text{where} \quad \Delta_{p,t}^1(\Gamma_t) &= \beta_p\delta_{3,t}\psi_{p,t+1}^1(\Gamma_t) \\ \text{and } \psi_{p,t+1}^1(\Gamma_t) &= \frac{\partial \mathbb{E}_t\big[V_{p,t+1}(\Gamma_{t+1})|CCT_t = 1\big]}{\partial \ln k_{t+1}}, \end{aligned}$$

[24]Indeed, the expression for $\gamma_t^1(r_t;\Gamma_t)$ is monotonically increasing in r_t, ranging between 0 (as $r_t \to r^{min} \equiv \frac{-\Delta_{c,t}^1(\Gamma_t)}{\lambda_2}$) and 1 (as $r_t \to +\infty$).

where $\psi^1_{p,t+1}(\cdot)$ represents future expected utility flows from the child's cognitive human capital as perceived by the parents. Inserting $r_t = r_t^*$ into our previous expression for $\gamma_t^1(r_t; \Gamma_t)$ yields the following expression for the parents' "first-best" solution implemented through their use of a CCT:

$$\gamma_t^1(r_t^*; \Gamma_t) \quad = \quad \frac{\Delta^1_{p,t}(\Gamma_t)}{\tilde{\alpha}_5 + \Delta^1_{p,t}(\Gamma_t)}.$$

This ratio, which denotes the proportion of the child's uncommitted time that is spent on self-investment, looks analogous to the one derived earlier in Section 2.2.1, except the numerator now contains the parents' (as opposed to the child's) marginal valuation of future child skills, and the denominator contains the parents' (as opposed to the child's) instantaneous preference for child leisure.

2.2.3 Optimal CCT Choice

Under the assumption that using a CCT involves an instantaneous utility cost $\zeta \geq 0$ and may (stochastically) affect the child's future discount factor, parents will only choose to implement a CCT if doing so maximizes their expected welfare from period t onward:

$$CCT_t(\Gamma_t) = \arg \max_{CCT \in \{0,1\}} \{V_{p,t}(\Gamma_t | a^0_{p,t}(\Gamma_t)), V_{p,t}(\Gamma_t | a^1_{p,t}(\Gamma_t))\}$$

where the second component includes the fixed utility cost, ζ. In the empirical analysis, we evaluate these value functions numerically in order to obtain the optimal CCT choice for every household-period and every state vector Γ_t. Despite the lack of closed form solutions we can still obtain some intuition about the parents' CCT choices by comparing the child's reaction functions for the cases with and without a CCT. All else equal, parents are more likely to use a CCT whenever the child's self-investment time in the absence of a CCT (i.e., $\tau^0_{c,t}$) is sufficiently different from the level that the parents would implement in the incentivized case (i.e., $\tau^1_{c,t}$). Relative to the no-CCT case, the CCT will induce the child to devote a larger fraction of their residual time to self-investment whenever the following condition holds (suppressing some inessential notation):

$$\gamma_t^1 > \gamma_t^0 \quad \iff \quad \frac{\Delta^1_{p,t}}{\varphi\lambda_1 + \Delta^1_{p,t}} > \frac{\Delta^0_{c,t}}{\lambda_1 + \Delta^0_{c,t}} \iff \Delta^1_{p,t} > \varphi\Delta^0_{c,t} \iff \beta_p\psi^1_{p,t+1} > \varphi\beta_{c,t}\psi^0_{c,t+1}$$

In the final period, $t = M$, our assumptions regarding both agents' terminal period welfare functions let us further simplify this expression by plugging in analytical expressions for $\psi^1_{p,M+1}(\cdot)$ and $\psi^0_{c,M+1}(\cdot)$. Then, $\gamma^1_M > \gamma^0_M$ if and only if:

$$\frac{\beta_p}{1-\beta_p}\frac{1-\varphi}{\varphi}\frac{\alpha_4}{\lambda_3} > \beta_{c,M}\mathbb{E}_M\big(\frac{1}{1-\beta_{c,M+1}}|CCT_M = 0\big) - \beta_p\mathbb{E}_M\big(\frac{1}{1-\beta_{c,M+1}}|CCT_M = 1\big)$$

All else equal, the last inequality becomes easier to satisfy if (*i*) parents are more forward-looking relative to the child, (*ii*) parents are less altruistic towards the child, (*iii*) parents care more about the child's cognitive skills relative to the child herself, and (*iv*) the use of a CCT has a less detrimental (or more positive) impact on the child's future discount factor process. In other words, CCTs are more effective at altering the child's self-investment time, and therefore more desirable from the parents' perspective, when there is a larger misalignment between the parents' and the child's preferences, and whenever the CCT does not crowd out the child's future patience (and hence future self-investments) too much by reinforcing myopic behavior today.

3 Data and Descriptive Statistics

We utilize data from the Panel Study of Income Dynamics (PSID) and the first three waves of the Child Development Supplements (CDS-I, CDS-II and CDS-III). The PSID is a longitudinal study that began in 1968 with a nationally representative sample of about 5,000 American families, with an oversample of black and low-income families. In 1997, the PSID began collecting data on a random sample of the PSID families that had children under the age of 13 (CDS-I). Data were collected for up to two children in this age range per family. While the PSID provides us with detailed information on household labor supply, wages, non-labor income and household demographics, the CDS collects additional information on child development and family dynamics, including various indicators of children's cognitive achievements, socio-emotional development and health; time use of both parents and children; home environment and parent-child relationships (e.g., information about the use of unconditional or conditional allowances, i.e., incentives). The entire CDS sample size in 1997 is approximately 3,500 children residing in 2,400 households. A follow-up study with these children and families was conducted in 2002-03 (CDS-II), and a third wave was released in 2007 (CDS-III). These children were between the ages of 8-18 in 2003. No new children were added to the study after the initial wave (Hofferth et al. (1998)).

Unfortunately, the PSID-CDS lacks quantifiable information on parental or child discount factors, both of which are important for our empirical analysis. To that purpose, we use two additional data sources to help us identify and estimate the empirical distributions of parental and child discount factors, respectively. For children, we use a unique experimental survey data set that includes detailed quantitative measures of patience and forward-lookingness for a sample of over 900 individuals between the ages of 10 and 30, with over 500 adolescents between the ages of 10 and 17. We are grateful to Laurence Steinberg from Temple University for providing us with access to this data. We discuss these data in more detail in Section 3.2 and Appendix C.2.

To calibrate the distribution of discount factors for parents in our model, we use survey data from the 2010 U.S. wave of the Osaka Preference Parameters Study, which collects detailed individual measures of time preferences for over 7000 U.S. adults, as well as other covariates such as educational attainment. We focus on the subsample of $N = 4625$ adults between the ages of 25 and 65, which corresponds to the age range of the parents we observe in the PSID-CDS. We refer to Section 3.3 and Appendix C.3 for more details and summary statistics of these data.

3.1 PSID-CDS Household Panel

All of our results are based on a selected sample of PSID households that satisfy several criteria. First, included households are intact over the observed period (i.e. only stable two-parent households). Second, households have either one or two children. We select only one child from each household (see below). Third, all children are biological; there are no adopted children and no step-parents. Fourth, all selected children are at least three years old during the first wave of the CDS (CDS-I in 1997), because we need access to an initial Letter Word (LW) score observation. Fifth, all selected children have an observed LW score in 1997 and in 2002. Some of these also have an observed LW score in 2007 (CDS-III) as well, although it is not required. Sixth, if a household has two eligible siblings satisfying the previous two requirements, we select the youngest. This has two potential advantages: parental labor supply is probably more responsive to the age of the youngest sibling than the age of the oldest sibling, and we also have a higher chance of observing the youngest sibling in CDS-III, which enriches the total sample. Finally, we keep only data observations for selected children whose age is between 0 and 16 at the interview date for waves 1 through 3.

Overall, these selection criteria yield a panel of $N = 247$ households (or children). Appendix C.1 contains more detailed information about the data variables and their construction. Tables 1, 2, 3 and 4 contain some descriptive statistics of our sample of households. Next, we discuss the various data components in more detail.

3.1.1 Child's Time Allocation

Starting in 1997, children's time diaries were collected along with detailed assessments of children's cognitive development. For two days within a week (one weekday and either Saturday or Sunday), children (with the assistance of the primary caregiver when the children were very young) filled out a detailed 24 hour time diary in which they recorded all activities during the day and who else (if anyone) participated with the child in these activities. At any point in time, the survey recorded the intensity of participation of the parents: mothers and fathers could be actively participating or engaging with the child, or simply around the child but not actively involved. In this paper, we partition productive child time into three categories of inputs: (1) active time with the mother, τ_1, (2) active time with the father, τ_2, and (3) the child's productive self-investment time, τ_c. Any time where both parents are actively participating with the child is divided proportionally between the two parents according to the relative ratio of their individual active time investments. Thus, active time with parent i ($i \in \{1, 2\}$) is defined as the sum of two components: (1) time where parent i is actively participating with the child while the other parent is either not around or only passively participating (which can be denoted τ_i^{alone}), and (2) a fraction $\frac{\tau_i^{alone}}{\tau_1^{alone}+\tau_2^{alone}}$ of the time where both parents are actively participating with the child. As such, the total parental time in investment activities ($\tau_p = \tau_1 + \tau_2$) uniquely comprises all activities where at least one parent is actively involved.[25] Finally, the child's self-investment time is defined as any "productive" activity where neither parent is actively involved, such as doing homework, studying, reading, solving puzzles or playing educational games. One or both of the parents could be present in the home during this time (e.g. passively monitoring the child), but neither were actively interacting with the child.[26] For each type of child investment time, we construct a weekly measure by multiplying the daily hours by 5 for the weekday report and 2 for the weekend day report (using a Saturday and Sunday report adjustment) and summing the total hours for each category of time.

It is likely that much of the child self-investment time we observe is due to homework assignments. Although the time spent in homework assignments is partially under the control of teachers or others outside of the household, the survey data created by and analyzed in Cooper et al. (1998) make clear that there is considerable variability in the amount of homework assigned by teachers in the same grade level, and that all homework assigned is not completed. Their survey instrument included questions only related to the completion of homework, and therefore does not measure any other type of "self-investment", such as additional reading a child might do outside of homework assignments. For these reasons we believe that our measure of self-investment time is considerably broader than the homework time measures used in previous studies, such as those surveyed in Cooper et al. (2006).[27] Although teachers' homework assignments are an important factor, we argue that the level of self-investment observed in a household is largely determined by the actions of the parents and the child.

Table 3 shows how these three types of time investment evolve as children age. Whereas we see a clear decrease in both mothers' and fathers' active time with their children as they

[25] This specification avoids double-counting joint parental time, which otherwise might cause violations of the child's time constraint.

[26] Analogously, active time where parents assist the child with homework or reading would be counted as parental time, not as self-investment.

[27] The data produced by the National Educational Longitudinal Study (NELS) and other secondary databases that are most frequently used in assessing the relationship between study time and outcomes specifically ask for time spent in completing homework. Thus broader measures of child self-investment time cannot be constructed using these data sources.

age, we also notice a clear increase in their average self-investment time. Taken together, the evolution of these three time investment choices highlights the intuitive notion that children become more independent as they age.

In the bottom row of Table 3, we show how the child's school time (s_t) evolves with child age. Although we observe school time in the CDS data, we believe these data to be relatively noisy. Table 15 shows the detailed distribution of reported school time for every child age. Given the implausibly wide data range of these reported school times, we only use the median of these reported values for every child age (denoted as s_t), and use that as a measure to define the child's effective time endowment at age t as $\tilde{T}_t = 112 - s_t$. Appendix C.1 contains more details on how we constructed school time.

3.1.2 Parental Labor Supply, Wages and Income

By linking the time survey data from the CDS to the labor supply and income data from the PSID, we can complete the parents' weekly time allocation. We define weekly labor hours of mothers and fathers (h_1 and h_2, respectively) as the total yearly reported number of labor hours for each spouse, divided by 52. Parental "leisure" is the residual time not spent working or with the selected child. This leisure time can include time spent exclusively with a non-sample child, if the household has more than 1 child, as well as any "passive" time spent with the sample child where the parent was not actively involved.[28] From the perspective of the sample child, this definition of parental leisure is inconsequential: time spent exclusively with other children does not directly influence their own development.

Using the main PSID data, we define hourly wages ($w_{1,t}$ and $w_{2,t}$) as the total yearly reported income from labor for each spouse, divided by their total yearly hours. Weekly non-labor income (I_t) is defined as the total yearly household income minus the yearly labor incomes reported for each spouse, divided by 52.

The monetary values have been deflated and are expressed in terms of 2007 dollars. All wage and income information pertaining to years where the sample child was between 0 and 16 years old is used in estimating the model. Note that since the PSID was only administered every two years after 1996, we do not have yearly data for labor supply and income. For the time use surveys and the children's test scores which we only observe every 5 years, these gaps in the data become even more salient. Our econometric implementation (which relies on the Method of Simulated Moments, discussed in Section 4) accommodates this data structure and makes use of all available information.

Table 1 shows the unconditional averages of the parents' hourly wages, weekly work hours and weekly non-labor income. Table 2 summarizes the parents' labor supply behavior for various child ages, both on the extensive and intensive margins. Whereas fathers' labor supply does not seem to vary much with the child's age, mothers tend to work more (along

[28]See DFW for an analysis of the allocation of time across siblings.

both margins) as children age and become more independent.[29] Finally, the first two rows of Table 3 contain the unconditional average work hours of either spouse, averaging over both working and non-working parents.

3.1.3 Cognitive Skill Measures

Given the wide range of ages to which the Woodcock Johnson Letter-Word (LW) test was administered in the CDS (starting at age 3), we use this test as our measure of child cognition.[30] We use the raw scores on this exam rather than the age-standardized scores, and allow our model to describe the development process. The test contains 57 items (so that in terms of our discussion in Section 4, $NQ = 57$), and the range of possible raw scores is from 0 to 57. Our econometric framework accommodates the discreteness and the particular floor and ceiling of this test score measure. Table 1 shows the unconditional average test scores taken from each CDS wave, averaging across all ages. Figure 2 shows how the average test scores increase smoothly with child age.

3.1.4 Conditional Allowances and CCTs

The CDS surveys include useful information on the use of periodic allowances by parents with children who are at least 6 years old. In particular, all three CDS waves asked the child's primary caregiver whether their child receives an allowance or not. Two of the three waves (CDS-II and CDS-III) also asked whether this allowance, if any, was contingent on the child doing their school work. Thus, the joint probability of these two events (i.e., giving a conditional allowance) can be interpreted as a measure of the prevalence of Conditional Cash Transfers (CCTs) in our sample of PSID parents.

Table 4 summarizes the parents' binary (Yes/No) responses to each of these questions in panels (a) and (b), as well as their joint probabilities in panel (c), defined as the product of the two. We divide the data into subgroups based on the child's age as well as on the parents' educational attainment level.[31] We notice two patterns in the likelihood of parents using a CCT. First of all, conditional allowances (or CCTs) are much more likely to be offered to younger children. In particular, 34% of households with children between ages 8 and 11 use CCTs, compared to 24% of households with children between ages 12 and 14, and only 14% of households with children aged 15 or 16. This age gradient is present within all parental education groups, although it is strongest among less educated parents. Second,

[29] Note that the composition of our sample includes households with two children for which, depending on the observable time survey and test score data for the children, we do not always select the youngest sibling. This may partially explain why we do not see a very clear upward trend in maternal labor force participation as the child ages, as has been documented in other studies.

[30] As we discuss in Appendix C.1, LW test scores show a very strong positive contemporaneous correlation (around 90 percent) to alternative measures of cognitive skills, which had more limited data availability.

[31] Given the available data, we only observe the use of conditional allowances for the 2002 and 2007 CDS waves, when all sample children were between 8 and 16 years old.

among households with younger children, CCTs are much more commonly used by parents with lower educational attainment. In the youngest age group, CCTs are used by nearly 50 percent of households with at most a high school degree, compared to less than 20 percent of households with a graduate degree, and roughly 33 percent of households with some college or a college degree.

We note that while the sample sizes in Table 4 are limited, the patterns are qualitatively similar when we look at allowance use for the full CDS sample that includes all surveyed households, including those that do not satisfy our fairly restrictive sample selection criteria. For example, among all 3732 children surveyed in the 2002 and 2007 CDS waves (many of whom appear in both waves), nearly 44% of children between ages 6 and 9 received a conditional allowance (i.e., 274 out of 625 children), whereas only 35% of children ages 10 or above were given a conditional allowance (i.e., 1089 out of 3107 children).

To further motivate our analysis, in Table 5 we present OLS estimates from three simple linear regression specifications in which the child's weekly hours of self-investment time is the dependent variable. In column (1), mirroring Figure 1b, we show that the education of the father is strongly positively related to the hours the child spends studying. The fact that children in higher SES households devote more time to study obviously has the potential to further increase human capital inequality across households. In column (2), we add the total amount of the weekly time investment of the parents, and we see that the child's investment time in herself is negatively related to the parent's investment time in her. This finding is consistent with the tradeoffs in time allocation characterized by our model. Finally, in column (3), we add the amount of allowance (in dollars per week) given by the parents to the child as an observable measure of just one way parents might incentivize child self-investment time. Even conditional on the other covariates, this variable has a positive association with child study time, suggesting that households providing child allowances also have higher levels of child self-investment.

3.2 Data on Child Discount Factors

We use experimental survey data provided by Laurence Steinberg to impute the empirical distribution of children's discount factors (see e.g., Steinberg et al, 2009). Appendix C.2 provides more details on these data, and on the procedure used to impute children's long-run discount factors while accounting for potential present-bias (i.e., using a simple model of quasi-hyperbolic discounting).

Figure 3 plots the empirical distribution of the imputed annual discount factors as a function of the survey respondents' age and parental characteristics. Panel 3a shows the mean and quartiles for the full sample, ranging from age 10 to 30. We notice a positive trend throughout adolescence, with mean and median discount factors increasing by roughly 1 percentage point per year up to age 20, after which the trend flattens out. We also notice substantial heterogeneity across agents, with the interquartile range in annual betas varying from around 0.50 at younger ages to around 0.30 at older ages. Panel 3b plots the average discount factor for the subset of adolescent respondents between ages 10 to 17, conditional on the educational attainment level of the respondent's parents. We find that there is a strong education gradient both among younger children (ages 10-12) as well as among older children (ages 13-17). In particular, children whose parents have at most a high school degree have average discount factors ranging between 0.62 and 0.67, whereas children whose parents have a graduate degree have an average discount factor of around 0.80. Children whose parents have an associate's or college degree fall in between, with discount factors around 0.7 to 0.72. We also find that the positive age trend documented in Panel 3a is mostly driven by children with lower-educated parents, suggesting that there is some catching-up over time in terms of forward-lookingness.

For more details on how we use these data in the econometric implementation of our model, we refer to Section 4 and Appendix D.1.1.

3.3 Data on Parental Discount Factors

To impute the empirical distribution of adults' discount factors, we use survey data from the 2010 U.S. wave of the Osaka Preference Parameters Study ("OPPS").[32] The main objective of these surveys was to measure various economic preferences; time preference, risk aversion, altruism, habit formation, etc. Our selected subsample of interest consists of 4625 individuals between 25 and 65 years old for whom we observe a valid educational attainment level, and who have a valid set of responses to our selected delayed gratification questions. Appendix C.3 provides more details on the procedure used to impute adults' long-run discount factors, and on how we use this data set in our econometric implementation.

Figure 4 shows the empirical distributions of the imputed annual discount factors, as a function of the respondents' educational attainment level. We notice a fairly weak but consistent schooling gradient, with the average discount factor ranging from 0.928 for individuals with a high school degree or less, to 0.935 for individuals with (some) college degree, and 0.945 for individuals with a graduate degree. The dispersion in discount factors is substantial within all education groups, with a standard deviation ranging from 0.057 to 0.071.

[32] Although the experimental survey from the Steinberg et al. (2009) survey also includes information on discount factors for young adults up to age 30, we do not use this information in our empirical specification of the parental discount factors. This is mostly due to sample size restrictions; if we restrict our attention to the subsample of adults who are at least 25 years old and have likely finalized their education, the Steinberg sample would leave us with only 147 adults, whereas the Osaka PPS surveyed over 7000 men and women between the ages of 18 and 99.

3.4 Data on Child Expenditures

Finally, some of the data moments targeted in our estimation procedure stem from the 2017 report on "Expenditures on Children by Families" from the U.S. Department of Agriculture, Center for Nutrition Policy and Promotion (Lino et al., 2017). The report, which is based on Consumer Expenditure Survey data from 2011-2015, estimates annual expenditures on a child for married couples and single families in the U.S. Total expenditures on the child include housing, food, transportation, clothing, health care, child care and education, and miscellaneous expenses like personal care items, entertainment, and reading materials. In our model, the corresponding measure of "total" child expenditures is given by the sum of the child's private consumption denoted by x_t, and productive investments denoted by e_t.[33] According to the CEX data, middle-income, married-couple families (broadly defined as having an annual income between 59,200 and 107,400 dollars, expressed in 2015 dollars) spend around 13,000 dollars per year – or 250 dollars per week – on the child, with only a modest positive trend in expenditures across child ages. In relative terms, these child expenditures make up around 21 percent of total household income.[34] In our estimation, we target both the average total weekly expenditure amount $(e + x)$, as well as the average fraction of household income $(\frac{e+x}{Y})$ as moments.

4 Econometric Implementation and Identification

We now discuss the functional form assumptions we make use of in estimation, how we use the combined PSID-CDS and discount factor survey data to identify the parameters characterizing the model, and describe the estimation method. Additional details can be found in Appendix D.

4.1 Econometric Specification

Preferences Both parents' and children's instantaneous utility functions are assumed to be constant over time and homogeneous across households. However, we allow for heterogeneity in agents' discount factors (both within and across households) which also enter into their terminal value functions at the end of the development period. The child's discount factor is allowed to be time-varying and endogenous. We assume that these discount factors follow a finite discrete types distribution with a parameterized first-order Markov transition matrix that is allowed to depend on CCT usage and the child's age. The parents' discount factor is

[33] Given the ambiguous interpretation of some child expenditure categories, we do not take an explicit stand on which expenditure categories pertain to the child's "private consumption" and which pertain to " productive investment".

[34] See Tables 8 and 9 from Lino et al., 2017, which report average child expenditures adjusted for household size and the age of the youngest child. This is useful since our PSID sample consists mostly of two-child households, of which we usually sample the youngest child.

drawn from a distribution that conditions on the father's educational level and this draw is fixed over time. We also allow the time-invariant fixed cost of using a CCT to vary across households, and assume that the distribution is exponential. See Appendix D.1.1 for further details.

Technology of Cognitive Development We allow the cognitive production function parameters of all inputs to vary monotonically with the age of the child. For the two parental time inputs, we allow the productivity to vary with the education levels of the parents. For the TFP process, we have adopted a flexible generalized logistic parameterization which imposes monotonicity in the child's age. Appendix D.1.2 provides further details.

Wage and Income Processes The hourly wage offer process for each parent depends on their age and education. The wage shocks are assumed to be serially uncorrelated draws but we allow for contemporaneous correlation in the mother's and father's wage offers, reflecting assortative mating or parents inhabiting the same local labor market.[35] See Appendix D.1.3 for more details.

We restrict non-labor income to be non-negative for all households and periods. We assume non-labor income is a function of the parent's ages and their schooling. The non-labor income parameters can be identified and estimated outside of the remaining model structure because the non-labor income process is strictly exogenous and observed for all households.[36] See Appendix D.1.4 for more details.

Measuring Child Cognitive Ability Child ability is assumed to be unobservable to the analyst, although we assume that it is observable by household members since it is a determinant of the household utility level and is a (potential) input into the decision-making process. Most cognitive test scores, such as the one used in our empirical work, are simple sums of the number of questions answered correctly by the test-taker.

Given a cognitive ability test consisting of NQ items of equal difficulty, the number of correct answers, $k_t^* \in \{0, 1, \ldots, NQ\}$ is distributed as a Binomial random variable with parameters $(NQ, p(k_t))$, where $p(k_t)$ is a nondecreasing function between 0 and 1,

$$p(k_t) = \frac{\exp\left(L_{0,t} + L_{1,t} \ln k_t\right)}{1 + \exp\left(L_{0,t} + L_{1,t} \ln k_t\right)}.$$

The randomness inherent in the test-taking process implies that the mapping between latent k and observed k^* is stochastic, and that a child of latent skill k_t has a positive probability

[35] Although we assume the shocks are independent over time, this is not implied by the model structure and is not a necessary assumption for identification of the key model parameters. We can allow for temporal dependence at some computational cost.

[36] We re-estimate all parameters in this process for each bootstrap re-sample so that inference for our full set of model parameters accounts for this first-step estimation.

of answering each question correctly. For a child of latent skill k_t, the expected number of questions they answer correctly is given by $p(k_t)NQ$. Our measurement model then achieves two goals: (i) we map a continuous latent child ability defined on $(0, \infty)$ into a discrete test score measure imposing the measurement floor at 0 and ceiling at $NQ < \infty$, and (ii) we allow for the possibility of measurement error so that a child's score may not perfectly reflect their latent ability. Previous research has often used linear (or log-linear) continuous measurement equations, e.g., Cunha and Heckman (2008), Cunha et al. (2010), Agostinelli and Wiswall (2016). Our approach differs by using a measurement process that explicitly recognizes the discrete and finite nature of the test score measure. We refer to Appendix D.1.5 for more details on the specification of the transformation function $p(k_t)$, and the measurement parameters (i.e., location and scale) that facilitate identification of the model.

4.2 Identification

We now discuss how the primitive parameters of the model can be recovered using the observed data. Given the complexity of the model, we focus on some key identification issues and provide some basic intuition, rather than attempting to provide a rigorous proof of identification.[37] In general, our identification strategy relies on the availability of rich PSID-CDS panel data on parental and child time use, wages, non-labor income, measures of child cognitive ability, and measures of conditional allowance (or CCT) use, as well as valuable cross-sectional data from the Steinberg and Osaka studies on children's and parental discount factors, respectively.

Even with these data we face several identification challenges: (1) the classic non-random selection problem associated with the observation of wage offers only for those who choose to work, (2) the fact that children's skills/ability are not directly observed but only imperfectly measured in our data, (3) the existence of gaps in the time series coverage of the panel data that we utilize, (4) the fact that the we do not have direct data linking households' choices of whether to use a CCT (as observed noisily through the use of conditional allowances in the CDS) to the parent's or child's discount factor, and (5) the fact that we do not have discount factor data for children younger than 10, even though the model simulates data for children as young as age three. This complicates the identification of the parameters governing the initial conditions of the child discount factor process. Appendix D.2 provides more technical details to the discussion provided below.

[37]Since we do not utilize a likelihood-based estimator and since the model is nonlinear in parameters, a rigorous demonstration of identification is problematic in any case.

4.2.1 Wage Offers and Non-Labor Income

The wage offer processes of the parents, although exogenous, cannot be estimated directly outside of the model due to endogenous selection. The log wage offer process is given by

$$\ln w_{h,i,t} = \Xi^{w}_{h,i,t}\boldsymbol{\eta}_i + \varepsilon_{h,i,t},\ h = 1, ..., N;\ i = 1, 2;$$

where $\Xi^{w}_{h,i,t} = [1\ age_{h,i,t}\ s_{h,i}]$ is a vector of observed covariates of parent i in household h at time t, $\boldsymbol{\eta}_i$ is a conformable parameter vector, and $\varepsilon_{h,i,t}$ is normally distributed with mean 0 and variance $\sigma^2_{w_i}$, and where $E(\varepsilon_{h,1,t}\varepsilon_{h,2,t}) = \sigma_{w_1,w_2}$. The disturbances are otherwise independently distributed over time and across households. Although we have access to wage observations for multiple periods, wage observations are non-randomly missing due to the significant number of corner solutions associated with labor supply choices. This type of systematic selection is particularly troublesome when preferences (e.g., discount factors) and technology parameters are heterogeneous in the population. In this case, observing a parent not supplying time to the labor market (and potentially spending more time investing in the child) is consistent with that parent having a low wage offer, or that household having a high discount factor such that they are willing to forgo current utility to be able to invest more time in the child, or that parent's time with the child being highly productive, or for a number of other reasons (e.g., the other spouse having a comparatively high wage or low time productivity). In order to "extrapolate" wages when a large number of households have at least one parent out of the labor force requires parametric assumptions on both parents' wage offer functions. Under our model specification, we can "correct" our estimator of model parameters for the non-randomly missing data using the DGP structure from the model. In this case, both the wage processes and the parameters characterizing preferences and production technologies must be simultaneously estimated.

This is not the case with respect to non-labor income. Since we have assumed that this process is strictly exogenous, and that $I_{h,t}$ is observed for all time periods in which non-labor income data is available for household h, we can estimate this process outside of the model. Identification is achieved both through time-series and cross-sectional variation in the PSID data.

4.2.2 Cognitive Skill Technology

It is typically straightforward to consistently estimate production function parameters if the output level and all inputs are observed, and if the choice of inputs is not a function of any stochastic unobserved component of the production technology. For our cognitive ability production technology, we allow for mis-measured child skills (the output of the skill technology) and an additively separable disturbance term. Identification first requires solving the generic problem of indeterminacy due to the fact that latent child ability/skill k does not have any natural units. As we discuss in Appendix D.1.5, this can be accomplished by normalizing the measurement (i.e., location and scale) parameters for one age, such that we can take the transformation function $p(k_t)$ (which translates latent ability into a probability of answering any test question correctly) as given.

For purposes of discussion, we will assume that we can observe measures of child ability at two ages, t and $t+1$. Then we can write the log production function for household h as

$$\ln k_{h,t+1} = \ln R_t + \mathbf{Z}_{h,t}\boldsymbol{\delta}_t + \phi_t \ln k_{h,t} + \varepsilon_{h,t}, \tag{12}$$

where $\varepsilon_{h,t}$ is the disturbance term for household h at age t, $\ln R_t$ is the age t log of TFP, $\mathbf{Z}_{h,t}$ is the row vector of log time and money inputs selected by household h in period t, $\boldsymbol{\delta}_t$ is a conformable column vector of technology parameters, and ϕ_t is a scalar parameter measuring persistence of the process at age t. Under our modeling assumptions, households all share the same production technology, with the exception that the time input productivities are allowed to depend on two parental characteristics, the schooling levels of the mother and father, in a parametric manner. For now, we will ignore this dependency, but will conclude this section by arguing that this generalized production function is still identified under our parametric assumptions.

Under the assumptions of Cobb-Douglas preferences and a Cobb-Douglas cognitive technology, we have shown that conditional on a binary CCT choice, $CCT_{h,t} \in \{0,1\}$ and on the child's (discrete) period-t discount factor, $\beta_{h,c,t} \in \mathcal{B}_c$, the remaining household choices (including inputs) at age t are not a function of the current level of child cognitive ability, so that conditional on $(CCT_{h,t}, \beta_{h,c,t})$ and on all other household characteristics, $\mathbf{Z}_{h,t}$ and $k_{h,t}$ are independent. Of course, unconditionally they are not, since both depend on state variables that exhibit temporal dependence.[38] Because of the structure of the model, the disturbance term $\varepsilon_{h,t}$ is mean independent of the household's choices and current period state variables, so that $E(\varepsilon_{h,t}|\mathbf{Z}_{h,t}, k_{h,t}, CCT_{h,t}, \beta_{h,c,t}) = 0$, $\forall(h,t)$,, and we assume $\varepsilon_{h,t}$ is independently distributed over time for each household.

If all of the arguments in $\mathbf{Z}_{h,t}$ are available in the PSID and CDS data, given measures of $k_{h,t+1}$ and $k_{h,t}$, conditional on $CCT_{h,t}$ and $\beta_{h,c,t}$, then the OLS estimator of the parameter vector $(\ln R_t \ \boldsymbol{\delta}_t \ \phi_t)$ is both unbiased and consistent. Even though for each household h we only have two or three points in time at which k and $\mathbf{Z}$ are measured, consistently estimating the age-specific parameter vector is possible given sufficiently large numbers of households at each age t, viewing the entire sample as constituting a synthetic cohort of children and parents

[38] Choice variables in $\mathbf{Z}_{h,t}$ are related to the level of child skill indirectly through the parental wage offers, non-labor income, CCT choices and household time preferences. For example, highly educated parents have higher wage offers and non-labor income and their children have higher levels of latent ability. Our model captures all of these fundamental relationships and includes observable data on CCT use and discount factors to facilitate identification.

progressing through the development process.[39] Appendix D.2.1 provides more details on the identification of the cognitive production technology particular to using the CDS dataset and its time gaps between consecutive panel waves.

4.2.3 Preferences, Parenting Style, and Monitoring Cost

Under certain ideal conditions regarding data availability (we observe leisure or can compute it for both parents and consumption of parents and children), all preference parameters can be nonparametrically identified. This is due to the simple structure of the choice problem and the fact that, given all of the state variables of the problem, including household preferences, the actions of the household are uniquely determined. We outline this argument in more detail in Appendices B and D.2.2.

However in this model identification is complicated by the heterogeneity in "parenting styles" generated by their decision of whether to use a CCT.[40] This means that there are two different mappings between the observed data and the model parameters, each corresponding to the choice of CCT use in the period. This problem is alleviated due to the fact the we make use of data that indicates the choice of parenting style in the period, which is whether the household makes use of conditional allowances. By letting households choose whether to use a CCT in each period, and then linking those simulated decisions to data moments on households' endogenous CCT use to their exogenous characteristics (i.e., the child's age and the parents' education level), we can resolve the indeterminacy associated with the behavioral regime of the household, and identify all model parameters while allowing for "mixing" between the two parenting styles, both across households as well as within households over time.

Given the gaps in the CDS data (we observe CCT use directly for some households in only some periods), we do not observe whether a particular household is using a CCT in any given period. However, we can identify a monitoring cost distribution using the observed frequency of CCT use (proxied using conditional allowances), and potentially even make it conditional on the child's age and other household demographics.[41]

[39]One issue is that we do not have complete data on expenditures on child specific goods, an element of **Z**. As we discuss below and detail in Appendix D.2, we lean on our model structure, in particular the implication that child good expenditures in any period are a given function of household income, up to some unknown preference parameters. These preference parameters are identified from other household behavior (time allocation choices), as discussed below.

[40]This problem is discussed at length in Del Boca and Flinn (2011), where households choose between these two different modes of behavior and where the choice of the household was unobservable. In their nonparametric setting, they showed that there existed two mappings from the observed actions to the parameters characterizing the household, one corresponding to each mode of behavior. In this case, each mapping must be used to determine the parameter values that would have generated it under the assumed form of behavior, and then each is evaluated to determine whether the assumed mode of behavior would have been the one chosen under those parameter values. In some cases, each of the two generated parameter vectors is consistent with the assumed behavioral rule being preferred by the household, which is essentially a case of multiple equilibria.

[41]Given the identification of the other model parameters, for each household, we can first compute the

Child Discount Factors Our model allows child discount factors to be both heterogeneous and endogenous to the parents' use of incentives (CCT). We parameterize this endogenous patience production relationship using a Markov transition matrix defined with respect to the 3 discount factor values we have defined. As specified in Appendix D.1.1, if the parents do not use a CCT in period t, the discount factor Markov transitions are fully determined by the four parameters $(b_0^{up}, b_1^{up}, b_0^{down}, b_1^{down})$. If parents do use a CCT, its effect is allowed to shift the overall level of the Markov transition probabilities (through parameters b_2^{up} and b_2^{down}, which are assumed to sum up to 0) as well as the age-dependence (through parameters b_3^{up} and b_3^{down}, which are also assumed to sum up to 0). This parsimonious structure helps us fit certain data moments relating to the proportion of parents that use a CCT (i.e., those who give their child an allowance contingent on doing school work), and the correlation between CCT use and household characteristics such as the child's age and the parents' educational attainment. For example, since we observe that CCTs are much more commonly used by parents who have younger children, we may infer that using a CCT is particularly harmful for older children, in terms of "crowding out" the child's intrinsic motivation by (stochastically) reducing their future discount factor. In this case, we would expect the interaction coefficient b_3^{up} to be negative, or vice versa, b_3^{down} to be positive. Analogously, the moments capturing the empirical distribution of children's discount factors as a function of their age, as shown in Figure 3, help us identify the remaining parameters characterizing the Markov transition matrix.

One final identification challenge follows from the fact that we cannot observe children's discount factors before age 10, since the Steinberg data set only covers children between the ages of 10 and 17. This requires making an additional assumption regarding the child's initial discount factor process, as specified in Appendix D.1.1. By construction, varying the ν parameters induces a different conditional distribution of child (initial) discount factors between ages 3 and 12. This distribution, in turn, affects all of the households' decisions, including parental time investments, child self-investment time, cognitive test scores and CCT choices, all of which we do observe both before and after age 10 through the CDS surveys which span ages 3 through 17. Moreover, the recursive Markov structure governing the child's discount factor implies that the child's initial discount factor (as determined by the ν parameters) will affect their entire sequence of future discount factors up to age 17. For example, if the ν parameters are such that children have high initial (latent) discount factors, then the persistence of the Markov process will propagate this by inducing counterfactually high discount factors between ages 10 and 17. By combining the PSID-CDS data moments with the additional moments generated from the Steinberg data we can jointly identify the

difference in period t value functions of the parents when they use a CCT and when they do not. Using simulation, we can then compute the probability of using a CCT for all households (conditional on observable variables). The parameter characterizing the monitoring cost distribution then uniquely determines the mapping between this simulated model probability and the sample frequency.

ν parameters together with all of the other parameters of the model.

4.3 Computation of the Model Solution

Our model does not have a simple closed form solution for all endogenous choices. Instead we use a mixed numerical and analytic solution, in which we compute some endogenous outcomes on a grid (i.e., the child's ability, k_t and discount factor, $\beta_{c,t}$), with numerical solutions for some choices (i.e., parental time investments and binary CCT choices), and, given these choices, analytical solutions for the remaining choices (i.e., labor supply, leisure, parental and child consumption, child expenditures, child self-investment time, and CCT reward parameters). We then substitute these choices into the utility function to determine the relative utility associated with each grid point, and use the maximum utility value choice as the optimal choice. We leave the details of the multiple-step model solution, including backward induction steps, to Appendix B.

4.4 Estimation

The estimator utilizes simulation methods. The Appendix details the simulation and estimation procedure. In brief, given a trial vector of primitive model parameters and the initial observed sample characteristics of each household, we simulate R sequences of household state variables (wage offers, non-labor income, parental discount factor, child's initial discount factor), agent actions (time allocation, expenditures, CCT choices), child ability and discount factors given R draws from the parametric random variable distributions. Using the simulated data set, we then compute the analogous simulated sample moments to those determined from the actual sample. We then use the difference in simulated versus data moments, weighted with a diagonal weight matrix reflecting the sample variance of each moment, to form a standard Method of Simulated Method of Moments objective function. We use numerical methods to find the minimum distance vector of parameters and discuss sample fit below.

The moments we target include averages and standard deviations of hours of labor for working mothers and fathers; labor force participation rates for mothers and fathers; productive investment time for mothers, fathers and children; child test scores and within-child changes in test scores over time; and frequency CCT use; computed conditionally for various child age and parental education categories. In addition, we target the average and standard deviation of accepted wages and the correlation in wages across parents. We also compute a number of contemporaneous and lagged correlations, between (i) current and future child test scores, (ii) parental time inputs and (changes in) child test scores, (iii) child self-investment time and (changes in) child test scores, (iv) parental schooling and child test scores, (v) parental schooling and parental/child investment time, (vi) CCT use and child age, parental education, parental/child investment time and child test scores. We also include external data moments related to children's average discount factors (conditional on the child's age and/or parental education level) and total child expenditures (in levels and as a fraction of household income), using the external data discussed in Sections 3.2 and 3.4, respectively. It is important to note that while we do not always observe child inputs, labor supply, wages, and income in the same periods, our simulation method allows us to combine moments from various points in the child development process into a single estimator.[42] Essentially the simulation procedure amounts to Monte Carlo integration of the unobserved data over the entire development process of the child.

The exogenous process determining household non-labor income is estimated outside of the model, using a separate Method of Simulated Moments routine that targets various moments related to the empirical non-labor income distribution, both unconditionally as well as conditional on parental age and education levels. Similarly, the conditional distribution of the parents' discount factors is calibrated exogenously using the Osaka PPS data discussed in Section 3.3.[43] We refer the reader to Section D.1.1 and Appendix C.3 for more details.

5 Model Estimates

5.1 Household Preference Parameters

Given the identification issues discussed in the previous section, we assume that the altruism parameter is fixed at the value $\varphi = \frac{1}{3}$ for all households. This value is relatively consistent with the broad range of parental altruism parameter estimates that are mainly found in the household macroeconomics literature.[44] This implies that parents value their child's utility about half as much as their private utility.

Any investments made by the parents and the child after the final age $M = 16$ are not explicitly modeled, but can be captured by the terminal period utility functions specified in Section 2.1.3. Note that while periods are in years, the model is specified in terms of weekly decisions, which are considered to be invariant within each yearly planning period.

[42] A full list of the moments utilized in the estimation procedure is available upon request.

[43] For children, the parameters governing the initial conditions and endogenous evolution of the discount factor process are jointly estimated with the rest of the model.

[44] Altruism parameters are estimated, or calibrated, most often in macroeconomic studies of household behavior. Moreover, it is usually the case that the altruism parameter is defined as $u_P + \varphi u_c$, whereas we have defined the parents "final" utility as a weighted average of their private utility and the child's utility, or $(1-\varphi)u_p+\varphi u_c$. In a model of the child's decision to leave the household, Kaplan (2012) finds an estimate of φ of 0.039. In Boar's (2021) study of parents' precautionary savings for the child's risky income, the calibrated value of φ is 0.201. In his study of bequests and inter-vivos transfers, Nishiyama (2002) finds a value of φ of 0.512. Finally, Barczyk and Kredler (2018), in a study of long-term care decisions for elderly parents, find that the calibrated value of parental altruism is 0.692. As can be seen, the range of the estimates is large, and depends on the application and specifics of the model. Our value of $\varphi = \frac{1}{3}$ in our weighted-average formulation implies a value of $\varphi = 0.5$ in the $u_p + \varphi u_c$ specification, which puts it in the middle of the range of values found in these studies.

Table 6 presents the estimates of the preference parameters for parents and children. The estimates suggest that parents put the most weight on the father's leisure (0.32), followed by private parental consumption (0.31), the mother's leisure (0.22), and finally child skills (0.15). Conversely, the child attaches a relatively low weight to their own consumption and skill level (about 0.10 and 0.07, respectively), and cares primarily about leisure (λ_1 approximately 0.83). However, the parents' relative preference order is altered when taking into account the fact that they are altruistic, which causes them to care more about child leisure than anything else. Indeed, the *total* parental weight on child leisure, $\tilde{\alpha}_5 = \varphi\lambda_1$ is approximately 0.28, giving it a higher total weight than the one on mother's leisure ($\tilde{\alpha}_1 = 0.15$), father's leisure ($\tilde{\alpha}_2 = 0.21$), parental consumption ($\tilde{\alpha}_3 = 0.21$), child skills ($\tilde{\alpha}_4 = 0.13$), or child consumption ($\tilde{\alpha}_6 = 0.03$).

Finally, we estimate that κ, the parameter representing the mean and standard deviation of the (exponential) CCT cost distribution, is approximately 0.0033. While difficult to directly interpret, we will show what this estimate implies in terms of simulated household behavior, and CCT use in particular, by presenting comparative statics exercises in Section 6.1. As was discussed in Section 3.1.4, our empirical measure of CCT use (as proxied by the proportion of PSID-CDS parents who give their child an allowance conditional on doing school work) decreases with the child's age and with parental education levels. Since the utility cost associated with implementing an CCT is assumed to remain constant for a given household over time, this trend is explained in terms of the decreasing value of using incentives as the child matures.

5.2 Technology of Cognitive Skill Formation Parameters

The Cobb-Douglas productivity parameters for household h associated with each productive input $j = 1, ..., 5$ ($\delta_{h,j,t}$) are allowed to change (monotonically) with the age of the child t, by specifying an intercept ($d_{j,0}$), an age slope parameter ($d_{j,1}$) for each input, and parental education slope parameters ($d_{1,2}$ and $d_{2,2}$) for the two parental time inputs. The total factor productivity parameters (R_t) are estimated using a more flexible generalized logistic specification. Panel (a) of Table 7 and Figure 5 display the estimates of these production technology parameters. Due to the exponential transformation function, the raw estimates of the intercepts and slopes relating to each input are difficult to interpret. A larger (i.e. less negative) intercept indicates a higher initial productivity at the first relevant child age, $t = 3$. A positive (negative) age slope estimate implies that the productivity increases (decreases) monotonically with child age.

In Panel 5a we plot the estimated sequences of all time inputs as a function of the child's age given average parental schooling levels in the sample. We find that active parental time inputs are highly productive at young ages, but become less productive as the child ages. Consistent with the literature, we estimate that maternal time is the most productive of all inputs for very young children. Paternal time is less productive than maternal time inputs at all child ages. The absolute levels and decreasing time trends in the productivity of maternal and paternal time are in accordance with the findings of DFW. Unsurprisingly, the productivity of the child's self-investment time is initially very low. When the child reaches adolescence, however, self-investment time gradually becomes more productive and even becomes more productive than both parental time inputs. Given our specification of the model, these relative trends make intuitive sense. Once children start formal schooling and become more influenced by teachers and fellow students, their time allocation starts to shift away from spending time with parents and other inputs start to become more relevant.

Panels 5b shows the degree to which parental education levels shift the productivity parameters at various child ages. Although years of schooling enters continuously into the parametric specification, the graphs focus on high school graduates versus college graduates (i.e. 12 years versus 16 years of schooling) for simplicity. For both sets of parents, we notice that schooling has a positive effect on the parent's individual time productivity. This suggests that all else equal, higher educated parents are more efficient at producing child cognitive skills.

Panels 5c and 5d show the estimated productivity parameters of the two remaining inputs as a function of child age: material child goods ($\delta_{4,t}$) and lagged child human capital ($\delta_{5,t}$). There is a positive trend in the productivity of weekly child expenditures (although this is hard to compare to the time inputs due to the different units of measurement: dollars versus hours per week). Conversely, we estimate that the productivity of lagged skills is fairly high and increases with age, starting off at 0.79 in early childhood and reaching 0.84 by age 16. The strong level of persistence in the estimated production technology suggests that the self-productivity of skills - or the principle that "skills beget skills" - is salient at all ages, in a pattern that is consistent with other evidence from the literature.

Panel 5e shows the estimated (log) total factor productivity process ($\ln R_t$) as a function of child age. We find that there is an overall positive trend at younger ages, stabilizing at around 0.92 after age 10.

5.3 Child Discount Factor Parameters

Panel (b) of Table 7, Figure 6a and Figure 6b display the parameter estimates relating to the child's endogenous discount factor process, i.e. the Markov transition probabilities as well as the initial conditions specification discussed in Section D.1.1. The estimates of the initial conditions process indicate that, all else equal, discount factors are increasing with the child's age and with the parents' educational attainment level. Analogously, the estimates of the Markov process capture the empirical observation that, on average, children's discount factors increase with age. Interestingly, the (stochastic) effect of using a CCT on the child's future discount factor is negative at all ages, but with a slightly worse effect for older children.[45] For example, the top row of Figure 6a indicates that for a child with a low initial discount factor, the parents' choice to use a CCT would increase the probability that the child's next-period discount factor remains low by approximately 15 percentage points at age 10, and by approximately 20 percentage points at age 16. Taking expectations, this implies that using a CCT would reduce the child's expected discount factor in the next period by 13 percent at age 11 (from 0.493 if the parents did not use a CCT at age 10, to 0.427 if they did), and by 17 percent at age 17 (from 0.503 if the parents did not use a CCT at age 16, to 0.418 if they did). Similarly, the bottom row of Figure 6a shows that for a 16-year old adolescent with a high discount factor, the usage of a CCT would decrease the probability that their next-period discount factor will remain high by approximately 15 percentage points. In expectation, the child's discount factor at age 17 would be around 0.957 if their parents did not use a CCT, and around 0.925 if they did. While this difference may not seem very large, it can meaningfully affect the child's future decisions and welfare, in part due to the complementarity between the child's final stock of cognitive skills and their final discount factor. Given our specification of the child's final-period utility function in Equation (2), the child's marginal utility of cognitive skills ($\lambda_3/(1-\beta_{c,17})$) would be roughly equal to 1.70 if $\beta_{c,17} = 0.957$, but would drop substantially to around 0.97 if $\beta_{c,17} = 0.925$. Since the parents' final-period utility function also depends (altruistically) on the child's final discount factor, this helps rationalize why parents with adolescent children are less likely to use a CCT than are parents with younger children, since (1) older children are more patient and therefore more willing to self-invest, and (2) the short-run incentive effect of using a CCT may no longer be sufficient to offset the negative effect on the child's future patience (and hence future self-investment) levels.

5.4 Wage and Non-labor Income Process Parameters

Table 8 provides the estimates of the wage and non-labor income processes. As specified in Section D.1.3, the (log) wage offer distribution for each spouse depends linearly on their observable characteristics, i.e. age and completed level of education.[46] We estimate that each additional year of education increases the wage offer by almost 9 percent for women and nearly 8 percent for men. Given the fairly small age range of the parents in our sample, we estimate a flat earnings profile that is similar for both spouses, where each additional year amounts to a mean wage increase of about 1.2 percent. Since most parents work full-time in all observed periods (especially fathers), we could interpret this as the average return on experience. When compounded over the relevant parental age range of $30-50$, this would amount to a total mean wage increase of about 27 percent for each spouse, which is not

[45] We would like to re-emphasize that the sign of the CCT effect was not constrained to be negative within our econometric specification.

[46] More complex polynomials, including e.g. an age-squared term, a birth cohort effect, or an interaction between age and education, had small and imprecisely estimated coefficients.

unreasonable. The remaining three wage parameter estimates $(\sigma_{w_1}, \sigma_{w_2}, \rho_{12})$ determine the variability of the periodic wage shocks of each spouse, which are allowed to be correlated with the wage shocks of the other spouse. We estimate that this correlation is strongly positive at 0.69. This could again be interpreted as evidence of positive assortative matching in the marriage market (or correlated local labor market shocks). Our estimates also suggest that the fathers' wage shocks are more volatile than the mothers' shocks.

Since there is no endogenous selection on non-labor income as there is on wages, the parameters of the non-labor income process can be estimated separately from the rest of the model. We refer to Section D.1.4 for more details on the two-step procedure used to estimate the non-labor income process, which is tailored to (1) fit the large observed fraction of households in the data that report a zero non-labor income in any given year, and (2) capture the observable heterogeneity in non-labor income draws. The first seven parameter estimates in part (c) of Table 8 indicate that the probability of having a strictly positive non-labor income in any period increases strongly with the father's education level, and is slightly concave in the father's age in that period. The remaining parameter estimates show that, conditional on having a strictly positive non-labor income draw, the mean non-labor income increases by about 4.5 percent each year, and does not seem to be significantly impacted by the education levels of the parents. The large standard deviation of the disturbances is not surprising given the small number of covariates used in the empirical specification.

5.5 Within-sample Fit

5.5.1 Test Scores

Figure E.4a plots the average Letter Word test score as a function of the child's age, where the solid line represents actual CDS data and the dotted line represents the simulated test scores under the model. The estimated model captures the S-shaped trend in the measured test scores almost exactly.[47]

While not depicted, the estimation also targets other moments related to the test score distribution, including the standard deviation of test scores at various ages, the averages of the within-child changes in test scores over time at various ages, the within-child correlation between current (time t) and future (time $t+5$) test scores at various ages, or the within-household correlation between child test scores and parental education levels. All of these moments were fitted very closely. For example, the standard deviation of test scores between ages 13 and 16 is 4.02 in the data, and 4.22 in the simulation. Similarly, the within-child correlation between time t and time $(t+5)$ test scores for ages $t = 3, ..., 7$ is 0.46 in the data,

[47] In general, there are two sources of "noise" in the average level of measured child skills at each age. First, the average child skill level at each age is an aggregate of the child skill levels for our sample of heterogeneous households. Second, we are displaying the simulated measure of child skills using our stochastic measurement model, and the simulated measure is noisy due to the fact that the number of simulation draws is finite.

and 0.42 in the simulation.

5.5.2 Time Allocations

Regarding the endogenous time allocation decisions of parents and children, the model is able to fit most of the basic patterns in the data. Table 9 displays the sample fit of the parental labor supply choices along the extensive and intensive margins, the two types of parental time investments and the child's self-investment time, broken down into four non-overlapping child age bins.

The model correctly predicts that mothers of older children are more likely to supply some positive amount of labor to the market. For fathers, the model slightly overpredicts the labor force participation rates at older ages, which could be due to the relatively low variation in paternal labor supply behavior in the data. When looking at the subset of working parents, the model fits the average weekly hours of supplied labor well for both spouses. For working mothers, the model again captures the positive trend in labor supply as children age.

Overall, the model fits the average levels of the two types of parental time investments observed in the data closely for most of the child age bins. Importantly, the strong negative trends in the observed maternal and paternal active time investments are reflected in the simulated data. The increasing trend in the child's self-investment time is also captured by the model, although the levels are slightly off for older children, possibly due to some outliers in the data, and the targeting of other moments besides simple means and standard deviations.

Figures E.4 and E.5 present a more detailed fit of the various time investments and labor supply choices, broken down by child age. Overall, the relatively close fit between the data moments and the simulated moments confirms that the results in Table 9 would be robust to different choices of the child age bins.

Although not shown in the table, the model also fits the standard deviations of the weekly labor supply and time investment choices at various child ages. For example, the standard deviation of children's weekly self-investment time between ages 13 and 16 is quite large but fitted relatively precisely (i.e., 6.98 in the data, 7.45 in the simulation).

The estimation also targets other moments linking weekly time investments to test scores, or to changes in test scores over time, or to parental education levels. For example, the contemporaneous correlation between the child's self-investment and their Letter Word score across all ages is roughly 0.36 in the data, and 0.49 in the simulation. Similarly, the correlation between the mother's active time invested at time t and the change in the child's test score between time t and time $(t+5)$ is roughly 0.28 in the data, and 0.40 in the simulation.

5.5.3 Wages and Non-labor Income

Table 10 shows how the simulated model is able to fit the basic moments of the empirical wage distribution fairly precisely. The basic unconditional moments of the exogenously estimated non-labor income process are matched precisely, including the fraction of strictly positive non-labor income observations. Any differences between the data moments and the simulated moments in this table are due to the fact that we are not only targeting these unconditional moments, but also several other moments of the wage and non-labor income distributions where we condition on parental age and education.

Figure E.6 shows a more detailed breakdown of the parents' hourly accepted wages and weekly non-labor income in the data and under the model, as a function of the parents' age or their completed education level. Although these conditional moments are relatively noisy due to small sample issues in the data, we find that the overall fit is reasonably good. For mothers, panels E.6a and E.6b indicate that the model captures the increasing wage trend in both the mother's age as well as her education level, although the data become noisier for older mothers due to smaller sample sizes.[48]. Panels E.6c and E.6d show that the model also captures the overall age and education gradients in fathers' accepted wages. For the non-labor income process, panels E.6e and E.6f show how various conditional moments are also fit relatively well.

5.5.4 Parental Use of CCTs

Table 11 presents some of the key moments relating to the parents' use of extrinsic motivation schemes, i.e., the binary CCT choice variable. While we somewhat overpredict the overall proportion of household-periods where a CCT is being used (0.24 in the data, 0.35 in the simulation), we broadly fit the correlation patterns between CCT use and various household characteristics, time investments, and child test scores. The overall negative correlation between parental CCT use and the child's age is well captured by the model (-0.20 in the data, -0.23 in the simulation). Similarly, the negative association between CCT use and parental education is qualitatively the same in the simulation, albeit the correlations are somewhat stronger in the data. The negative correlation between CCT use and same-period child test scores is also closely fitted (-0.19 in the data, -0.21 in the simulation). Finally, the weakly positive correlations between CCT use and parental time investments are also replicated in the simulation. Despite our relatively limited and noisy empirical measure of the use of parental incentives, these results suggest that the model is able to capture most of the relevant trends and variation in CCT use.

[48] We do not simulate wages for very young parents, since the model is only specified for children of age 3 and above.

5.5.5 Parental and Child Discount Factors

As discussed in Section D.1.1, the parental discount factor process is assumed to be exogenous and is estimated separately from the rest of the model. Panel (a) of Table 12 presents the model fit of this process. The left two columns show data moments, and the right two columns show simulated moments. Overall, the key moments of the simulated distribution of parental discount factors (i.e., means and standard deviations conditional on parental education level) are very close to those of the empirical distribution in the Osaka PPS data, exhibiting the same conditional averages, dispersion, and education gradient.

Panel (b) of Table 12 presents the key moments of the distribution of children's discount factors, conditional on the child's age and on the parents' educational attainment. Since the survey by Steinberg et al. (2009) only sampled children who were at least 10 years old, we cannot report any data moments for children younger than 10, but we still report their simulated moments for completeness. Overall, the model is able to fit the levels and dispersion of children's discount factors, as well as the positive age and education gradients that we observe in the data. We estimate that the average child's discount factor grows monotonically from around 0.4 in early childhood (ages $3-5$) to around 0.8 in adolescence (ages 13-17).

While not directly targeted as moments in the estimation, we note that the correlation in the Steinberg et al. data between the child's discount factor and the parent's years of schooling was weakly positive at around 0.05, which is comparable to the correlations we find in the simulated data (which were 0.05 for mother's education and 0.10 for father's education). Moreover, the empirical correlation between the children's discount factor and their cognitive ability (as measured by their IQ score) was roughly 0.18 between ages 10 and 17, which is comparable to the correlation between children's discount factors and our measure of the child's ability (i.e., the Letter Word score) in our simulated data set ($corr = 0.13$).

5.5.6 Consumption, Expenditures, Leisure and CCT Choices

In Figure E.7, we present some simulated moments of variables which are either not explicitly used in the estimation or not observed in our data set. Each of the six panels shows average simulated values as a function of the child's age. Panel E.7a shows how the average weekly household income (Y_t) increases steadily with child age, and how this trend is reflected in each of the three expenditure components, although parental consumption is by far the largest spending category. Overall, total expenditures on the child (i.e., private consumption x as well as productive investments e) amount to about 260 dollars per week, or around 16 percent of the total household budget, which is in line with other findings in the literature based on comparable samples from CEX data (see also our discussion in Section 3.4).

Panel E.7b plots the weekly (residual) leisure time enjoyed by each household member.

Although mothers tend to spend more time with children than fathers, lower female labor supply causes the average mother's leisure time to be greater than the father's by approximately 10 to 12 hours per week.[49] Due to the increase in the child's exogenously defined school time at age 6, the child's residual leisure time takes a sharp drop at that age.[50]

Panel E.7c shows the average fraction of simulated households choosing to implement a CCT at each age, conditional on the father's educational attainment level. Consistent with the PSID-CDS data patterns on the use of conditional allowances shown in Table 4, this fraction decreases as children transition from childhood to adolescence, with more highly educated parents being less likely to use a CCT at nearly all ages. Panel E.7d shows how the average reward function elasticity (r_t) chosen by the subset of parents who implement a CCT increases markedly as children age, with more highly educated parents offering larger incentives. Our estimates support the notion that, as children age, fewer parents choose to incentivize their children by offering a contingent reward system, while those that choose to do so offer higher, and not lower, incentives to study.

Finally, Panels E.7e and E.7f exhibit the average reward function intercepts (i.e., b_t) and resulting total child consumption (i.e., x_t) conditional on the father's educational attainment. We notice that the intercept parameter (which could be interpreted as a consumption floor for the child) decreases with the child's age. However, this pattern is completely driven by the subset of parents who choose to implement a CCT, and who tend to incentivize the child by reducing the consumption floor (low b_t) while increasing the reward for studying (high r_t). The resulting amount of consumption (i.e., x_t) is more or less stable over time, with a moderately positive trend in child age. While more educated parents are less likely to use CCTs overall, their higher incomes on average still allow them to offer the child more private consumption (e.g., through unconditional allowances) than less educated households.

6 Comparative Statics Analysis

6.1 Varying the Costs of CCTs

In our benchmark model, we assumed that the parents' choice of whether or not to use a CCT could affect their utility through two distinct channels: (1) the instantaneous fixed utility cost of using a CCT, defined as ζ, and (2) the stochastic effect of using a CCT on the child's future discount factor, through the parameters (b_2^{up}, b_3^{up}) that enter the Markov process specified in D.1.1. In this section, we consider the comparative statics effect of altering these parameters associated with parental CCT choices.

[49] Note that this could be partially attributed to the absence of household production in the model.

[50] Although irrelevant for the estimation procedure or any of the other results, this sudden drop in child leisure could be partially smoothed out by recognizing that children's average sleep time decreases as they age.

Table 13 shows the results of this exercise, where we compare simulated household choices and outcomes under the baseline model to those produced under four alternative scenarios. Since this is a comparative statics exercise, all other parameters of the model remain fixed.

First, in column (1), we consider the case where the instantaneous utility cost, ζ, is infinitely large for all households (i.e., the parameter κ governing the exponential cost distribution is allowed to go to infinity). In this case, CCTs are essentially unavailable (or undesirable) for all households, and the child's discount factor simply evolves over time according to a simplified Markov process that only depends on the child's age and initial conditions. In this case, the child's average discount factor at age 17 would increase substantially to around 0.88, compared to 0.81 in the baseline model. However, given the reduction in CCT use and child self-investment time, the average final stock of (log) cognitive skills would decrease by roughly 0.17 log-units, or 40 percent of a standard deviation. Children's welfare, as measured by their average Bellman value across all ages, would increase in this scenario, likely due to having more leisure and becoming more forward-looking. Conversely, parental altruistic welfare would decrease, since they can no longer use incentives to influence the child's behavior.

In the second scenario, in column (2), we assume that CCTs are costless for all households (i.e., $\kappa = 0$) but can still endogenously affect the child's future patience level. The fraction of household-periods where CCTs are used increases only modestly in this situation, by approximately 2 percentage points on average. As a result, average child self-investment time would only increase marginally, and the child's final-period outcomes or household (approximate) welfare levels would not be much affected relative to baseline. However, the simulated correlations between CCT use and parental characteristics or the child's contemporaneous test score (all of which are moments targeted in the estimation) would become more negative, which suggests that allowing for $\kappa \geq 0$ ultimately improves the model fit, conditional on keeping all other model parameters fixed (i.e., in a "local" identification sense).

Third, in column (3), we consider the case where CCTs are costly (i.e., κ is equal to its baseline estimate), but the child's discount factor process is assumed to be independent of the parents' CCT choice (i.e., the intercept and slope parameters b_2^{up} and b_3^{up} in the Markov process are assumed to be zero). Now, the fraction of household-periods where a CCT is used increases dramatically to 90 percent, causing an increase in the final stock of cognitive skills of approximately 0.34 log-units, or over 78 percent of a standard deviation. The child's final-period discount factor, now unaffected by any past use of CCTs, would increase to 0.88, which is identical to the scenario in column (1) where CCTs were simply unavailable. Despite the sharp increase in child self-investment time (and hence a decrease in child leisure), child welfare levels would still be larger than in the baseline, as would parental welfare levels.

Finally, in column (4), we consider a scenario that combines the alternatives considered in columns (2) and (3), such that CCTs are assumed to be costless and no longer affect the child's patience (i.e., $\kappa = b_2^{up} = b_3^{up} = 0$). In this case, as is expected, all parents would use a CCT in all periods. Analogously to our results in column (3), this would cause children to choose substantially larger amounts of study time and end up with larger final stocks of both cognitive and non-cognitive skills, with both parents and children enjoying more welfare. The similarity of these results to those in column (3) reflects the fact that under our baseline estimates, the main reason parents prefer not to use CCTs is due to their negative "crowding-out" effect on the child's future patience levels, with the fixed instantaneous utility cost playing a relatively minor role. At the same time, the overall comparison of columns (1) through (4) indicates that restricting the channels through which CCTs can affect parental welfare and/or child discount factors would have significant counterfactual implications on various (targeted) moments, including measures of cognitive ability, children's patience, time investments, CCT use, and the correlation between the use of CCTs and other observables such as the child's age, the parents' education level, and contemporaneous test scores.

6.2 Parental Schooling and Child Outcomes

In the baseline model parental demographics appear as a shifter of various primitive parameters: (1) in the budget constraint, where age and education are a determinant of hourly wage offers and weekly non-labor income, (2) in the technology of cognitive skill formation, where education levels affect the productivity of parental time inputs, and (3) in time preferences, where the conditional distributions of both the parents' as well as the child's (initial) discount factors are allowed to depend on the father's education level. Although the estimated model is able to broadly match the conditional heterogeneity patterns observed in the data (e.g., in wages, non-labor income, time use, CCT use, test scores and discount factors), it is unclear to what extent this heterogeneity in choices and outcomes is driven by the across-household variation in wages, non-labor income, productivity or time preferences. In this section, we investigate the role of the household's socio-economic status in accounting for the cross-sectional heterogeneity in household choices and child outcomes. For simplicity, the analysis focuses mainly on breaking down the heterogeneity in children's final-period stocks of cognitive skills and discount factors, conditional on the father either having at most a high-school degree ("low SES") or a graduate degree ("high SES").[51]

The simulated results are shown in Table 14, which compares the baseline model to four alternative specifications which gradually restrict the amount of heterogeneity in the model primitives. The first alternative specification, summarized in column (1), shuts down the age-based heterogeneity in parental wage offers and non-labor income levels. Effectively, this is done by inserting the median age for each spouse (i.e., 35 for mothers and 36 for fathers) into the wage offer and non-labor income equations, as defined in Appendix D.1.3. Since

[51] All results are robust to alternative specifications where SES is defined using either the mother's schooling level, or both parents' schooling levels, or using different definitions of "low" versus "high" SES.

higher-educated parents tend to be older (conditional on the child's age), this reduces the simulated wage gap between high SES and low SES households by roughly 17 percent for mothers and 13 percent for fathers, and reduces the household income gap by 17 percent. However, we notice very little effect on the simulated gaps in child outcomes, parental time inputs, or CCT use.

The second specification, in column (2), additionally shuts down the education-based heterogeneity in wage offers and non-labor income, by fixing every parent's schooling level to the median education level observed in the data (i.e., 15 years for mothers and 14 years for fathers). In this case, all parents receive the same wage offers by gender, up to an i.i.d. wage shock with gender-specific variance. In this case, the simulated average wage gaps between high and low SES households become very small (less than 3 percent in absolute value).[52] Somewhat surprisingly, the results indicate that shutting down wage and income heterogeneity has little effect on the resulting gaps in child outcomes, parental time investments or CCT use. As such, our model suggests that these gaps may be hard to remedy through pure income-based cash transfers, as they are primarily driven by differences in parental time productivity, (time) preferences, and initial conditions.

The third specification, in column (3), further restricts the model such that parental time productivity parameters are the same for all households, conditional on the child's age.[53] Consistent with the baseline estimates indicating a strong positive effect of education on parental time productivity, shutting down this channel substantially reduces the gap in parental investment time between high and low SES households (i.e., by roughly 44 percent relative to baseline). As a result, the gap in the average level of cognitive skills at the end of childhood also shrinks significantly, by around 61 percent relative to baseline.

Finally, column (4) considers an even more restricted specification that no longer allows the distribution of parental discount factors, or the distribution of the child's initial discount factors, to depend on the parent's schooling level. This greatly reduces the simulated gaps in children's final-period cognitive skills (84 percent relative to baseline) as well as final-period discount factors (79 percent relative to baseline).[54] The remaining gap by SES reflects differences in initial cognitive ability, taken as exogenous. Children from high SES households tend to have higher initial test scores, and this initial heterogeneity can persist over time given the dynamics in cognitive skill formation. Moreover, the complementarity between cognitive

[52]Since the table depicts the average *accepted* wages, any residual difference is because of labor supply selection, and small-sample noise in the wage error terms.

[53]Analogously to column (2), the models in columns (3) and (4) simply insert the parents' median schooling levels into the functional forms describing the parental time productivities and/or the distributions of parent/child (initial) discount factors, as specified in Appendix D.1.2 and D.1.1, respectively.

[54]While the SES-based gap in the average child discount factors may not seem very large in absolute size, we note that small differences can still cause meaningful differences in final-period marginal valuations, which are related to the "perpetuity" factor $1/(1-\beta_{c,17})$. For example, a gap of only three percentage points in the annual discount factor (i.e., 0.79 versus 0.82) translates to a difference in the "perpetuity" factor of over 16 percent (4.76 versus 5.56), which can meaningfully impact future investment behavior.

skills and patience levels (seen, for example, in the final-period utility functions) may cause spillovers between the two outcomes, through parenting choices made during childhood and adolescence. Indeed, column (4) further shows that there are non-negligible residual gaps in CCT use and parental time investments (around 15 and 11 percent, respectively) driven by initial conditions. Our model thus shows that even when households are ex-ante identical in terms of wages, productivity and preferences, parents still make different parenting choices depending on their child's initial stock of skills, and as a result experience different future outcomes.

7 Conclusion

We observe that investments of children in themselves, primarily by studying independently, increase as the child matures and gradually becomes more forward-looking, while the time children spend actively engaging with their parents decreases. We develop a model of the child's cognitive and non-cognitive growth in which both parents and children take an active role. Within our modeling framework, there exist three primary mechanisms that produce such a result. The first is associated with estimated increases in the productive value of the child's study time as they age. We estimate that the child's time inputs become more valuable than the parents' at later stages of the development period. A second mechanism is the increase in the child's valuation of future events, reflected in a increasing discount factor sequence over the development period, a phenomenon that has been documented in research conducted by developmental psychologists. As the child becomes more forward-looking, the incentive to invest in their own human capital increases. The third mechanism is related to the choice of parenting styles. By paying a short-run utility cost, parents can incentivize their child to invest the optimal time in self-investment (from the parents' perspective) by linking their investment time at age t to their consumption in the period.

Our estimates of the model parameters are sensible and fit the data reasonably well given the large number of moments utilized and the model's complexity. Allowing for endogeneity of the child's discount factor was a particularly important and novel feature, as was the use of data on the use of conditional allowances by parents as a proxy for CCTs. We found that higher educated parents were no more likely to use such incentives, and that the use of these incentives declined with the age of the child in the pre-teen years for all SES groups. Viewed through the lens of the model, this is mainly attributable to the cost of using CCTs in terms of the reduction in the child's future discount factor. Thus we found that the second and third mechanisms described above were quantitatively important in accounting for the timing and patterns of the use of CCTs by parents.

In a counterfactual exercise we examined the paths through which parental heterogeneity in education impacts the distribution of child outcomes, i.e. final-period skill and patience levels. The quantitatively most important paths were associated with the productivity of parental investment time and the impact of parental education on both their own as well as their children's discount factors. Even though the use of CCTs is more affordable to higher educated or wealthier households, the fact that the children in such households tend to be more forward-looking and thus have more intrinsic motivation to invest leads their parents to offer fewer extrinsic incentives. More generally, our model and empirical work demonstrates the need to examine cognitive and non-cognitive development jointly in order to understand child development. This paper constitutes a step in this direction.

And finally, our work suggests several areas of future work, in particular new data collections. The data we use on whether the receipt of a regular allowance tied to satisfactory school work (or school performance) is clearly an imperfect measure of the use of CCTs by parents. Additional measures of parental and child relationships, linked to important other household choices and demographics, would be valuable. So too would be additional data on child non-cognitive development, on discount rates and other attributes such as executive function, again linked to important household choices and demographics. Finally, interventions and experiments, in particular those that can be expected to alter child-parent interactions, can be an important source of data for the next generation of child development models.

References

[1] Agostinelli, F. and M. Wiswall, 2020. "Estimating the technology of children's skills formation", NBER Working Paper 2244.

[2] Arrondel, L., 2013. "Are "daddy's boys" just as rich as daddy? The transmission of values between generations, *Journal of Economic Inequality*, 11(4), 439-471.

[3] Bagger, J. F. Fontaine, F. Postel-Vinay, J.M. Robin, 2014. "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics", *American Economic Review*, 104, 6:1551-96.

[4] Barczyk, B. and M. Kredler, 2018. "Evaluating Long-Term Care Policy Options, Taking the Family Seriously", *Review of Economic Studies*, 85(2), 766-809.

[5] Becker, G. S., 1962. "Investment in Human Capital: A Theoretical Analysis", *Journal of Political Economy*, University of Chicago Press, vol. 70:9-29.

[6] Becker, G. S. and N. Tomes, 1976. "Child Endowments and the Quantity and Quality of Children", *Journal of Political Economy*, University of Chicago Press, vol. 84(4): S143-62, August.

[7] Ben-Porath, Y., 1967. "The Production of Human Capital and the Life Cycle of Earnings", *Journal of Political Economy*, 75(4):352-365.

[8] Boar, C., 2021. "Dynastic Precautionary Savings", *Review of Economic Studies*, 88(6), 2735-2765.

[9] Bonke, J., 2013. "Do Danish children and young people receive pocket money?", Study Papers 57, Rockwool Foundation Research Unit.

[10] Brenoe, A. and T. Epper, 2022. "Parenting values and the intergenerational transmission of time preferences", *European Economic Review*, 148, Pages 10408.

[11] Brown, H. and M. van der Pol, 2015. "Intergenerational transfer of time and risk preferences", *Journal of Economic Psychology*, 49, 187-204.

[12] Carneiro, P. and J.J. Heckman, 2003. "Human Capital Policy", in: J. Heckman and A. Krueger, Inequality in America: What Role for Human Capital Policy?, MIT Press.

[13] Castillo M., Ferraro P., Jordan, J. and R. Petrie, 2011. "The today and tomorrow of kids: Time preferences and educational outcomes of children", *Journal of Public Economics*, 2011, vol. 95, issue 11, pages 1377-1385.

[14] Castillo, M., Jordan, J. and R. Petrie, 2019. "Discount Rates of Children and High School Graduation", *The Economic Journal*, Royal Economic Society, vol. 129(619), pages 1153-1181.

[15] Cesarini, D., Dawes, C., Johannesson, M., Lichtenstein, P., and B. Wallace, 2009. "Genetic Variation in Preferences for Giving and Risk Taking", *Quarterly Journal of Economics*, 124(2), 809-842.

[16] Cooper, H., Lindsay, J. J., Nye, B., and S. Greathouse, 1998. "Relationships among attitudes about homework, amount of homework assigned and completed, and student achievement", *Journal of Educational Psychology*, 90(1), 70-83.

[17] Cooper, H., Robinson J. C. and E.A. Patall, 2006. "Does homework improve academic achievement? A synthesis of research, 1987–2003", *Review of Educational Research*, 76(1): 1–62.

[18] Cosconati M., 2009. "Parenting style and the development of human capital in children", Dissertation available from ProQuest, AAI3363272.

[19] 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.

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

[21] Currie, J., and D. Almond, 2011. "Human capital development before age five", in David Card and Orley Ashenfelter (eds.), *Handbook of Labor Economics*, vol. 4B, 1315-1486, Amsterdam: Elsevier.

[22] Daruich, D., 2018. "The Macroeconomic Consequences of Early Childhood Development Policies", FRB St.LOuis Working Paper Issue 2018-29

[23] Dauphin, A., El Lahga, A.R., Fortin, B., and G. Lacroix 2011. "Are Children Decision-Makers within the Household?", *Economic Journal*, 121:871-903.

[24] Deci, E., 1971."Effects of externally mediated rewards on intrinsic motivation", *Journal of Personality and Social Psychology*, 18(1), 105–115.

[25] Deci, E. L., Koestner, R. and R.M. Ryan, 1999. "A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation", *Psychological Bulletin*, 125(6), 627-668.

[26] Del Boca, D. and C. Flinn, 2011. "Endogenous household interaction", *Journal of Econometrics*, 166, issue 1, p. 49-65.

[27] Del Boca, D., Flinn, C., and M. Wiswall, 2014. "Household choice and child development", *Review of Economic Studies*, 81:137-185.

[28] Del Boca, D., Monfardini, C. and C. Nicoletti, 2017. "Parental and child time and cognitive development of adolescents", *Journal of Labor Economics,* 35(2):565-608.

[29] De Nardi, C. and G. Fella, 2017. "Saving and wealth inequality", *Review of Economic Dynamics*, 26, 280-300.

[30] Doepke, M. and Zilibotti, F., 2017. "Parenting with Style: Altruism and Paternalism in Intergenerational Preference Transmission", *Econometrica*, 85(5): 1331-1371.

[31] Doepke, M., Sorrenti G., and Zilibotti, F., 2019. "The Economics of Parenting," *Annual Review of Economics*, Vol. 11: 55-84.

[32] Dohmen, T., Falk, A., Huffman, D., and U. Sunde, 2012. "The Intergenerational Transmission of Risk and Trust Attitudes", *Review of Economic Studies*, 79(2), 645-677.

[33] Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and U. Sunde, 2018. "Global Evidence on Economic Preferences", *Quarterly Journal of Economics*, 133(4), 1645-1692.

[34] Fang, H. and D. Silverman, 2009. "Time-inconsistency and Welfare Program Participation: Evidence from the NLSY", *International Economic Review*, 50, 1043-1077.

[35] Fiorini, M., and M. Keane, 2014. "How the Allocation of Children's Time Affects Cognitive and Noncognitive Development", *Journal of Labor Economics*, 32(4), 787 - 836.

[36] Flinn, C.J., Gemici A., S. Laufer, 2017. "Search, Matching, and Training", *Review of Economic Dynamics*, 25, 260-297.

[37] Flinn, C. and J. Mullins, 2015. "Labor Market Search and Schooling Investment", *International Economic Review*, 56 (2), 359-398.

[38] Frederick, S., G. Loewenstein and T. O'Donoghue, 2002. "Time Discounting and Time Preference: A Critical Review", *Journal of Economic Literature*, 40(2): 351-401.

[39] Greene, D. and M. Lepper, 1974. "Effects of Extrinsic Rewards on Children's Subsequent Intrinsic Interest", *Child Development*, 45(4), 1141-1145.

[40] Hendricks, L., 2007. "How important is discount rate heterogeneity for wealth inequality?", *Journal of Economic Dynamics and Control*, 31(9), 3042-3068.

[41] Hofferth, S., Davis-Kean, P., Davis, J., and Finkelstein, J., 1998. "User guide for the Child Development Supplement of the Panel Study of Income Dynamics", http://psidonline.isr.umich.edu/CDS/usergd.html.

[42] Kaplan, G., 2012. "Moving Back Home: Insurance Against Labor Market Risk", *Journal of Political Economy*, vol. 120, no. 3.

[43] Knowles, J. and A. Postlewaite, 2005. "Do children learn to save from their parents?", Unpublished manuscript.

[44] Kooreman, P., 2007. "Time, money, peers, and parents; some data and theories on teenage behavior", *Journal of Population Economics*, 20(1): 9-33.

[45] Kosse, F. and F. Pfeiffer, 2012. "Impatience among preschool children and their mothers", *Economics Letters*, 115(3), 493-495.

[46] Krusell, P. and A.A. Smith, 1998. "Income and wealth heterogeneity in the macroeconomy", *Journal of Political Economy*, 106(5), 867-896.

[47] Lentz, R. and N. Roys, 2015. "Training and Search On the Job", NBER Working Papers 21702, National Bureau of Economic Research.

[48] Lino, M., Kuczynski, K., Rodriguez, N., and T. Schap, 2017. "Expenditures on Children by Families, 2015", Miscellaneous Publication No. 1528-2015. U.S. Department of Agriculture, Center for Nutrition Policy and Promotion.

[49] Lord, F. M. and M. Novick, 1968. “Statistical Theories of Mental Test Scores”, Reading MA: Addison-Welsley Publishing Company.

[50] Lundberg, S., J. Romich, and K. Tsang, 2009. “Decision-making by children”, *Review of Economics of the Household*, 7, 1-30.

[51] Macartney, H., 2016. “Incentives in Education and Teacher Effort”, mimeo.

[52] Mullins, J., 2022. “Designing Cash Transfers in the Presence of Children’s Human Capital Formation”, Working paper.

[53] Nishiyama, S., 2002. “Bequests, Inter Vivos Transfers, and Wealth Distribution”, *Review of Economic Dynamics*, 5(4), 892-931.

[54] Reynolds, B., Leraas, K., Collins, K., and S. Melanko, 2009. “Delay discounting by the children of smokers and nonsmokers”, *Drug and Alcohol Dependence*, 99, 350-353.

[55] Steinberg, L., Brian, A., Cauffman, E., Graham, S., Wooland, J. and A. Banich, 2009. “Age Differences in Future Orientation and Delay Discounting”, *Child Development*, 80:28-44.

[56] Verriest, E., 2022. “Household Labor Supply, Child Development and Childcare Policies in the United States”, Working paper.

[57] Wasmer, E. 2006. “General versus Specific Skills in Labor Markets with Search Frictions and Firing Costs”, *American Economic Review*, American Economic Association, vol. 96(3), pages 811-831.

[58] Webley, P. and E. Neyhus, 2006. “Parents’ influence on children’s future orientation and saving”, *Journal of Economic Psychology*, 27(1), 140–164.

[59] Woodcock, R., and M. Johnson, 1989. “Tests of Achievement, Standard Battery” [Form B], Chicago, IL: Riverside Publishing (Houghton Mifflin).

Appendices

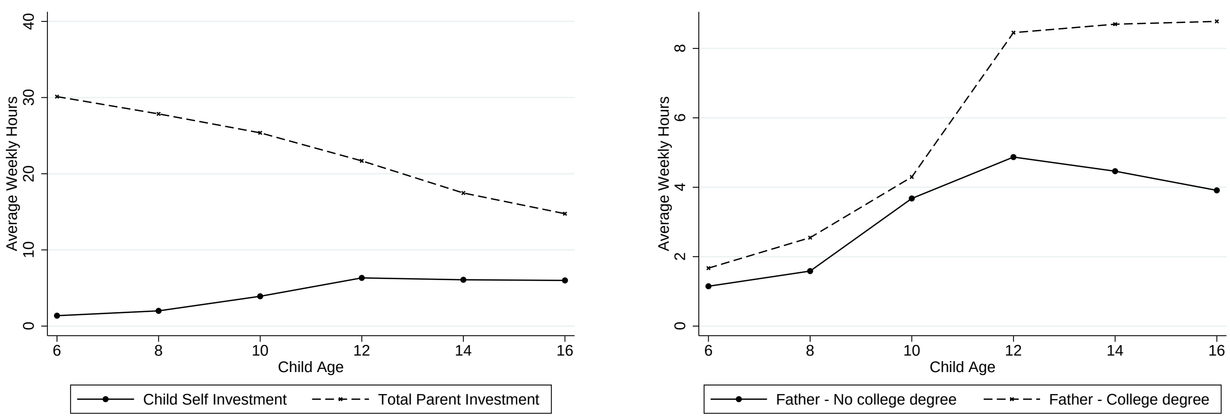
A Tables and Figures

A.1 Descriptive Statistics

Figure 1: Child Self-Investment Time by Child Age and Parental Characteristics

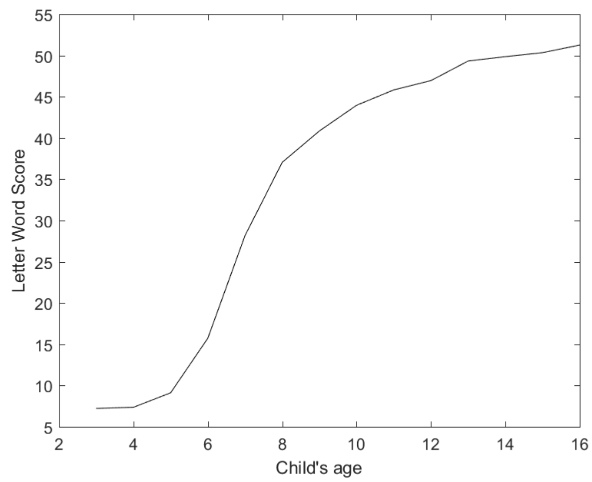
(a) Child and Parental Time Allocation, by Child Age

(b) Child Self-investment Time, by Father's Education



Notes: "Child self-investment time" is defined as the time spent on educational activities (homework, studying, reading, ...) by the child alone, with neither of the parents actively involved. "Total parent investment time" is defined as the total time spent in all parental "productive" activities where the child is actively engaged with at least one parent (including both educational and non-educational activities). All plotted data points are simple averages taken over two-year bins of child age (6-7, 8-9, etc.).
Source: PSID-CDS combined sample from 1997, 2002 and 2007 interviews and PSID core data between 1986 and 2010.

Figure 2: Average Child's Letter Word Score



Source: CDS combined sample from 1997, 2002 and 2007 interviews.

Table 1: Descriptive Statistics

PSID-CDS: 1997, 2002, 2007			
	Mean	**Std.**	**N**
Mother's age in 1997	33.31	8.32	247
Father's age in 1997	34.52	9.58	247
Mother's education	14.63	2.09	247
Father's education	14.19	2.31	247
Child's age in 1997	6.76	2.33	247
Letter Word raw score in 1997	23.37	15.49	247
Letter Word raw score in 2002	46.60	5.40	247
Letter Word raw score in 2007	50.44	3.42	109
PSID-Core: 1986 - 2010			
	Mean	**Std.**	**N**
Mother's work hours per week	23.29	16.45	2466
Father's work hours per week	42.64	11.52	2407
Mother's hourly wage	20.13	13.32	1852
Father's hourly wage	28.79	18.38	2306
Non-labor income per week	88.77	160.29	2356

Notes: Parental work hours, wages and non-labor income statistics are averaged over all years where the child is between 0 and 16 years old, ranging from 1986 to 2010.
Source: PSID-CDS combined sample from 1997, 2002 and 2007 interviews and PSID core data between 1986 and 2010.

Table 2: Parental Labor Supply by Child Age

	Fraction Working >0 Hours				
Child Age	**3**	**4-5**	**6-8**	**9-11**	**12-16**
Mothers working	0.796	0.766	0.787	0.816	0.849
Fathers working	0.990	0.985	0.982	0.985	0.959
Both working	0.781	0.746	0.758	0.794	0.805
	Avg. Hours Working (>0 Hours)				
Child Age	**3**	**4-5**	**6-8**	**9-11**	**12-16**
Mothers	27.77	28.87	28.97	29.99	31.75
Fathers	43.57	43.14	44.56	43.33	43.58

Notes: Upper half of the table shows labor force participation rates. Bottom half shows average labor hours conditional on working positive hours.
Source: PSID-CDS combined sample from 1997, 2002 and 2007 interviews and PSID core data between 1986 and 2010.

Table 3: Time Allocation by Child Age (Average Hours per Week)

Child Age	**3**	**4-5**	**6-8**	**9-11**	**12-16**
Mother's Work Hours	22.10	22.12	22.81	24.47	26.97
Father's Work Hours	43.13	42.49	43.78	42.70	41.79
Mother's Active Time	30.77	27.73	21.20	17.82	11.51
Father's Active Time	16.21	11.18	8.00	8.40	6.49
Child's Self-Investment Time	0.00	0.67	1.42	3.38	6.13
School Time	11.25	11.89	27.77	31.42	34.62

Notes: Parental work hours, wages and non-labor income statistics are averaged over all years where the child is between 0 and 16 years old, ranging from 1986 to 2010.
Source: PSID-CDS combined sample from 1997, 2002 and 2007 interviews and PSID core data between 1986 and 2010.

Table 4: (Un)conditional Allowances, by Child Age and Household Characteristics

			Fraction	
Child Age	**8-11**	**12-14**	**15-16**	**All ages**
(a) Does your child receive an allowance?				
All households	0.742	0.541	0.442	0.569
High School or less	0.750	0.649	0.444	0.610
(Some) college	0.750	0.522	0.413	0.553
Graduate	0.714	0.423	0.500	0.536
Sample size	93	109	104	306
(b) Is the allowance contingent on child doing his/her school work?				
All households	0.464	0.441	0.326	0.420
High School or less	0.625	0.583	0.313	0.531
(Some) college	0.433	0.292	0.368	0.370
Graduate	0.267	0.455	0.273	0.324
Sample size	69	59	46	174
(c) Joint probability of giving an allowance conditional on school work				
All households	0.344	0.239	0.144	0.239
High school or less	0.469	0.378	0.139	0.324
(Some) college	0.325	0.152	0.152	0.205
Graduate	0.190	0.192	0.136	0.174
Sample size	93	109	104	306

Notes: The table shows the fraction of households who answer "Yes" to the questions: (a) "Does your child receive an allowance?", and (b) "Is the allowance contingent on your child doing his/her school work?". Panel (c) shows the joint probabilities of using a conditional allowance, defined as the product of the corresponding fractions in panel (a) and (b). The various rows break down the responses by the father's education level (high school or less, some college or college degree, or graduate level). Note: Question (b) was not asked in 1997, and therefore all numbers only include data from 2002 and 2007, when all children were at least 8 years old.
Source: PSID-CDS combined sample from 2002 and 2007.

Table 5: Child Self-investment Time - OLS regressions

	(1) Study time	(2) Study time	(3) Study time
Father's years of schooling	0.386***	0.411***	0.558***
	(0.100)	(0.102)	(0.120)
Parental investment time		-0.0425***	-0.0549***
		(0.0135)	(0.0189)
Weekly allowance (dollars)			0.108*
			(0.0613)
Observations	572	572	404

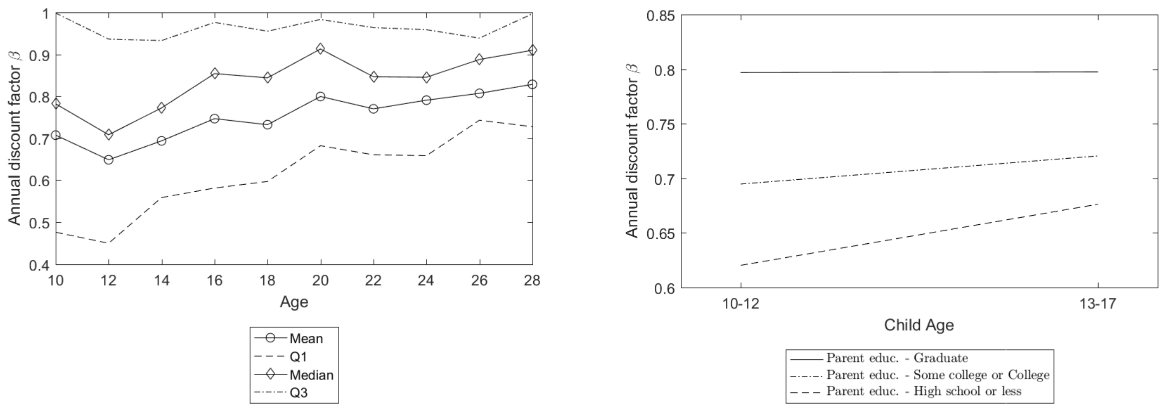
Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Each regression also contains a constant term, the child's age, and the child's age squared. Standard Errors are heteroskedasticity robust. The dependent variable in all columns, "Child self-investment time", is defined as the weekly time spent on educational activities (homework, studying, reading, ...) by the child alone, with neither of the parents actively involved. "Parental investment time" is defined as the total weekly time spent in all parental "productive" activities where the child is actively engaged with at least one parent (including both educational and non-educational activities). The difference in sample sizes between columns 1 and 2 and column 3 is because allowance information is not available for all households in our sample. If an allowance is not given, the value of weekly allowance is set to 0.
Source: PSID-CDS combined sample from 1997, 2002 and 2007 interviews and PSID core data between 1986 and 2010.

Figure 3: Empirical Distribution of Adolescents' Discount Factors

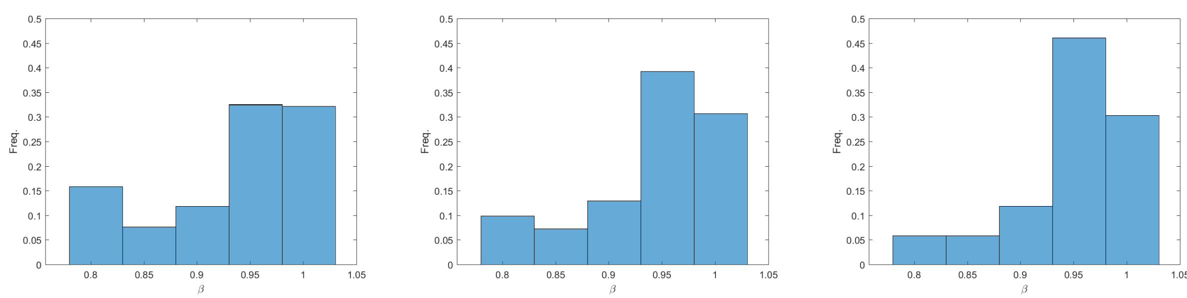
(a) Full sample, by Age **(b)** Adolescents, by Age and SES



Notes: Panel (a) shows summary statistics (average, median, 1st and 3rd quartile) of the empirical distribution of annual long-run discount factors, by age of the individual (full sample size $N = 904$). Panel (b) shows the average discount factor for the subset of adolescents between ages 10 and 17, divided by age category and by parental SES (sample size $N = 502$). **Source**: Experimental survey data from Steinberg et al. (2009).

Figure 4: Empirical Distribution of Adults' Discount Factors

(a) High school or less **(b)** (Some) College **(c)** Graduate



Notes: The graphs shows relative frequencies of adults' discount factors for various education groups. "High school or less" is defined as having either 11 or 12 years of schooling, "(Some) college" is defined as having either an associate's degree, a bachelor's degree, or some years of college but no degree, and "Graduate" is defined as having some postgraduate, a Master's or a doctoral degree. Sample sizes in each group are equal to 1005, 2677 and 943, respectively. **Source**: Osaka Preference Parameters Study, 2010 U.S. Survey. Only adults between ages 25 and 65 with known educational attainment and valid survey responses.

A.2 Parameter Estimates

Table 6: Preference Parameter Estimates

(a) Parental preferences	**Estimate**	**SE**
Mother's Leisure (α_1)	0.218	(0.00305)
Father's Leisure (α_2)	0.318	(0.00362)
Consumption (α_3)	0.311	(0.00564)
Child Quality (α_4)	0.153	(0.00068)
(b) Child preferences	**Estimate**	**SE**
Leisure (λ_1)	0.828	(0.01047)
Consumption (λ_2)	0.099	(0.00615)
Child Quality (λ_3)	0.073	(0.00438)
(c) Other preferences	**Estimate**	**SE**
Parental altruism (φ)	0.333	-
Mean/Std. of CCT Utility Cost (κ)	0.003	(0.00064)

Notes: SEs are standard errors computed using a cluster bootstrap sampling each household with replacement. Parameters without SE are assumed (not estimated) values.

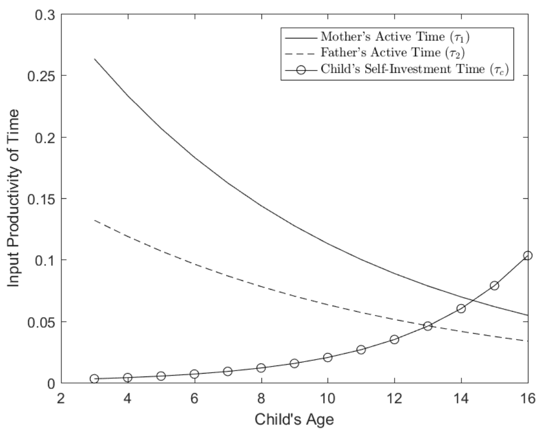
Table 7: Technology Parameter Estimates

		Estimate	SE
(a) Cognitive Human Capital			
Mother's Active Time (δ_1)	Intercept $d_{1,0}$	-1.408	(0.00779)
	Slope - Age $d_{1,1}$	-0.121	(0.00115)
	Slope - Educ. $d_{1,2}$	0.030	(0.00086)
Father's Active Time (δ_2)	Intercept $d_{2,0}$	-2.256	(0.01811)
	Slope - Age $d_{2,1}$	-0.105	(0.00153)
	Slope - Educ. $d_{2,2}$	0.038	(0.00250)
Child's Self-Investment Time (δ_3)	Intercept $d_{3,0}$	-6.598	(0.05655)
	Slope - Age $d_{3,1}$	0.271	(0.00394)
Child Expenditures (δ_4)	Intercept $d_{4,0}$	-7.154	(0.24507)
	Slope - Age $d_{4,1}$	0.072	(0.00283)
Last Period's Child Quality (δ_5)	Intercept $d_{5,0}$	-0.254	(0.00095)
	Slope - Age $d_{5,1}$	0.005	(0.00006)
Total Factor Productivity (R_t)	$d_{6,0}$	0.95594	(0.04368)
	$d_{6,1}$	2.51444	(0.00923)
	$d_{6,2}$	1.24575	(0.25336)
	$d_{6,3}$	5.28224	(0.06627)
(b) Child Discount Factors			
Upward transitions (D^{up})	Intercept b_0^{up}	-0.788	(0.02277)
	Slope - Age b_1^{up}	0.017	(0.00076)
	Slope - CCT b_2^{up}	-0.421	(0.03449)
	Slope - Interaction b_3^{up}	-0.037	(0.00318)
Downward transitions (D^{down})	Intercept b_0^{down}	-2.178	(0.07462)
	Slope - Age b_1^{down}	0.011	(0.00067)
	Slope - CCT b_2^{down}	0.421	-
	Slope - Interaction b_3^{down}	0.037	-
Initial conditions - $Pr(\beta_{c,t_0^h} = \beta_c^1 \| t_0^h, s^h)$	Intercept ν_0^1	10.831	(0.24501)
	Slope - Age ν_1^1	-0.278	(0.00816)
	Slope - Parent Educ. ν_2^1	-0.458	(0.01215)
Initial conditions - $Pr(\beta_{c,t_0^h} = \beta_c^2 \| t_0^h, s^h)$	Intercept ν_0^2	2.598	(0.13306)
	Slope - Age ν_1^2	-0.177	(0.01908)
	Slope - Parent Educ. ν_2^2	-0.239	(0.01520)

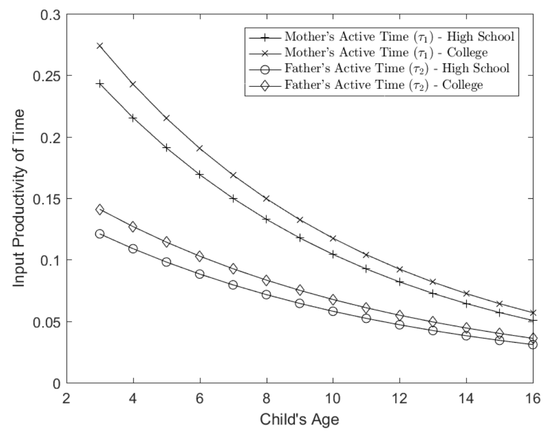
Notes: SEs are standard errors computed using a cluster bootstrap sampling each household with replacement.

Figure 5: Estimated Cognitive Human Capital Parameters by Child Age

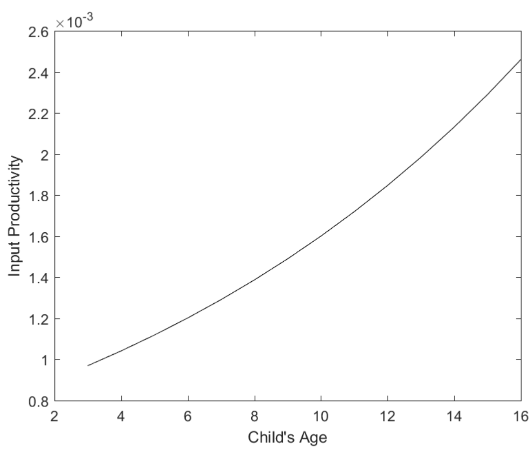
(a) All Time Inputs (avg.)



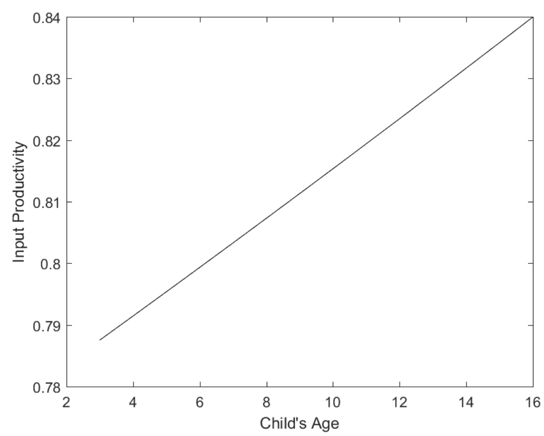
(b) Parental Time Inputs, by Educ.



(c) Child Goods



(d) Lagged Child Skills



(e) TFP

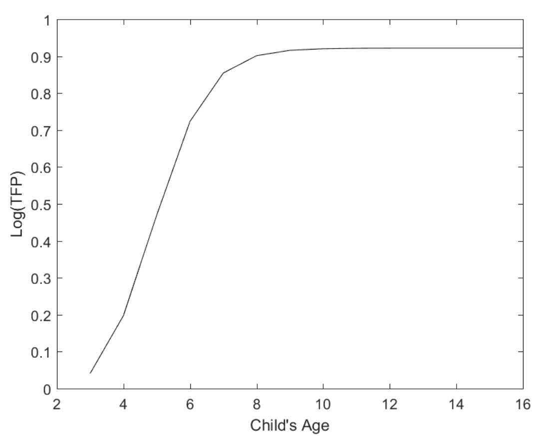
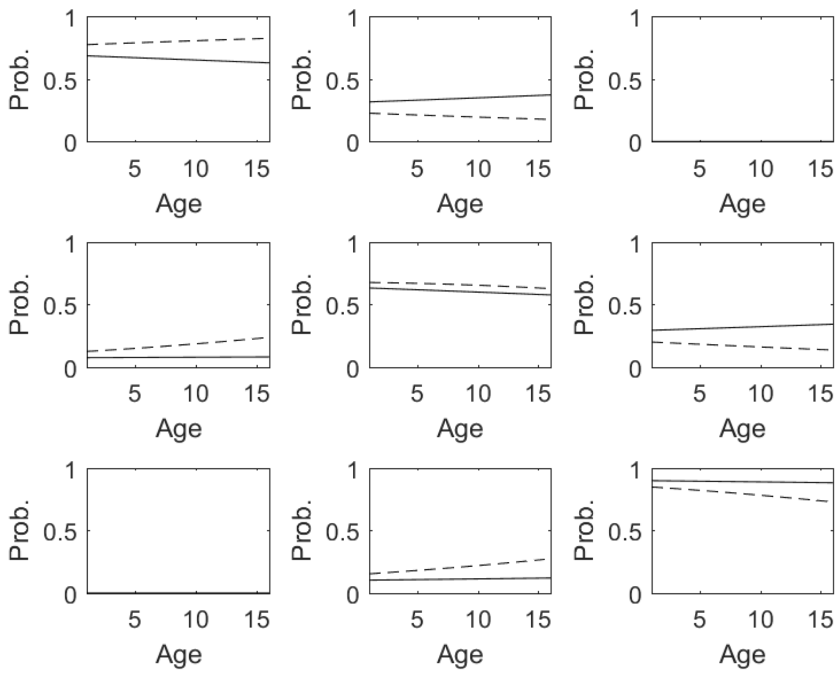
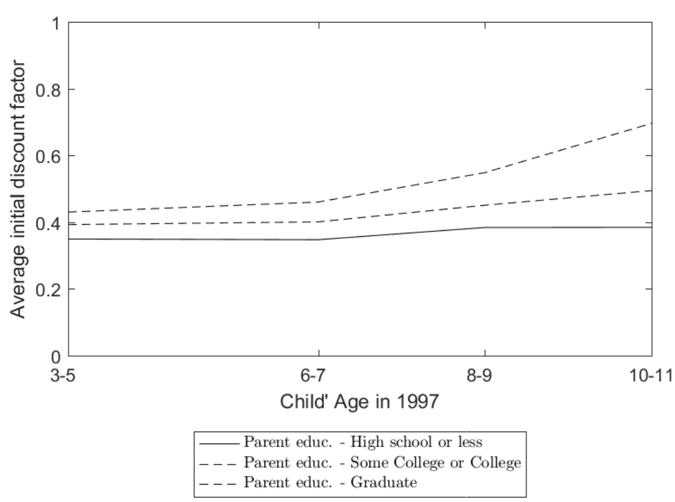


Figure 6: Estimated Child Discount Factor Parameters by Child Age

(a) Markov Transition Probabilities, CCT vs. No CCT



(b) Initial Conditions



Notes: Panel (a) shows Markov transition probabilities from each starting state $\beta_{c,t} \in \{Low, Middle, High\}$ (rows) to each future state $\beta_{c,t+1} \in \{Low, Middle, High\}$ (columns), by age and CCT use. Solid lines correspond to not using a CCT, and dashed lines correspond to using a CCT. Panel (b) shows the average of the child's initial discount factor conditional on the child's age in 1997 (when simulations start) and the father's schooling.

Table 8: Wage and Income Parameter Estimates

(a) Mother's Log Wage Offer	**Estimate**	**SE**
Intercept ($\eta_{0,1}$)	0.941	(0.01442)
Mother's Age ($\eta_{1,1}$)	0.012	(0.00023)
Mother's Education ($\eta_{2,1}$)	0.088	(0.00057)
Standard Deviation of Shock (σ_{w_1})	0.439	(0.01392)
Correlation with Father's Wage Shock (ρ_{12})	0.687	(0.01941)
(b) Father's Log Wage Offer	**Estimate**	**SE**
Intercept ($\eta_{0,2}$)	1.678	(0.01203)
Father's Age ($\eta_{1,2}$)	0.012	(0.00026)
Father's Education ($\eta_{2,2}$)	0.077	(0.00087)
Standard Deviation of Shock (σ_{w_2})	0.650	(0.02230)
(c) Latent Non-Labor Income	**Estimate**	**SE**
Logit - Intercept (μ_1)	-5.012	(0.11379)
Logit - Mother's Age (μ_2)	0.000	(0.00009)
Logit - Mother's Age Squared (μ_3)	0.000	(0.00007)
Logit - Mother's Education (μ_4)	0.000	(0.00002)
Logit - Father's Age (μ_5)	0.115	(0.00337)
Logit - Father's Age Squared (μ_6)	-0.001	(0.00010)
Logit - Father's Education (μ_7)	0.237	(0.00673)
Conditional - Intercept (μ_8)	2.205	(0.15222)
Conditional - Mother's Age (μ_9)	0.000	(0.00004)
Conditional - Mother's Age Squared (μ_{10})	-0.000	(0.00004)
Conditional - Mother's Education (μ_{11})	0.003	(0.00849)
Conditional - Father's Age (μ_{12})	0.045	(0.00413)
Conditional - Father's Age Squared (μ_{13})	-0.000	(0.00005)
Conditional - Father's Education (μ_{14})	-0.001	(0.00073)
Standard Deviation of Shock (σ_I)	1.299	(0.03250)

Notes: SEs are standard errors computed using a cluster bootstrap sampling each household with replacement.

A.3 Model Fit

Table 9: Sample Fit of Time Allocations by Child Age

(a) Probability Work > 0 Hours

	Mother		Father	
Child Age	**Data**	**Simulated**	**Data**	**Simulated**
3-5	0.777	0.751	0.987	0.988
6-8	0.787	0.798	0.982	0.993
9-12	0.811	0.819	0.984	0.993
13-16	0.863	0.840	0.954	0.995

(b) Hours Worked if Work > 0 Hours (Avg.)

	Mother		Father	
Child Age	**Data**	**Simulated**	**Data**	**Simulated**
3-5	28.45	27.95	43.30	42.85
6-8	28.97	28.70	44.56	44.08
9-12	29.92	29.75	43.66	43.71
13-16	32.22	31.22	43.29	44.35

(c) Active Time with Child (Avg.)

	Mother		Father	
Child Age	**Data**	**Simulated**	**Data**	**Simulated**
3-5	28.05	29.24	11.71	12.34
6-8	21.20	21.55	8.00	10.06
9-12	17.11	19.89	8.56	9.85
13-16	11.51	15.46	6.12	8.45

(d) Child Self-Investment Time (Avg.)

Child Age	**Data**	**Simulated**
3-5	0.60	0.73
6-8	1.42	1.41
9-12	4.01	3.04
13-16	5.81	10.72

Notes: Data is actual data. Simulated is the model prediction at estimated parameters.
Source: PSID-CDS combined sample from 1997, 2002 and 2007 interviews and PSID core data between 1986 and 2010.

Table 10: Sample Fit of Wages and Non-Labor Income

(a) Hourly Wages	**Mother**		**Father**	
	Data	**Simulated**	**Data**	**Simulated**
Average	20.13	20.07	28.79	29.50
Standard deviation	13.32	13.74	18.38	16.22
(b) Weekly Non-labor Income				
	Data	**Simulated**		
Average	89.12	87.05		
Standard deviation	160.98	163.39		
Fraction with $I > 0$	0.74	0.73		

Notes: Data is actual data. Simulated is the model prediction at estimated parameters.
Source: PSID-CDS combined sample from 1997, 2002 and 2007 interviews and PSID core data between 1986 and 2010.

Table 11: Sample Fit of Binary CCT Choice

	Data	**Simulated**
Average CCT use	0.239	0.353
Correlation with child's age	-0.197	-0.226
Correlation with mother's years of schooling	-0.208	-0.047
Correlation with father's years of schooling	-0.173	-0.079
Correlation with child's Letter Word score	-0.189	-0.206
Correlation with mother's weekly time investment	0.061	0.086
Correlation with father's weekly time investment	0.023	0.042

Notes: Data is actual data. Simulated is the model prediction at estimated parameters. Since actual data on CCT use is only available for children between ages 8 and 16 (in the CDS-II and CDS-III waves), we report all simulated moments for the same age range. All correlations are contemporaneous.
Source: PSID-CDS combined sample from 1997, 2002 and 2007 interviews and PSID core data between 1986 and 2010.

Table 12: Discount Factors of Parents and Children - Data vs. Model Simulation

(a) Parents/Adults	**Data (Osaka-PPS)**		**Model Sim.**	
	Mean	**St.Dev.**	**Mean**	**St.Dev.**
All individuals	0.936	0.064	0.936	0.061
High school or less	0.928	0.071	0.930	0.063
Some college or College	0.935	0.064	0.936	0.061
Graduate	0.945	0.057	0.943	0.057
(b) Children	**Data (Steinberg et al.)**		**Model Sim.**	
	Mean	**St.Dev.**	**Mean**	**St.Dev.**
Ages 3-5	-	-	0.422	0.192
High school or less	-	-	0.381	0.127
Some college or College	-	-	0.431	0.202
Graduate	-	-	0.456	0.225
Ages 6-9	-	-	0.555	0.263
High school or less	-	-	0.524	0.247
Some college or College	-	-	0.551	0.263
Graduate	-	-	0.605	0.275
Ages 10-12	0.680	0.310	0.666	0.281
High school or less	0.620	0.325	0.619	0.277
Some college or College	0.695	0.313	0.673	0.281
Graduate	0.797	0.211	0.726	0.274
Ages 13-17	0.710	0.283	0.766	0.261
High school or less	0.676	0.293	0.739	0.267
Some college or College	0.720	0.279	0.775	0.256
Graduate	0.797	0.244	0.790	0.254

Notes: Panel (a) shows the mean and standard deviation of adults' annual discount factors, conditional on their own educational attainment. Panel (b) shows the mean and standard deviation of children's annual discount factors, conditional on their age and their parents' educational attainment. The data for adults stems from the 2010 U.S. survey of the Preference Parameters Study of Osaka University, for the subsample of adults between ages 25-65 (N = 4625). The data for children stems from the experimental surveys collected by Steinberg et al. (2009), where we restrict attention to children between ages 10-17 (N = 344). The first two columns show data moments, whereas the last two columns show summary statistics of the simulated (imputed) annual discount factors for our sample of PSID-CDS parents and children, respectively.

A.4 Comparative Statics

Table 13: Comparative Statics - CCT Parameters

	Baseline	**(1)**	**(2)**	**(3)**	**(4)**
Model specifications					
- Util. Cost Mean/St.Dev. κ	0.003	$+\infty$	0	0.003	0
- Disc. factor process - CCT slope b_2^{up}	-0.421	-0.421	-0.421	0	0
- Disc. factor process - $CCT \times age$ slope b_3^{up}	-0.037	-0.037	-0.037	0	0
Simulated Outcomes					
Final Child Quality $\log(k_{17})$	7.29	7.12	7.28	7.62	7.63
Final Discount Factor $\beta_{c,17}$	0.81	0.88	0.80	0.88	0.88
Parents' Bellman value V_p, ages 3-16	71.95	70.93	71.94	73.73	73.76
Child's Bellman value V_c, ages 3-16	34.21	37.13	34.09	35.93	35.93
Other Simulated Moments					
Fraction using CCT, ages 8-16	0.35	0.00	0.37	0.90	1.00
Parental Investment Time $\tau_{p,t}$, ages 8-16	27.15	27.36	27.13	27.36	27.36
Self-investment Time $\tau_{c,t}$, ages 8-16	6.43	3.61	6.45	12.00	12.16
Corr. (CCT_t, age_t), ages 8-16	-0.23	-	-0.25	0.04	-
Corr. $(CCT_t, educ_2)$, ages 8-16	-0.08	-	-0.09	0.03	-
Corr. (CCT_t, LW_t), ages 8-16	-0.21	-	-0.24	0.20	-

Notes: "Baseline" corresponds to the baseline model. Column (1) corresponds to the alternative model where CCTs are infinitely costly (i.e., $\kappa = \infty$). Column (2) considers the case where CCTs are costless (i.e., $\kappa = 0$), but still endogenously affect the child's future discount factor. Column (3) considers the case where CCTs are costly but no longer affect the child's future discount factor (i.e., $b_2^{up} = b_3^{up} = 0$). Column (4) considers the case where CCTs are costless and no longer affect the child's future discount factor (i.e., $\kappa = b_2^{up} = b_3^{up} = 0$). All experiments were done using $R = 10$ simulated data sets.

Table 14: Comparative Statics - Heterogeneity based on Demographics

	Baseline	**(1)**	**(2)**	**(3)**	**(4)**
Model specifications					
- Wages/NLI = f(age)	Yes	No	No	No	No
- Wages/NLI = f(educ.)	Yes	Yes	No	No	No
- Time prod. = f(educ.)	Yes	Yes	Yes	No	No
- Discount factor distr. = f(educ.)	Yes	Yes	Yes	Yes	No
- Fix child's initial test score	Yes	Yes	Yes	Yes	Yes
Simulated Outcomes					
Final Quality $\log(k_{17})$, Low SES	7.155	7.154	7.157	7.244	7.266
Final Quality $\log(k_{17})$, High SES	7.416	7.416	7.414	7.345	7.308
Gap relative to baseline (%)	100.0	100.3	98.7	38.6	15.9
Final Discount Factor $\beta_{c,17}$, Low SES	0.794	0.794	0.794	0.794	0.804
Final Discount Factor $\beta_{c,17}$, High SES	0.818	0.818	0.818	0.818	0.809
Gap relative to baseline (%)	100.0	100.0	100.0	98.0	20.9
Household Income, Low SES	1603.9	1596.2	1890.0	1879.9	1876.4
Household Income, High SES	2320.2	2192.8	1827.1	1838.5	1850.6
Gap relative to baseline (%)	100.0	83.3	-8.8	-5.8	-3.6
Mother's hourly wage, Low SES	16.719	16.595	19.128	19.130	19.139
Mother's hourly wage, High SES	23.362	22.100	19.282	19.322	19.288
Gap relative to baseline (%)	100.0	82.9	2.3	2.9	2.3
Father's hourly wage, Low SES	23.107	23.033	27.617	27.617	27.617
Father's hourly wage, High SES	36.100	34.324	27.294	27.290	27.293
Gap relative to baseline (%)	100.0	86.9	-2.5	-2.5	-2.5
Fraction using CCT, Low SES	0.439	0.439	0.439	0.437	0.410
Fraction using CCT, High SES	0.349	0.349	0.349	0.349	0.396
Gap relative to baseline (%)	100.0	100.0	100.0	97.5	15.1
Parental investment time (τ_p), Low SES	27.435	27.453	27.504	28.632	29.053
Parental investment time (τ_p), High SES	32.517	32.520	32.382	31.355	29.604
Gap relative to baseline (%)	100.0	99.7	96.0	53.6	10.8

Notes: "Baseline" corresponds to the baseline model. Column (1) corresponds to the alternative model where parental wage offers and non-labor income (NLI) no longer depend on parental age. Column (2) considers the case where parental wages and NLI no longer depend on parental age or education levels. Column (3) considers the case where parental wages, NLI and investment time productivities no longer depend on parental characteristics. Column (4) considers the case where parental wages, NLI, investment time productivities, and parent/child (initial) discount factor distributions no longer depend on parental characteristics. "Low SES" considers simulated households where the father has at most a high school degree, whereas "High SES" looks at households where the father has a graduate degree. All experiments were done using $R = 10$ simulated data sets.

B Model Solution

For both the parents and the child, the decision rules are solved using backward recursion beginning from the end of the development process, time M. Recall that at the start of period $t = 0, ..., M$, the state vector is given by $\Gamma_t = (w_{1,t}, w_{2,t}, I_t, k_t, \beta_{c,t})$, which includes hourly wage offers, non-labor income, and the child's initial stocks of cognitive and non-cognitive skills. The parents' choice set is given by $\mathbf{a}_{p,t} = (h_{1,t}, h_{2,t}, l_{1,t}, l_{2,t}, \tau_{1,t}, \tau_{2,t}, c_t, x_t, e_t, CCT_t, r_t, b_t\}$, which includes labor supply, consumption and investment choices, and CCT parameters. As a Stackelberg follower, the child only chooses how to allocate their residual time budget between self-investment time, $\tau_{c,t}$, and leisure time, $l_{c,t}$, conditional on observing the parents' actions $\mathbf{a}_{p,t}$, and the reward function given by equation (5).

B.1 The final period problem

B.1.1 The child's problem.

The child solves the following problem in period $t = M$:

$$\begin{aligned} V_{c,M}(\Gamma_M|\mathbf{a}_{p,M}) \quad = \quad & \max_{\tau_{c,M}|\mathbf{a}_{p,M},\mathcal{C}_{c,M}} u_c(l_{c,M}, x_M, k_M) \\ & +\beta_{c,M}\mathbb{E}_M\left(V_{c,M+1}(k_{M+1}, \beta_{c,M+1})|\tau_{c,M}, \mathbf{a}_{p,M}, \Gamma_M\right) \end{aligned}$$

subject to the child's time constraint, the deterministic law of motion for k_{M+1} given by (4), and the stochastic Markov process for $\beta_{c,M+1}$ defined in (7). Given the child's final-period welfare function defined in (2), the conditional expectation term can be rewritten as

$$\begin{aligned} & \sum_{j=1}^{Z} Pr(\beta_{c,M+1} = \beta_c^j|\beta_{c,M}, M, CCT_M)\frac{\lambda_3 \ln k_{M+1}}{1-\beta_c^j} \\ \equiv \quad & \psi_{c,M+1}(CCT_M, \beta_{c,M}) \ln k_{M+1} \end{aligned}$$

where $\psi_{c,M+1}(\cdot)$ denotes the child's expected marginal utility of future human capital:

$$\begin{aligned} \psi_{c,M+1}(CCT_M, \beta_{c,M}) \quad & = \quad \frac{\partial \mathbb{E}_M(V_{c,M+1})}{\partial \ln k_{M+1}} \\ & = \quad \lambda_3 \sum_{j=1}^{Z} Pr(\beta_{c,M+1} = \beta_c^j|\beta_{c,M}, CCT_M, M)\frac{1}{1-\beta_c^j} \end{aligned} \tag{13}$$

When choosing their optimal study time, the child takes their current discount factor and their parents' actions (including CCT_M) as given. Taking the first order condition of $V_{c,M}(\cdot)$ with respect to $\tau_{c,M}$ yields the following reaction function:

$$\tau^*_{c,M}(\mathbf{a}_{p,M}, \beta_{c,M}) \quad = \quad \gamma_M(r_M, CCT_M, \beta_{c,M})(\tilde{T}_t - \tau_{p,M}) \tag{14}$$

$$\text{where} \quad \gamma_M(r_M, CCT_M, \beta_{c,M}) \equiv \frac{\lambda_2 r_M + \Delta_{c,M}(CCT_M, \beta_{c,M})}{\lambda_1 + \lambda_2 r_M + \Delta_{c,M}(CCT_M, \beta_{c,M})}$$
$$\text{and} \quad \Delta_{c,M}(CCT_M, \beta_{c,M}) \equiv \beta_{c,M}\psi_{c,M+1}(CCT_M, \beta_{c,M})\delta_{3,M}$$

where $\gamma_M(\cdot) \in (0,1)$ denotes the fraction of the child's residual time budget that they choose to spend on self-investment time (as opposed to leisure). Given the properties of the production, utility and reward functions, the choice of self-investment time is independent of all of the parents' decisions, with the exception of (1) the total time they spend interacting with the children, $\tau_{p,M}$, the effect of which is to reduce (or "crowd out") the child's effective time endowment; (2) the use of a CCT, which may alter the child's future discount factor, and therefore their expected marginal value of human capital, $\psi_{c,M+1}(\cdot)$; and (3) the child's "wage" rate r_M, which corresponds to the elasticity of child consumption with respect to child study time, as defined in (5). The fact that the intercept term b_M (i.e., the "consumption floor" parameter in the parents' reward function) drops out will prove useful in deriving some of the results below. Note that when the parents do not use a CCT (i.e., $r_M = 0$), this solution simplifies to the special case in which the parents make a fixed transfer of x_M to the child that is not tied to the child's investment time. Clearly, the solution to the child's problem is increasing in r_M and the child can be induced to spend virtually all of its residual time in investment as r_M becomes arbitrarily large.[55]

For future reference, we introduce the following notation for each case, $CCT_t \in \{0,1\}$:

$$\psi_{c,M+1}(CCT_M, \beta_{c,M}) = \begin{cases} \psi^0_{c,M+1}(\beta_{c,M}) & \text{if } CCT_M = 0 \\ \psi^1_{c,M+1}(\beta_{c,M}) & \text{if } CCT_M = 1 \end{cases} \tag{15}$$

$$\Delta_{c,M}(CCT_M, \beta_{c,M}) = \begin{cases} \Delta^0_{c,M}(\beta_{c,M}) = \beta_{c,M}\psi^0_{c,M+1}(\beta_{c,M})\delta_{3,M} & \text{if } CCT_M = 0 \\ \Delta^1_{c,M}(\beta_{c,M}) = \beta_{c,M}\psi^1_{c,M+1}(\beta_{c,M})\delta_{3,M} & \text{if } CCT_M = 1 \end{cases} \tag{16}$$

$$\gamma_M(CCT_M, \beta_{c,M}) = \begin{cases} \gamma^0_{c,M}(\beta_{c,M}) \equiv \frac{\Delta^0_{c,M}(\beta_{c,M})}{\lambda_1 + \Delta^0_{c,M}(\beta_{c,M})} & \text{if } CCT_M = 0 \\ \gamma^1_M(r_M, \beta_{c,M}) \equiv \frac{\lambda_2 r_M + \Delta^1_{c,M}(\beta_{c,M})}{\lambda_1 + \lambda_2 r_M + \Delta^1_{c,M}(\beta_{c,M})} & \text{if } CCT_M = 1 \end{cases} \tag{17}$$

where the $\psi_{c,M+1}$ terms denote the child's marginal valuation of future (cognitive) human capital depending on the parents' binary CCT choice, the $\Delta_{c,M}$ terms denote the child's (discounted) future marginal benefit of self-investing today, and the γ_M terms denote the fraction spent by the child in self-investment for each CCT case. To the extent that using a CCT decreases future patience capital in a probabilistic sense (i.e., $\psi^1_{c,M+1} < \psi^0_{c,M+1})$), then the "total" effect of the CCT on γ_t is ambiguous; choosing $r_t > 0$ works to increase the child's self-investment time whereas the reduction in future patience decreases it.

[55]Of course, this could never be optimal, since the parents would not want their child to have zero leisure as long as they are altruistic ($\varphi > 0$).

B.1.2 The parents' problem.

Given the child's reaction function, the parents solve the following problem:

$$V_{p,M}(\Gamma_M) = \max_{\mathbf{a}_{p,M}|\tau^*_{c,M}(\mathbf{a}_{p,M}),\mathcal{C}_{p,M}} \tilde{u}_p(\mathbf{a}_{p,M}) + \beta_p \mathbb{E}_M\big(V_{p,M+1}(\Gamma_{M+1}|\mathbf{a}_{p,M},\Gamma_M\big)$$

subject to the usual time and budget constraints, the laws of motion for k_{M+1} and $\beta_{c,M+1}$, and the CCT reward function (5), and with $\tilde{u}_p(\cdot)$ as defined in (6). Given the parents' final-period welfare function defined in (3), the expectation term can be rewritten as the sum of a parental component and a child component, weighted by the altruism parameter, φ:

$$\begin{aligned}
\mathbb{E}_M\big(V_{p,M+1}(k_{M+1},\beta_{c,M+1})|\mathbf{a}_{p,M},\Gamma_M\big) &= \sum_{j=1}^{Z} Pr(\beta_{c,M+1}=\beta_c^j|\beta_{c,M},CCT_M,M) \\
&\quad \times\Big(\frac{(1-\varphi)\alpha_4}{1-\beta_p}+\frac{\varphi\lambda_3}{1-\beta_c^j}\Big)\ln k_{M+1} \\
&= \big[(1-\varphi)\frac{\alpha_4}{1-\beta_p}+\varphi\psi_{c,M+1}(CCT_M,\beta_{c,M})\big]\ln k_{M+1} \\
&\equiv \psi_{p,M+1}(CCT_M,\beta_{c,M})\ln k_{M+1}
\end{aligned}$$

where $\psi_{p,M+1}(\cdot)$ denotes the parents' expected marginal value of future child human capital:

$$\begin{aligned}
\psi_{p,M+1}(CCT_M,\beta_{c,M}) &= \frac{\partial \mathbb{E}_M(V_{p,M+1})}{\partial \ln k_{M+1}} \\
&= \big[(1-\varphi)\frac{\alpha_4}{1-\beta_p}+\varphi\psi_{c,M+1}(CCT_M,\beta_{c,M})\big]
\end{aligned} \tag{18}$$

where $\psi_{c,M+1}(\cdot)$ is the child's marginal valuation, as defined in (13). For future reference, we introduce the following shorthand notation for each discrete CCT case:

$$\psi_{p,M+1}(CCT_M,\beta_{c,M}) = \begin{cases}\psi^0_{p,M+1}(\beta_{c,M}) & \text{if } CCT_M=0\\ \psi^1_{p,M+1}(\beta_{c,M}) & \text{if } CCT_M=1\end{cases} \tag{19}$$

$$\Delta_{p,M}(CCT_M,\beta_{c,M}) = \begin{cases}\Delta^0_{p,M}(\beta_{c,M})=\beta_p\psi^0_{p,M+1}(\beta_{c,M})\delta_{3,M} & \text{if } CCT_M=0\\ \Delta^1_{p,M}(\beta_{c,M})=\beta_p\psi^1_{p,M+1}(\beta_{c,M})\delta_{3,M} & \text{if } CCT_M=1\end{cases} \tag{20}$$

where the $\Delta_{p,M}$ terms denote the parents' (discounted) future marginal benefit of current child investment time.

Conditional closed form solutions. Given the log-additively separable model structure, we can derive closed form solutions for all parental choice variables, except for the binary CCT choice and the two parental time inputs. To solve the parents' problem, we can substitute out (1) c_M for the period M budget constraint, (2) $l_{1,M}$, $l_{2,M}$ and $l_{c,M}$ for the individual time constraints, (3) $\ln k_{M+1}$ for the production technology, and (4) $\tau_{c,M}$ for the child's optimal reaction function derived in the previous paragraph. In order to simplify the first order conditions with respect to the remaining continuous choices $\{h_{1,t}, h_{2,t},\ \tau_{1,t}, \tau_{2,t},\ e_t, x_t, r_t, b_t\}$, note that the parents jointly choose the triple $\{x_M, r_M, b_M\}$ subject to the reward function specified in (5) and the child's reaction function (14), with $r_t = 0$ if $CCT_t = 0$. For the case where $CCT_M = 1$, we can plug in equation (14) and rearrange terms to find the intercept:

$$\begin{aligned} b_M &= \ln(x_M) - r_M \ln(\tau^*_{c,M}(\mathbf{a}_{p,M}, \beta_{c,M})) \\ &= \ln(x_M) - r_M \ln(\gamma^1_M(r_M, \beta_{c,M})) - r_M \ln(\tilde{T}_M - \tau_{p,M}) \end{aligned}$$

where $\gamma^1_M(\cdot)$ is defined in (17), and $\tau_{p,M} = \tau_{1,M} + \tau_{2,M}$. Conditional on $\{x_M, r_M, \tau_{1,M}, \tau_{2,M}\}$, this pins down the optimal choice of b_M. Taking first order conditions with respect to $\{e_M, x_M\}$ and using the budget constraint yields the following solutions for consumption and expenditures:

$$c^\star_M = \frac{\tilde{\alpha}_3}{\tilde{\alpha}_3 + \tilde{\alpha}_6 + \beta_p \psi_{p,M+1}(CCT_M, \beta_{c,M})\delta_{4,M}} Y_M \tag{21}$$

$$x^\star_M = \frac{\tilde{\alpha}_6}{\tilde{\alpha}_3 + \tilde{\alpha}_6 + \beta_p \psi_{p,M+1}(CCT_M, \beta_{c,M})\delta_{4,M}} Y_M \tag{22}$$

$$e^\star_M = \frac{\beta_p \psi_{p,M+1}(CCT_M, \beta_{c,M})\delta_{4,M}}{\tilde{\alpha}_3 + \tilde{\alpha}_6 + \beta_p \psi_{p,M+1}(CCT_M, \beta_{c,M})\delta_{4,M}} Y_M \tag{23}$$

where $Y_M = w_{1,M}h_{1,M} + w_{2,M}h_{2,M} + I_M$ denotes total household income. After substituting out these choices and the child's reaction function in the parents' value function, we can solve for the remaining choices, $(\tau_{1,M},\ \tau_{2,M},\ h_{1,M},\ h_{2,M},\ CCT_M,\ r_M)$. The log-additively separable model structure allows us to derive semi-closed form solutions for labor and leisure, conditional on $(\tau_{1,M}, \tau_{2,M}, CCT_M, \beta_{c,M})$. Assuming interior solutions for labor, we obtain:[56]

$$\begin{aligned} l^*_{1,M} &= \frac{\tilde{\alpha}_1}{\chi_M}\frac{TY_M}{w_{1,M}}, \quad h^*_{1,M} = T - l^*_{1,M} - \tau_{1,M}, \\ l^*_{2,M} &= \frac{\tilde{\alpha}_2}{\chi_M}\frac{TY_M}{w_{2,M}}, \quad h^*_{2,M} = T - l^*_{2,M} - \tau_{2,M}, \\ \text{where } \chi_M &\equiv \tilde{\alpha}_1 + \tilde{\alpha}_2 + \tilde{\alpha}_3 + \tilde{\alpha}_6 + \beta_p \psi_{p,M+1}(CCT_M, \beta_{c,M})\delta_{4,M} \end{aligned}$$

and where TY_M denotes total "potential" period-M income conditional on time investment choices:

$$TY_M \equiv w_{1,M}(T - \tau_{1,M}) + w_{2,M}(T - \tau_{2,M}) + I_M.$$

[56] If one or both of the resulting labor choices is negative, the corresponding corner solutions are analogous and straightforward.

We can derive total disposable household income by using the budget constraint:

$$\begin{aligned} Y_M &= w_{1,M} h^*_{1,M} + w_{2,M} h^*_{2,M} + I_M \\ &= \frac{\chi_M - \tilde{\alpha}_1 - \tilde{\alpha}_2}{\chi_M} TY_M \end{aligned}$$

Then, we can rewrite the budget allocation choices (21)-(23) as a function of total potential income:

$$c^\star_M = \frac{\tilde{\alpha}_3}{\chi_M} TY_M, \quad x^\star_M = \frac{\tilde{\alpha}_6}{\chi_M} TY_M, \quad e^\star_M = \frac{\beta_p \psi_{p,M+1}(CCT_M, \beta_{c,M}) \delta_{4,M}}{\chi_M} TY_M,$$

The remaining choice variables are $(\tau_{1,M}, \tau_{2,M}, r_M, CCT_M)$, where $CCT_M \in \{0, 1\}$ is binary.

Conditional on the parents not using a CCT, we denote the agents' optimal choice vector as $a^0_{p,M}$. When $CCT_M = 0$, the parents are constrained to choose $r^0_M = 0$, and the choice of $b^0_M = \ln(x^0_M)$ becomes redundant once the child's consumption level is determined by equation (22). Conversely, conditional on the parents using a CCT ($CCT_M = 1$), we can denote the agents' optimal choices as $a^1_{p,M}$. In this case, the parents incur a fixed utility cost ζ, and can freely choose the reward elasticity, r_M.

Conditional on $\beta_{c,M}$ and a binary choice $CCT_M = j$ ($j \in \{0, 1\}$), we know the scalar values of $\psi^j_{c,M+1}(\beta_{c,M})$, $\psi^j_{p,M+1}(\beta_{c,M})$, and $\Delta^j_{c,M}(\beta_{c,M})$ as defined above. For the case where $CCT_M = 1$, we can take the first order condition with respect to the reward elasticity, r_M, and solve it in closed form:

$$r^*_M(\beta_{c,M}) = \frac{\Delta^1_{p,M}(\beta_{c,M}) - \varphi \Delta^1_{c,M}(\beta_{c,M})}{\varphi \lambda_2}$$

where $\Delta^1_{c,M}(\cdot)$ and $\Delta^1_{p,M}(\cdot)$ denote each agent's marginal benefit of child self-investment time, as defined in (16) and (20). Under this incentive contract, the child's reaction function (14) evaluated at $r_M = r^*_M$ implies:

$$\begin{aligned} \tau^1_{c,M} = \tau^*_{c,M}(\tau^1_{p,M}, r^*_M; \beta_{c,M}) &= \frac{\lambda_2 r^*_M + \Delta^1_{c,M}}{\lambda_1 + \lambda_2 r^*_M + \Delta^1_{c,M}} (\tilde{T}_t - \tau^1_{p,M}) \\ &= \frac{\Delta^1_{p,M}(\beta_{c,M})}{\tilde{\alpha}_5 + \Delta^1_{p,M}(\beta_{c,M})} (\tilde{T}_t - \tau^1_{p,M}) \\ &\equiv \gamma^1_M(r^*_M, \beta_{c,M})(\tilde{T}_t - \tau^1_{p,M}) \end{aligned}$$

where $\tilde{\alpha}_5 = \varphi \lambda_1$ denotes the parents' altruistic preference for child leisure, and $\tau^1_{p,M}$ is the parents' optimal total investment time conditional on using a CCT. The child's resulting fraction of study time is given by $\gamma^1_M(\cdot)$ (defined in equation (17)) evaluated at $r_M = r^*_M$.

Numerical solutions. The parents' optimal time investments (and hence, all other endogenous choices) will typically differ depending on whether they use a CCT or not, since the parents' CCT choice may affect the child's future discount factor, $\beta_{c,M+1}$, which in turn affects the marginal valuations of next-period human capital $(\psi_{c,M+1}, \psi_{p,M+1})$ upon which all of the household choices depend. Therefore, we must solve for $(\tau_{1,M}, \tau_{2,M})$ for the cases where $CCT_M = 0$ and the case where $CCT_M = 1$, respectively. Given the complex model structure – with parental time inputs entering both the parents' and child's time constraints – we cannot solve for $(\tau_{1,M}, \tau_{2,M})$ in reduced form, and instead use a numerical solver. This involves a grid search procedure where, for a given simulated household-period, we define a grid of size $K \times Z \times 2$, i.e., $K = 20$ possible values for the child's initial stock of cognitive skills k_M; $Z = 3$ possible values of the child's initial discount factor $\beta_{c,M} \in \{\beta_c^1, \beta_c^2, \beta_c^3\}$; and 2 possible CCT choices $CCT_M \in \{0, 1\}$. For each grid point $(k_M, \beta_{c,M}, CCT_M)$, we solve for all the household's remaining choices, and calculate the parents' indirect value function at each grid point. A convenient feature of our functional form assumptions is that the parents' value function is separable in initial (log) child capital. In particular, the term that multiplies $\ln k_M$ is

$$\tilde{\alpha}_4 + \beta_p \psi_{p,M+1}(CCT_M, \beta_{c,M})\delta_{5,M}$$

which depends on the flow utility of cognitive skills, $\tilde{\alpha}_4$, and the self-productivity of current skills, $\delta_{5,M}$. This implies that conditional on $(CCT_M, \beta_{c,M})$, all of the household's remaining optimal choices are independent of the initial level of k_M.

However, the parents' choice of whether or not to use a CCT will depend on both k_t as well as $\beta_{c,t}$. This complicates the solution relative to a simpler model where the child's discount factor is exogenously fixed and where all period t- choices are independent of k_t (see, e.g., DFW). The current model is richer by allowing for various complementarities between the current and future stocks of both cognitive and non-cognitive skills. First of all, current cognitive skills affect future cognitive skills through the self-productivity parameter, $\delta_{5,t}$. Second, current patience levels (as well as current CCT choices) affect future patience levels through the endogenous Markov transition probabilities. Third, cognitive and non-cognitive skills are complementary in the sense that each agent's (marginal) valuation of future cognitive skills, as defined in (13) and (18), is strictly increasing in the child's stock of non-cognitive skills, as summarized by the discount factor.

Optimal CCT choice. For a given initial state vector, $\Gamma_M = (w_{1,M}, w_{2,M}, I_M, k_M, \beta_{c,M})$, the parents' optimal CCT choice is derived numerically, by comparing their indirect value functions for the case where $CCT_M = 0$ and $CCT_M = 1$, respectively. So far we have derived the parents' optimal time and budget allocations for each case:

$$\mathbf{a}_{p,M}(\Gamma_M, CCT_M) = \begin{cases} a^0_{p,M}(\Gamma_M) \equiv \{\tau^0_{1,M}, \tau^0_{2,M}, l^0_{1,M}, l^0_{2,M}, h^0_{1,M}, h^0_{2,M}, c^0_M, e^0_M, x^0_M\} \\ a^1_{p,M}(\Gamma_M) \equiv \{\tau^1_{1,M}, \tau^1_{2,M}, l^1_{1,M}, l^1_{2,M}, h^1_{1,M}, h^1_{2,M}, c^1_M, e^1_M, r^*_M, b^*_M\} \end{cases}$$

where the superscripts correspond to the binary choice $CCT_M \in \{0,1\}$. Since using a CCT involves an instantaneous utility cost $\zeta \geq 0$ for the parents, they will only choose to implement a CCT if doing so maximizes their expected welfare conditional on the initial state vector Γ_M:

$$CCT_M(\Gamma_M) = \arg\max_{CCT\in\{0,1\}} \{V_{p,M}(\Gamma_M | a^0_{p,M}(\Gamma_M)), V_{p,M}(\Gamma_M | a^1_{p,M}(\Gamma_M))\}$$

where the second component includes the fixed utility cost, ζ. In the empirical analysis, we evaluate these value functions numerically in order to obtain the optimal CCT choice for every household-period and every state vector Γ_M.

Interpretation. Intuitively, parents should be more likely to want to use a CCT whenever the child's self-investment time in the absence of a CCT is sufficiently different relative to the incentivized case. Note that the child's optimal self-investment time in each case equals:

$$\tau^*_{c,M}(\tau_{p,M}, \beta_{c,M}) = \begin{cases} \gamma^0_M(\beta_{c,M})(\tilde{T}_t - \tau^0_{p,M}) = \frac{\beta_{c,M}\delta_{3,M}\psi^0_{c,M+1}(\beta_{c,M})}{\lambda_1+\beta_{c,M}\delta_{3,M}\psi^0_{c,M+1}(\beta_{c,M})}(\tilde{T}_t - \tau^0_{p,M}) \\ \gamma^1_M(\beta_{c,M})(\tilde{T}_t - \tau^1_{p,M}) = \frac{\beta_p\delta_{3,M}\psi^1_{p,M+1}(\beta_{c,M})}{\varphi\lambda_1+\beta_p\delta_{3,M}\psi^1_{p,M+1}(\beta_{c,M})}(\tilde{T}_t - \tau^1_{p,M}) \end{cases}$$

where $\tau^j_{p,M} = \tau^j_{1,M} + \tau^j_{2,M}$ denotes total parental time investments if $CCT_M = j \in \{0,1\}$. Without closed form solutions for the parental time investments, we cannot say definitively whether the child's self-investment time is greater in the case with or without a CCT. However, we can get some intuition by comparing the two fractions, $\gamma^0_M(\cdot)$ and $\gamma^1_M(\cdot)$, for a given initial discount factor $\beta_{c,M}$. Relative to the no-CCT case, the child will devote a larger fraction of their residual time to self-investment under the CCT whenever the following holds (suppressing some notation):

$$\begin{aligned} \gamma^1_M > \gamma^0_M \iff & \quad \frac{\Delta^1_{p,M}}{\varphi\lambda_1 + \Delta^1_{p,M}} > \frac{\Delta^0_{c,M}}{\lambda_1 + \Delta^0_{c,M}} \iff \Delta^1_{p,M} > \varphi\Delta^0_{c,M} \\ \iff & \quad \frac{\beta_p}{1-\beta_p}\frac{1-\varphi}{\varphi}\alpha_4 > \beta_{c,M}\psi^0_{c,M+1} - \beta_p\psi^1_{c,M+1} \\ \iff & \quad \frac{\beta_p}{1-\beta_p}\frac{1-\varphi}{\varphi}\frac{\alpha_4}{\lambda_3} > \beta_{c,M}\mathbb{E}_M\Big(\frac{1}{1-\beta_{c,M+1}}|CCT_M = 0\Big) \\ & \quad -\beta_p\mathbb{E}_M\Big(\frac{1}{1-\beta_{c,M+1}}|CCT_M = 1\Big) \end{aligned}$$

where the second and last lines follow from the definitions in (13), (16), (18) and (20). All else equal, the last inequality becomes easier to satisfy if (i) parents are more forward-looking relative to the child (i.e., β_p increases relative to $\beta_{c,M}$), (ii) parents are less altruistic towards the child (i.e., φ decreases), (iii) parents care more about the child's cognitive skills relative to the child herself (i.e., α_4 increases relative to λ_3), and (iv) the use of a CCT has a less detrimental (or more positive) impact on the child's future discount factor, $\beta_{c,M+1}$. Intuitively, a CCT is more effective at altering the child's self-investment time (relative to the no-CCT Stackelberg equilibrium) whenever there is a substantial misalignment between the parents' and the child's preferences, and whenever the CCT does not "backfire" too much in terms of making the child more myopic in the future.

B.2 Backward induction algorithm

In the empirical implementation, we use the following computational procedure to solve the period M problem, and perform the backward induction to period $t = M - 1$.

Step 0. Fix a set of model parameters, exogenous household characteristics (parental age and education, child's age etc.), a randomly drawn parental discount factor β_p, randomly drawn sequences of wage offers and non-labor income shocks, and an exogenous grid of $K \times Z$ possible values for the child's cognitive human capital stock, $k_t \in \{k^1, ..., k^K\}$ and discount factor, $\beta_{c,t} \in \{\beta_c^1, ..., \beta_c^Z\}$. Starting in period $t = M$, the algorithm proceeds as follows.

Step 1. Conditional on household characteristics, draw a random triple of wages and non-labor income $(w_{1,t}, w_{2,t}, I_t)$. Everything below implicitly conditions on this set of exogenous state variables.

Step 2. Define a first loop over initial state variable $\beta_{c,t} \in \{\beta_c^1, ..., \beta_c^Z\}$. For each CCT choice, $CCT_t \in \{0, 1\}$, solve for the household's remaining choices; solving numerically for the parental time inputs $(\tau_{1,t}, \tau_{2,t})$, and deriving closed-form solutions for all other parental choices $(h_{1,t}, l_{1,t},\ h_{2,t}, l_{2,t},\ c_t, e_t,\ x_t, b_t,\ r_t)$, and the child's optimal response $\tau_{c,t}$.[57] Conditional on $(\beta_{c,t}, CCT_t)$, we calculate the parents' "partial" indirect value function, denoted $\tilde{V}_{p,t}(\beta_{c,t}; CCT_t)$, which contains all utility components except those multiplying the initial stock of cognitive skills, $\ln k_t$. For $t = M$, this gives:

$$\begin{aligned}\tilde{V}_{p,t}(\beta_{c,t}; CCT_t) = \tilde{\alpha}_1 \ln l^*_{1,t} + \tilde{\alpha}_2 \ln l^*_{2,t} + \tilde{\alpha}_3 \ln c^*_t + \tilde{\alpha}_5 \ln l^*_{c,t} + \tilde{\alpha}_6 \ln x^*_t - \mathbb{1}[CCT_t = 1]\zeta \\ + \beta_p \psi_{p,t+1}(CCT_t, \beta_{c,t}) \Big(\ln R_t + \delta_{1,t} \ln \tau^*_{1,t} + \delta_{2,t} \ln \tau^*_{2,t} + \delta_{3,t} \ln \tau^*_{c,t} + \delta_{4,t} \ln e^*_t, \Big)\end{aligned}$$

[57] This is a computationally intensive step, but only needs to be done $2Z$ times, i.e. for every $(\beta_{c,t}, CCT_t)$ grid point.

which includes the fixed utility cost of using a CCT, ζ. The log-separable Cobb-Douglas model structure implies that all household choices (except for the binary CCT choice itself) do not depend on k_t directly, which simplifies the numerical computation. The only source of uncertainty stems from the future realization of the child's discount factor, $\beta_{c,M+1}$, which is summarized by the expected marginal valuation term $\psi_{p,M+1}(\cdot)$ defined in (18).

Step 3. Define a second loop over initial state variable $k_t \in \{k^1, ..., k^K\}$. Add back in the separable term multiplying $\ln k_t$, to calculate the "complete" value functions for period $t = M$:

$$V_{p,t}(k_t, \beta_{c,t}; CCT_t) = \tilde{V}_{p,t}(\beta_{c,t}; CCT_t) + \ln k_t \Big(\tilde{\alpha}_4 + \beta_p \psi_{p,t+1}(CCT_t, \beta_{c,t}) \delta_{5,t} \Big)$$

By comparing this indirect value function for the case where $CCT_t = 0$ or 1, we can derive the optimal choice $CCT_t^*(k_t, \beta_{c,t})$, and all other corresponding endogenous choices, which only depend upon k_t indirectly through their dependence on CCT_t^*. The child's indirect value function can be calculated analogously.

Step 4. Plugging in all optimal choices, define the $K \times Z$ matrices of indirect values, $V_{p,t}(k_t, \beta_{c,t})$ and $V_{c,t}(k_t, \beta_{c,t})$. In order to keep the backward induction process computationally tractable, we approximate these value functions by projecting them onto a first-order polynomial of the initial state variables, $\ln k_t$ and $\beta_{c,t}$: [58]

$$\hat{V}_{j,t}(k, \beta_c) = \hat{z}^j_{0,t} + \hat{z}^j_{1,t} \ln k + \hat{z}^j_{2,t} \beta_c + \hat{z}^j_{3,t} \ln k \times \beta_c$$

where the interaction term $z^j_{3,t}$ captures the skill complementarity for each agent $j \in \{p, c\}$. Given OLS estimates $\{\hat{z}^j_{0,t}, \hat{z}^j_{1,t}, \hat{z}^j_{2,t}, \hat{z}^j_{3,t}\}$ for each j, define the approximated value functions:

$$\hat{V}_{j,t}(k, \beta_c) = \hat{z}^j_{0,t} + \hat{z}^j_{1,t} \ln k + \hat{z}^j_{2,t} \beta_c + \hat{z}^j_{3,t} \ln k \times \beta_c$$

For any period $t \leq M$, this implies the following partial derivative for each agent j:

$$\frac{\partial \hat{V}_{j,t}}{\partial \ln k_t} \quad = \quad \hat{z}^j_{1,t} + \hat{z}^j_{3,t} \beta_{c,t}$$

Taking the expectation of these objects at time $t - 1$ conditional on $(\beta_{c,t-1}, CCT_{t-1})$, we can define each agent $j = p, c$'s expected marginal valuation of future (i.e., period-t) cognitive ability:

$$\psi_{j,t}(CCT_{t-1}, \beta_{c,t-1}) \quad \equiv \quad \mathbb{E}_{t-1} \Big(\frac{\partial \hat{V}_{j,t}}{\partial \ln k_t} \Big) = \sum_{i=1}^{Z} Pr(\beta_{c,t} = \beta^i_c | \beta_{c,t-1}, CCT_{t-1}) \, \big(\hat{z}^j_{1,t} + \hat{z}^j_{3,t} \beta^i_c \big)$$

[58] Throughout the simulations, the R-squared of these regressions is usually over 0.99.

$$= \quad \hat{z}^j_{1,t} + \hat{z}^j_{3,t}\mathbb{E}_{t-1}\big(\beta_{c,t}|\beta_{c,t-1}, CCT_{t-1}\big) \tag{24}$$

which follows from our implicit assumption that the only uncertainty at time $t-1$ stems from the stochastic draw of $\beta_{c,t}$ conditional on $(\beta_{c,t-1}, CCT_{t-1})$.[59] In other words, each agent's marginal value of period-t cognitive skills can be written as a deterministic function of known regression coefficients, known state variables and choices made in the previous period $t-1$. Thus, at each step in the backward recursion, we replace each agent j's (expected) continuation value function with their respective approximation:

$$\begin{aligned} \mathbb{E}_t\big(\hat{V}_{j,t+1}(k_{t+1}, \beta_{c,t+1})|\tau_{c,t}, \mathbf{a}_{p,t}, \Gamma_t\big) \quad &= \quad \psi_{j,t+1}(CCT_t, \beta_{c,t}) \ln k_{t+1} + \hat{z}^j_{0,t+1} \\ &\quad +\hat{z}^j_{2,t+1}\mathbb{E}_t(\beta_{c,t+1}|\beta_{c,t}, CCT_t) \end{aligned} \tag{25}$$

where everything is essentially deterministic; future skills k_{t+1} are given by the law of motion (4); the regression coefficients $\hat{z}^j_{t+1}$ are known with perfect foresight, and the child's expected future discount factor $\mathbb{E}_t(\beta_{c,t+1}|\beta_{c,t}, CCT_t)$ is straightforward to calculate given the Markov transition matrix and initial conditions specified in Section D.1.1. Given these approximate value functions, we can now go to the previous period $t = M - 1$.

Since $\beta_{c,t+1}$ only depends on $(CCT_t, \beta_{c,t})$, the period $t < M$ solution follows a very similar procedure as for period M, where the only meaningful differences are (i) using approximated rather than exact continuation values, and (ii) different reduced-form expressions for each agent's marginal valuation of future cognitive skills, $(\psi_{p,t+1}(\cdot), \psi_{c,t+1}(\cdot))$. As before, the solution algorithm first solves for all optimal time and budget allocation choices conditional on $(CCT_t, \beta_{c,t})$, then adds back in the terms multiplying $\ln k_t$ to solve for the optimal CCT choice, and finally runs a new set of surface regressions to approximate the period-t value functions. Conveniently, the fact that the period-t problem remains (almost) separable in $\ln k_t$ implies that the total number of grid points for which we have to numerically solve the model for each simulated household h is limited to $2 \times Z \times (M + 1 - t_{h,0})$.[60] The initial level of k_t only directly affect the optimal choice of $CCT_t \in \{0, 1\}$, which in turn affects all other endogenous decision variables.[61]

[59]Indeed, under our assumptions that (i) the technology of cognitive skill formation is deterministic and (ii) households have perfect foresight with respect to future wages and non-labor income draws, it follows that there is no uncertainty about future value functions. Therefore, we can treat the period-t regression coefficients $(\hat{z}^p_{1,t}, \hat{z}^p_{3,t}, \hat{z}^c_{1,t}, \hat{z}^c_{3,t})$ as known at any prior time $s \leq t$.

[60]In particular, we solve the model for 2 possible CCT choices, Z possible values of $\beta_{c,t}$, and $(M+1-t_{h,0})$ time periods, where $t_{h,0}$ denotes the child's initial age in 1997, when we initiate the model taking the child's initial cognitive test score as given.

[61]We could extend the model to allow for a direct effect of parental investment time $\tau_{p,t}$ (or other productive inputs) on the child's future discount rate $\beta_{c,t+1}$. However, identification concerns aside, this addition would significantly increase the computational complexity by a factor of K, since we would now require a numerical solution for every triple $(CCT_t, \beta_{c,t}, k_t)$, as opposed to the current model where we only need a numerical solution for every pair $(CCT_t, \beta_{c,t})$. The reason is that total parental investment time $\tau_{p,t}$ would also enter into the $\psi_{p,t+1}(\cdot)$ term multiplying $\ln k_t$.

The (approximate) period-t problem solved by the child is:

$$\begin{aligned} V_{c,t}(\Gamma_t|\mathbf{a}_{p,t}) &= \max_{\tau_{c,t}|\mathbf{a}_{p,t},\mathcal{C}_{c,t}} u_c(l_{c,t}, x_t, k_t) + \beta_{c,t}\mathbb{E}_t\left[\hat{V}_{c,t+1}(\Gamma_{t+1})|\tau_{c,t}, \mathbf{a}_{p,t}, \Gamma_t\right] \\ &= \max_{\tau_{c,t}|\mathbf{a}_{p,t},\mathcal{C}_{c,t}} u_c(l_{c,t}, x_t, k_t) + \beta_{c,t}\Big[\psi_{c,t+1}(CCT_t, \beta_{c,t}) \ln k_{t+1} + \hat{z}^c_{0,t+1} \\ &\qquad + \hat{z}^c_{2,t+1}\mathbb{E}_t(\beta_{c,t+1}|\beta_{c,t}, CCT_t)\Big] \end{aligned}$$

subject to the usual constraints and laws of motion, where the second line follows from the approximation defined in (25), and where $\psi_{c,t+1}(\cdot)$ is defined recursively in (24) for $t \leq M-1$. Similarly, the parents' period-t problem becomes:

$$\begin{aligned} V_{p,t}(\Gamma_t) &= \max_{\mathbf{a}_{p,t}|\tau^*_{c,t}(\mathbf{a}_{p,t}),\mathcal{C}_{p,t}} \tilde{u}_{p,t} + \beta_p\mathbb{E}_t\left[\hat{V}_{p,t+1}(\Gamma_{t+1})|\mathbf{a}_{p,t}, \Gamma_t\right] \\ &= \max_{\mathbf{a}_{p,t}|\tau^*_{c,t}(\mathbf{a}_{p,t}),\mathcal{C}_{p,t}} \tilde{u}_{p,t} + \beta_p\Big[\psi_{p,t+1}(CCT_t, \beta_{c,t}) \ln k_{t+1} + \hat{z}^p_{0,t+1} \\ &\qquad + \hat{z}^p_{2,t+1}\mathbb{E}_t(\beta_{c,t+1}|\beta_{c,t}, CCT_t)\Big] \end{aligned}$$

subject to the usual time and budget constraints and laws of motion, with $\tilde{u}_{p,t}$ defined in (1), and where we substitute in (25) in the second line. Since both the instantaneous utility functions, u_c and $\tilde{u}_p$, and the child's future stock of skills, k_{t+1}, are log-separable in k_t, both the child's and the parents' problem can be solved analogously to the period-M case.

C Data Appendix

C.1 Primary data set: PSID-CDS Household Panel

All our results are based on a selected sample of households from the Panel Study of Income Dynamics (PSID) and its Child Development Supplement (CDS). We consider PSID data from 1986 to 2010, and the first three CDS waves collected in 1997, 2002 and 2007. After merging the PSID and CDS data and applying our sample selection criteria (discussed in Section 3), we obtain a final data set of $N = 247$ households with children between the ages of 0 and 16 years old. The corresponding panel includes household-year level information on the following variables: (1) household identifier, (2) year, (3) number of child (i.e., birth order), (4) mother's age, (5) father's age, (6) family size, (7) mother's education, (8) mother's weekly labor supply, (9) mother's hourly wage, (10) father's weekly labor supply, (11) father's hourly wage, (12) weekly non-labor income, (13) child's age, (14) Letter Word raw score, (15) father's education, (16) joint parental active time, (17) mother's individual active time, (18) father's individual active time, (19) child's self-investment time, (20) total school time, (21) child's effective time endowment, (22) child's age in 1997, and (23) various questions related to unconditional and conditional allowances.

C.1.1 Joint parental time

We divide all joint time (where both parents are actively engaged with the child) between "mother's active time" and "father's active time", according to the relative proportions of these two individual parental time inputs, i.e., we assume each parent's active time investment is defined as the sum of their individual active time (where the other parent is not actively involved) and a fraction of the joint parental active time, where the mother's fraction is defined as the ratio between her individual active time, and the sum of the mother's and the father's individual active time. This ensures that there is no double-counting of joint parental time, which could otherwise violate the child's time constraint.

C.1.2 Measures of cognitive ability

The PSID-CDS contains several measures of children's cognitive ability, including the Letter Word score (LW), the Applied Problems score (AP) and the Passage Comprehension score (PC). We use the Letter Word (LW) score because it is administered to the broadest age range of children (starting at age 3). We also found that the two alternative measures are very strongly correlated with the Letter Word (LW) Score. Based on the subsample of child-year observations where test scores are observed simultaneously, the correlation between the LW and AP scores is around 0.93, and the correlation between the LW and the PC scores is around 0.87, which suggests that our anchoring procedure and model estimates would likely be robust to using these alternative ability measures.

C.1.3 Allowance use and CCTs

From the CDS interviews with the primary caregivers, we have data on (i) whether or not parents give their child an allowance (Yes/No), (ii) if so, how much the allowance is in dollars, and (iii) if so, whether the allowance is made contingent on the child doing his/her school work (Yes/No). Based on these questions, we construct our binary proxy measure for parents using a Conditional Cash Transfer (CCT) as the act of giving an allowance that is contingent on doing school work. In particular, we set $CCT = 0$ if the parents do not give an allowance, or if they give an unconditional allowance that is not contingent on doing school work. Only when the parents give an allowance contingent on school work do we define $CCT = 1$. In the estimation, we specify moments related to the fraction of households who use CCTs, conditional on age of the child and educational attainment of the parents. We do not use moments related to the dollar amounts of the allowance due to possible noise in the reporting (e.g., whether the amount is given daily, weekly, bi-monthly etc.).

C.1.4 Censoring and truncation

Actual data. Obvious reporting errors in the parental wage and labor supply data were resolved in the following way. For a given spouse in a given year, we replace the reported labor income and labor supply by missing values if (1) the reported labor income is positive but the reported labor hours are 0, (2) if the reported labor hours are positive but the labor income is 0, or (3) if either reported labor hours or labor income is missing.

If the non-labor income in any given year (calculated as the residual yearly income after subtracting both spouses' labor income) was either negative or above 1000 dollars per week, we replace all the corresponding hourly wage, labor hours and non-labor income data by missing values for that year.

If the labor supply for a given spouse was above 80 hours per week, we truncate that observation at 80. If an hourly wage rate for a given spouse was either less than $5 per hour or more than $150 per hour, we replace that observation by a missing value. However, we keep all the other information pertaining to that household.

Simulated data. All simulated data are being censored in exactly the same way as the original data. Hence, if the original data contain a missing value or a censored observation for some variable at some child age, then the simulated data will have a missing value in the corresponding cell (i.e. in all R corresponding cells, since we simulate $R > 1$ data sets). Similarly, whenever the simulations yields a corner solution for labor supply, we censor the corresponding simulated wage. However, we do not censor extreme simulated wage draws (i.e. below $5 or above $150 per hour).

Given our estimation procedure for the non-labor income process, simulated non-labor income draws cannot be negative. In the event that they exceed $1000 per week, we truncate that draw at $1000. Note that we cannot replace these extreme draws by a missing value (as we did for the actual data), since we always need a real-numbered (non-missing) value of non-labor income to simulate household choices in each period.

C.1.5 School time

We believe the reported school time data from the CDS to be relatively noisy, as can be seen in Table 15, which shows the distribution of reported school time at each child age t. Given the implausibly wide data range of these reported school times, we only use the median of these reported values (conditional on child age t), and use that as a measure to define the child's effective time endowment at age t as $T_{c,t} = 112 - med(s_t)$. To construct school time s_t, we use combined CDS data from 1997, 2002 and 2007, and define total school time as the sum of "regular" school time and "other" school time. These two subcomponents were constructed based on the following CDS time categories:

1. Regular school time: All time use with activity code
   - 5090: Student (full-time); attending classes; school if full-time student.
   - 5091: Daycare/nursery school for children not in school.
   - 5092-5093: School field trips inside/outside of regular school hours.
2. Other school time: all activities taking place at school with activity code
   - 5190-5193: Other classes, courses, lectures, being tutored.
   - 5680: Daycare/nursery before or after school only.
   - 6130-6138: Attending a before or after school club (math, science, drama, debate, band, ...).

Detailed descriptive statistics of these schooling components are available upon request. Finally, we note that time spent with babysitters, time spent at daycare before or after school, or time spent in home care from a non-household member (CDS activity code 4870) is not counted as school time.

Table 15: Distribution of Total Weekly School Time s_t by Child Age

Age	Mean	Std.	Min	P25	**Median**	P75	Max	NrZeros	NrObs
$t = 3$	11.250	17.866	0.000	0.000	0.000	26.042	47.083	6	9
$t = 4$	9.845	16.152	0.000	0.000	0.000	15.833	55.000	23	36
$t = 5$	13.725	16.830	0.000	0.000	0.000	29.583	56.250	21	40
$t = 6$	24.534	17.404	0.000	0.000	32.500	37.083	47.083	9	34
$t = 7$	31.739	9.956	0.000	32.500	33.333	35.000	45.417	1	23
$t = 8$	28.274	14.201	0.000	30.833	33.458	35.000	48.333	6	38
$t = 9$	31.814	12.061	0.000	30.833	34.167	37.812	50.000	5	59
$t = 10$	31.719	10.948	0.000	31.667	33.750	36.250	47.500	4	62
$t = 11$	30.629	14.027	0.000	30.771	34.583	38.750	56.833	6	53
$t = 12$	32.246	16.583	0.000	33.333	37.500	41.417	53.750	4	22
$t = 13$	31.948	12.409	0.000	32.500	34.583	37.604	45.833	4	37
$t = 14$	37.421	14.753	0.000	35.000	37.125	43.750	74.833	5	54
$t = 15$	35.792	15.259	0.000	34.375	37.500	43.750	60.833	7	56
$t = 16$	33.281	17.038	0.000	30.792	37.500	44.063	62.500	8	49

C.2 Data on Children's Discount Factors

We are grateful to Laurence Steinberg from Temple University for providing access to a unique experimental survey data set that collects detailed measures of patience and forward-lookingness. The sample consists of a cross-section of approximately 900 children and adults between ages 10 and 30. For our purposes, we focus on the subsample of children who are at most 17 years old, leaving us with 502 observations. Each individual was asked to respond to a number of delayed gratification questions with varying time delays, ranging from one day to one year. We use the two survey questions where the time delay is either 6 months or 12 months. The two questions were essentially of the following type: "If you could choose between either receiving X dollars today, or 1000 dollars in six (resp. twelve) months, what would you prefer?" By using a sequence of follow-up questions, the survey narrows down each individual's indifference (or "switching") point, which we denote by the random variable $X_{i,t} \in [0, 1000]$, where i denotes the individual and $t \in \{6, 12\}$ denotes the time delay in months. For example, if respondent i prefers 801 dollars today over 1000 in 12 months, but prefers 1000 dollars in 12 months over 799 dollars today, then the measured indifferent point $X_{i,12} \approx 800$. If we assume survey respondents have concave, log-utility preferences (as we do in our model) and time-consistent intertemporal preferences (i.e., exponential discounting), then we could simply calculate everyone's annualized discount factor by exploiting the following indifference condition:

$$\ln(X_{i,12}) = \beta_{i,12}\ \ln(1000) \Rightarrow \beta_i = \frac{\ln(X_{i,12})}{\ln(1000)}$$

However, it has often been shown that experimental survey respondents do not exhibit time-consistent preferences, and may suffer from present-bias (see Frederick, Loewenstein and O'Donoghue (2002) for a critical review). In particular, we expect survey responses to be present-biased when the delayed gratification questions are phrased as either receiving some amount "right now" or "in 12 months". However, we would not expect responses to exhibit present-bias when the question is phrased as either receiving some amount "in one month" or "in 13 months", because then both amounts are delayed (as is the case in the Osaka Preference Parameters Study, discussed in Appendix C.3). In standard models of quasi-hyperbolic discounting, an individual's time preferences can be summarized by a pair (δ, β), where we let δ denote the short-term discount factor which applies only to one-period-ahead payoffs, and let β denote the long-term discount factor which applies to all future payoffs. In this two-parameter framework, we can separately identify the long-run discount factor β (which is our object of interest) by exploiting the fact that we observe two survey questions with different time delays. For each respondent i, the two indifference points $(X_{i,6}, X_{i,12})$ solve the following system:

$$\begin{aligned}
\ln(X_{i,6}) &= \delta_i \beta_{i,6} \ln(1000) \\
\ln(X_{i,12}) &= \delta_i \beta_{i,12} \ln(1000)
\end{aligned}$$

where $\beta_{i,6}$ and $\beta_{i,12}$ denote the semi-annual and annual discount factors, respectively. Since standard discounting implies that $\beta_{i,6} = \sqrt{\beta_{i,12}}$, we can identify each respondent's annual discount factor as follows:

$$\beta_{i,12} = \left(\frac{\ln(X_{i,12})}{\ln(X_{i,6})}\right)^2$$

Figure 3 shows summary statistics of the resulting empirical distribution, by age of the individual and (for adolescents only) by SES of the individual's parents.

C.3 Data on Parental Discount Factors

To calibrate the distribution of parental discount factors in our model, we use survey data from the 2010 U.S. wave of the Osaka Preference Parameters Study ("PPS"). We focus mainly on the following sequence of questions that allows the elicitation of an annualized discount factor: "Suppose you have the option to receive either 100 dollars in one month ("Option A"), or receive a different amount (X) in thirteen months ("Option B"). What would you choose?" The survey proceeds by setting the delayed payoff X equal to 100, 102, 104, 106, 110, 120, 140, 180 or 250 dollars, respectively. We retain only individuals who entered a valid set of responses, by exhibiting a single "switching point". This includes individuals who either always choose Option A, those who always choose Option B, and those who always choose Option A up to some amount X, and then always choose Option B for amounts larger than X. For this sample of "rational" adults, we define the random variable X_i as the upper bound on their switching point (i.e., someone who prefers Option A when $X = 104$ but prefers Option B when $X = 106$ would have a switching point $X_i = 106$ as an upper bound). For the small subset of individuals who always choose Option A (resp. always Option B), we set the switching point equal to 100 (resp. 300). Then, assuming individuals have concave, log-utility preferences, we can infer their annualized discount factors from their (approximate) indifference condition:

$$\ln(100) = \beta_i \ln(X_i) \iff \beta_i = \frac{\ln(100)}{\ln(X_i)}$$

We assume these imputed discount factors to be free of present-bias, since both Option A and B are delayed by at least one month. Therefore, the elicited long-run annual discount factors for adults are comparable to those elicited for children and adolescents based on the Steinberg et al. (2009) survey, as discussed in Appendix C.2. Figure 4 shows the empirical distributions of the imputed annual discount factors for adults conditional on their educational attainment level, as reported in the PPS.

Model simulation. Table 16 shows the discrete probability distribution for each education type (see also Section D.1.1 for more details on the exogenous specification of parental discount factors in the model). When simulating the model, we generate $N \times R$ random households, where N denotes the number of households in our PSID sample ($N = 247$), and R denotes the number of random draws per household ($R = 10$). For each household $h = 1, ..., N$, we take the father's educational attainment level as given from the PSID, and generate R random draws from the appropriate conditional distribution shown in Table 16. The resulting model fit is discussed in Section 5.5, and summarized in Panel (a) of Table 12.

Table 16: Conditional Distribution of Parental Discount Factors

	Educational attainment level s		
Mass point	**High school or less**	**(Some) College**	**Graduate**
$Pr(\beta_p = \beta_p^1 \mid s)$	0.353	0.301	0.235
$Pr(\beta_p = \beta_p^2 \mid s)$	0.224	0.279	0.293
$Pr(\beta_p = \beta_p^3 \mid s)$	0.423	0.420	0.472
Sum	1	1	1

Notes: The table shows the discretized conditional distribution of adults' discount factors for various education groups, where the grid points are ordered from low to high such that $\beta_p^1 < \beta_p^2 < \beta_p^3$, and where "s" denotes educational attainment level. "High school or less" is defined as having either 11 or 12 years of schooling, "(Some) college" is defined as having either an associate's degree, a bachelor's degree, or some years of college but no degree, and "Graduate" is defined as having some postgraduate, a Master's or a doctoral degree. The probabilities sum up to 1 within each column, and correspond to the likelihood of being assigned either a low, medium or high discount factor conditional on educational attainment. The calibrated grid points are $\beta_p^1 \approx 0.85$, $\beta_p^2 \approx 0.95$, and $\beta_p^3 \approx 0.99$.
Source: Osaka Preference Parameters Study, 2010 U.S. Survey. Only adults between ages 25 and 65 with known educational attainment and valid survey responses.

D Econometric Implementation

D.1 Econometric Specification

D.1.1 Preferences

Both parents' and children's instantaneous utility functions are assumed to be constant over time and across households. However, we allow for heterogeneity in agents' discount factors (both within and across households) which enter into their final-period utility functions. In particular, the child's discount factor is allowed to be time-varying and endogenous, and the parents' discount factor is assumed to be exogenous but allowed to vary with educational attainment. We also allow the fixed cost of using a CCT to vary across households.

Instantaneous utility. For the parents' preferences, the four-dimensional vector $\alpha = (\alpha_1 \ \alpha_2 \ \alpha_3 \ \alpha_4)'$ is defined such that $\sum_j \alpha_j = 1$, $\alpha_j > 0$, $j = 1, \ldots, 4$. Similarly, the child's preferences are given by a three-dimensional vector $\lambda = (\lambda_1 \ \lambda_2 \ \lambda_3)'$, where $\sum_j \lambda_j = 1$, $\lambda_j > 0$, $j = 1, 2, 3$. These restrictions are standard and ensure that utility is increasing in each argument and the summations impose an inconsequential normalization on the utility functions of the parents and the child. In the econometric implementation, we estimate five real-valued parameters $v = (v_1 \dots v_5)'$ that map into the preference parameters as follows:

$$\begin{aligned}
\alpha_1 &= \frac{\exp(v_1)}{1+\exp(v_1)+\exp(v_2)+\exp(v_3))}, \quad \alpha_2 = \frac{\exp(v_2)}{1+\exp(v_1)+\exp(v_2)+\exp(v_3))}, \\
\alpha_3 &= \frac{\exp(v_3)}{1+\exp(v_1)+\exp(v_2)+\exp(v_3))}, \quad \alpha_4 = \frac{1}{1+\exp(v_1)+\exp(v_2)+\exp(v_3))}, \\
\lambda_1 &= \frac{\exp(v_4)}{1+\exp(v_4)+\exp(v_5)}, \quad \lambda_2 = \frac{\exp(v_5)}{1+\exp(v_4)+\exp(v_5)}, \quad \lambda_3 = \frac{1}{1+\exp(v_4)+\exp(v_5)}.
\end{aligned}$$

CCT utility cost. We also estimate a scalar parameter κ, which governs the distribution of the instantaneous utility costs associated with implementing a CCT. The utility cost, $\zeta \geq 0$, is assumed to stay fixed over time but can vary across households.[62] For each household, the value of ζ is drawn from an exponential distribution with mean and standard deviation $\kappa > 0$. This parameter is assumed to be the same for all households.

Child discount factors: Markov process. As specified in equation (7), the child's next-period discount factor, denoted $\beta_{c,t+1}$, is determined endogenously and stochastically, as a function of the child's current discount factor, $\beta_{c,t}$, the child's age, t, and the parents' current CCT choice, $CCT_t \in \{0, 1\}$. In the application, we assume that in any given period $t = 1, ..., M + 1$, $\beta_{c,t}$ can take one of finitely many values, denoted by the set

$$\mathcal{B}_c = \{\beta_c^1, ..., \beta_c^Z\}, \quad \text{where} \quad 0 < \beta_c^1 < ... < \beta_c^Z < 1.$$

These values are calibrated exogenously based on the subsample of children between 10 and 17 years old from the survey data set of Steinberg et al. (2009). In particular, we assume that $Z = 3$, and we define each grid point β_c^j $(j \in \{1, 2, 3\})$ as the average annual discount factor within the j'th tercile of the empirical distribution (see Appendix C.2 and Figure 3 for more details). The resulting grid points are:

$$\beta_c^1 = 0.343, \ \beta_c^2 = 0.773, \ \beta_c^3 = 0.982.$$

[62]It is straightforward computationally to allow ζ to vary with child age, but we believe that this characteristic of households is likely to be highly persistent.

At the start of period $t+1$, the child's discount factor is determined by taking a stochastic draw from an endogenous $Z \times Z$ Markov transition matrix:

$$Pr(\beta_{c,t+1} = \beta_c^{j'} | \beta_{c,t} = \beta_c^j, t, CCT_t) = \begin{cases} p^{up}(j, t, CCT_t) & \text{if } j' = j+1 \\ p^{stay}(j, t, CCT_t) & \text{if } j' = j \\ p^{down}(j, t, CCT_t) & \text{if } j' = j-1 \\ 0 & \text{otherwise} \end{cases}$$

where j corresponds to the period-t grid point, j' corresponds to the period-$t+1$ grid point, and where $p^{up} + p^{stay} + p^{down} = 1$ for all values of (j, t, CCT_t). This structure assumes that the child's discount factor cannot move up or down more than one level within a single period.[63] In the bottom state, $j = 1$, we restrict $p^{down} = 0$, and similarly in the top state, $j = Z$, we restrict $p^{up} = 0$ for all t. We assume the Markov matrix takes the following flexible parametric form:

$$\begin{array}{llll} \text{Bottom state } j = 1: & p^{up} = \dfrac{D^+}{1+D^+}, & p^{stay} = \dfrac{1}{1+D^+}, & p^{down} = 0 \\ \text{Middle state } j = 2: & p^{up} = \dfrac{D^+}{1+D^+ + D^-}, & p^{stay} = \dfrac{1}{1+D^+ + D^-}, & p^{down} = \dfrac{D^-}{1+D^+ + D^-} \\ \text{Top state } j = 3: & p^{up} = 0, & p^{stay} = \dfrac{1}{1+D^-}, & p^{down} = \dfrac{D^-}{1+D^-} \end{array}$$

where D^+ and D^- are shorthand notation to denote two strictly positive functions of the child's age and the parents' binary CCT choice:

$$\begin{aligned} D^+ &= \exp(b_0^{up} + b_1^{up} \times t + b_2^{up} \times CCT_t + b_3^{up} \times t \times CCT_t), \\ D^- &= \exp(b_0^{down} + b_1^{down} \times t + b_2^{down} \times CCT_t + b_3^{down} \times t \times CCT_t) \end{aligned}$$

To facilitate identification using our available data sources, we impose symmetry on the parameters that govern the effect of using a CCT on the probability of the child's discount factor going up or down in the next period:

$$b_2^{up} = -b_2^{down} \quad \text{and } b_3^{up} = -b_3^{down}$$

In particular, if using a CCT reduces the child's future discount factor (in a probabilistic sense), we would expect b_2^{up} to be negative, or vice versa, b_2^{down} to be positive. Similarly, if the "crowding out effect" of using a CCT exacerbates with age, we would expect the

[63] This assumption is made to simplify the computational burden, but could easily be relaxed. Similarly, we could allow for more than $Z = 3$ grid points, but this would require a larger and more detailed data sample on child discount factors.

interaction term b_3^{up} to be negative, or vice versa, b_3^{down} to be positive.[64] This leaves us with six parameters to estimate: $\{b_0^{up}, b_1^{up}, b_2^{up}, b_3^{up}, b_0^{down}, b_1^{down}\}$.

Child discount factors: initial conditions. Given the availability of cognitive test scores and time survey data from the CDS-I, CDS-II and CDS-III recorded in 1997, 2002 and 2007, respectively, we simulate our model $R = 10$ times for each household $h = 1, ..., N$ between ages $t = t_{h,0}$ and $t = M$, where $t_{h,0}$ denotes the child's age in 1997 when we first observe the child in the CDS, and where $M = 16$ denotes the last stage of the developmental process. Whereas we can use the child's observed test score in 1997 as an initial condition for the cognitive development process (see below), the CDS does not include an equivalent quantitative measure of the child's discount factor. Moreover, the experimental survey data set from Steinberg et al. (2009) only records time preferences of children and adolescents who are at least 10 years old, whereas the children in our PSID-CDS sample are observed at ages as early as 3 years old in 1997. Therefore, to initialize the model, we specify an initial discount factor for the child in each simulated household, taking the child's initial age in 1997, $t_{h,0} \in \{3, ..., 12\}$, and the father's education level, $s_{h,2}$, as given. For each simulated household h, we specify the child's initial discount factor, $\beta_{c,t_{h,0}}$, as a random draw from a probability distribution that is allowed to depend on $t_{h,0}$ and $s_{h,2}$:

$$\begin{aligned} Pr(\beta_{c,t_{h,0}} = \beta_c^j | t_{h,0}, s_{h,2}) &= \frac{\exp(\nu_0^j + \nu_1^j t_{h,0} + \nu_2^j s_{h,2})}{1 + \sum_{j=1}^{Z-1} \exp(\nu_0^j + \nu_1^j t_{h,0} + \nu_2^j s_{h,2})}, \quad j = 1, ..., Z-1 \\ Pr(\beta_{c,t_{h,0}} = \beta_c^Z | t_{h,0}, s_{h,2}) &= \frac{1}{1 + \sum_{j=1}^{Z-1} \exp(\nu_0^j + \nu_1^j t_{h,0} + \nu_2^j s_{h,2})} \end{aligned}$$

For the case with $Z = 3$ grid points, this implies six new parameters to estimate.

Parental discount factors. As discussed in Section 2.1.7, we assume that parental discount factors, denoted by $\beta_p \in (0, 1)$, are time-invariant, exogenous, heterogeneous in the population, and potentially correlated with the parents' educational attainment level. We assume β_p can take on a finite number of values, denoted by the set

$$\mathcal{B}_p = \{\beta_p^1, ..., \beta_p^Z\}, \quad \text{where} \quad 0 < \beta_p^1 < ... < \beta_p^Z < 1.$$

In our application, the grid points and conditional distributions for parental discount factors are calibrated exogenously (outside of the model) based on the subsample of adults between ages 25 and 65 from the Osaka Preference Parameters Study (see Appendix C.3 for more details). For simplicity, we assume that $Z = 3$, i.e., the same number of grid points as for

[64] This flexible specification allows us to rationalize why our data simultaneously show that children's average discount factors tend to increase with age, whereas the use of conditional allowances (or CCTs) decreases with age.

children. We define each grid point β_p^j ($j \in \{1, 2, 3\}$) as the average annual discount factor within the j'th tercile of the (unconditional) empirical distribution. The resulting values are:

$$\beta_p^1 = 0.847,\ \beta_p^2 = 0.947,\ \beta_p^3 = 0.990.$$

When simulating our model, each simulated household takes a random draw from this set, with a conditional probability distribution that is allowed to depend on the father's educational attainment level. For simplicity, we consider three parental education levels or "types": high school or less; some college or college degree; and graduate degree. Figure 4 shows the empirical distribution of discount factors for each education group, indicating a small but positive correlation between forward-lookingness and educational attainment. Then, for each type, we approximate the empirical distribution with a discrete probability distribution that consists of (i) the three mass points given by the support set $\mathcal{B}_p$, and (ii) a probability mass distribution that corresponds to the proportion of observations that fall within each of the unconditional tercile cutoffs, which were used to define the grid points in $\mathcal{B}_p$.

Table 16 shows the resulting discrete probability distribution for each education type. We notice that as educational attainment increases, so does the probability mass associated with larger discount factors. In particular, the probability mass associated with the lowest discount factor ($\beta_p^1 \approx 0.85$) decreases monotonically with education, from over 35 percent for individuals with at most a high school degree to less than 24 percent for graduates. Conversely, the probability mass associated with the middle discount factor ($\beta_p^2 \approx 0.95$) increases monotonically, from around 22 percent for individuals with at most a high school degree to over 29 percent for graduates. The probability mass associated with the highest discount factor ($\beta_p^3 \approx 0.99$) ranges from 42 percent for individuals with a college degree or less, to over 47 percent for graduates.

Appendix C.3 and Section 5.5 contain more details on the model simulations and fit of the parental discount factors. To assess model fit, panel (a) in Table 12 shows how the conditional distributions in the model simulations are well-calibrated to the Osaka survey data.

Taken together, the distribution of household preferences is thus characterized by 18 free parameters to be estimated jointly within the model. It is important to note that we do not attempt to estimate the altruism parameter, but instead are fixing it at $\varphi = \frac{1}{3}$. Attempting to estimate this parameter would make identification of the remaining parameters more challenging.

D.1.2 Technology of Cognitive Skill Formation

We allow the production function parameters of all inputs to vary with the age of the child and, for the two parental time inputs, also with respect to the education levels of the parents in household h. We economize on parameters by assuming that the input-specific productivity parameters are given by

Mother's investment time:	$\delta_{h,1,t} = \exp(d_{1,0} + d_{1,1}t + d_{1,2}s_{h,1}),$
Father's investment time:	$\delta_{h,2,t} = \exp(d_{2,0} + d_{2,1}t + d_{2,2}s_{h,2}),$
Child's self-investment time:	$\delta_{3,t} = \exp(d_{3,0} + d_{3,1}t),$
Investment goods:	$\delta_{4,t} = \exp(d_{4,0} + d_{4,1}t),$
Lagged ability:	$\delta_{5,t} = \exp(d_{5,0} + d_{5,1}t),$
Total Factor Productivity:	$R_t = d_{6,0} + \dfrac{d_{6,1} - d_{6,0}}{1 + \exp(-d_{6,2}(t - d_{6,3})}$

where h indexes the household and $s_{h,1}$ and $s_{h,2}$ are the mother and father's years of schooling, respectively. The productivity of each input is restricted to be monotonic in the age of the child ($t = 1, ..., M$), with the characteristics of the parents potentially entering as shifters of the profile. Parental schooling levels, meant to roughly capture the human capital levels of the parents, only appear in the δ productivity parameters connected with each parent's respective time input. Our specification is intended to capture the possibility that parents with higher human capital can provide higher quality time inputs to their children. For the Total Factor Productivity process (R_t), we have adopted a flexible generalized logistic parametrization which imposes monotonicity in the child's age.[65]

D.1.3 Wage Offers

The hourly wage offer processes for each parent $i = 1, 2$, in household h are specified as follows:

$$\ln w_{h,i,t} = \eta_{0,i} + \eta_{1,i} age_{h,i,t} + \eta_{2,i} s_{h,i} + \varepsilon_{h,i,t}$$

where $age_{h,i,t}$ and $s_{h,i}$ denote the age and education of parent i when the child is age t. $\eta_{1,i}$ is the age coefficient for parent i, and $\eta_{2,i}$ is the labor market "return" to schooling for parent i. The wage shocks are assumed to be serially uncorrelated draws from the following joint distribution:

$$(\varepsilon_{h,1,t}, \varepsilon_{h,2,t}) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{w_1} & \sigma_{w_1,w_2} \\ \sigma_{w_1,w_2} & \sigma^2_{w_2} \end{bmatrix} \right)$$

where σ_{w_1,w_2} ($= \rho_{12}\sigma_{w_1}\sigma_{w_2}$) is the contemporaneous covariance between the parental wage shocks. Any correlation between these disturbances ($\rho_{12} \neq 0$) could arise through assortative mating on unobservable determinants of wages and from the parents inhabiting the same

[65] Note that if the slope/steepness parameter $d_{6,2}$ is positive, the other parameters can be interpreted as the lower limit ($\lim_{t\to-\infty} R_t = d_{6,0}$), the upper limit ($\lim_{t\to+\infty} R_t = d_{6,1}$) and the inflection point ($\lim_{t\to d_{6,2}} R_t = \frac{d_{6,0}+d_{6,1}}{2}$), respectively. In the estimation, we restrict $(d_{6,0}, d_{6,1}) \in (0, 10)$, $d_{6,2} \in (-4, 4)$ and $d_{6,3} \in (-20, 20)$.

local labor market. Although we assume the shocks are independent over time, this is not implied by the model structure and is not a necessary assumption for identification of the key model parameters. We can allow for temporal dependence at some computational cost.

D.1.4 Non-Labor Income

We restrict non-labor income $I_{h,t}$ to be non-negative for all households and periods. We assume non-labor income is a function of the parent's ages and their schooling. In particular, we first define the probability of household h having a strictly positive non-labor income level in period t as

$$Pr(I_{h,t} > 0) = \frac{\exp(\zeta_{h,t})}{1 + \exp(\zeta_{h,t})}$$

where $\zeta_{h,t}$ denotes a polynomial in age and education of both spouses:

$$\zeta_{h,t} = \mu_1 + \mu_2 age_{h,1,t} + \mu_3 age^2_{h,1,t} + \mu_4 s_{h,1} + \mu_5 age_{h,2,t} + \mu_6 age^2_{h,2,t} + \mu_7 s_{h,2}.$$

If non-labor income is positive, the level of non-labor income is determined as follows:

$$\ln I_{h,t} = \mu_8 + \mu_9 age_{h,1,t} + \mu_{10} age^2_{h,1,t} + \mu_{11} s_{h,1} + \mu_{12} age_{h,2,t} + \mu_{13} age^2_{h,2,t} + \mu_{14} s_{h,2} + \varepsilon^I_{h,t},$$

where the shock $\varepsilon^I_{h,t}$ is assumed to be i.i.d. $N(0, \sigma^2_I)$ for all h, t.

The non-labor income parameters can be identified and estimated outside of the remaining model structure because the non-labor income process is strictly exogenous and $I_{h,t}$ is observed for all households. We estimate the 15 parameters governing this non-labor process in a first step, and bootstrap all parameters of the full model repeating this first step for each bootstrap data draw in order to account for sampling variation in this initial step. To simulate random non-labor income draws, we first draw from a Bernoulli distribution with probability $Pr(I_{h,t} > 0)$ to determine if income is positive or 0, and, if positive, then draw again from a Normal shock distribution to simulate the level of non-labor income.

D.1.5 Measuring Child Cognitive Ability

Rather than assume the stock of child cognitive ability (or skills) are perfectly measured in our data, we allow for only imperfect measures. To derive the mapping between unobserved (latent) child ability, k_t, and measured child ability, k^*_t in our data, we build on the approach utilized by psychometricians (see, e.g., chapter 17 in Lord and Novick, 1968). Consistent with prior research on this subject, we consider child ability to be inherently unobservable to the analyst, although we assume that it is observable by household members, since it is a determinant of the household utility level and is a (potential) input into the decision-making process.

Most cognitive test scores, such as the one used in our empirical work, are simple sums of the number of questions answered correctly by the test-taker. If a child of age t has an ability level of k_t, the probability that they correctly answer a question of difficulty d is $p(k_t, d)$. It is natural to assume that p is non-decreasing in its first argument for all d and is non-increasing in its second argument for all k_t. Taking the model to data, we assume that the test used in the empirical analysis consists of equally "difficult" questions, and we drop the argument d for simplicity. Related papers that consider more general item response specifications (e.g., two-parameter logistic or "2PL" models) show that results are similar (see e.g., Verriest, 2022).

Given a cognitive ability test consisting of NQ items of equal difficulty, the number of correct answers, $k_t^* \in \{0, 1, \ldots, NQ\}$ is distributed as a Binomial random variable with parameters $(NQ, p(k_t))$. Note that the randomness inherent in the test-taking process implies that the mapping between latent k and observed k^* is stochastic, and that a child of latent skill k_t has a positive probability of answering each question correctly. For a child of latent skill k_t, the expected number of questions they answer correctly is given by $p(k_t)NQ$.

In order to identify the model, we do have to take a position on the form of the function $p(k_t)$. We choose the following nondecreasing function:

$$p(k_t) = \frac{\exp(L_{0,t} + L_{1,t} \ln k_t)}{1 + \exp(L_{0,t} + L_{1,t} \ln k_t)} \tag{26}$$

where $L_{0,t}$ and $L_{1,t} > 0$ are measurement (i.e., location and scale) parameters.[66]

For model identification, we must first solve the generic problem of indeterminacy due to the fact that latent child skills (k) do not have any natural units. This is accomplished by normalizing the measurement parameters $(L_{0,\tilde{t}}, L_{1,\tilde{t}})$ for one age $\tilde{t}$. Given that our estimation (and backwards induction procedure) also requires discretizing the (log) child skills process on a grid, we choose a convenient normalization such that the log of the lowest grid point $(\ln(k_{min}))$ is zero, where k_{min} corresponds to the latent skill level of a child whose probability of answering any question correctly (at some age $t = \tilde{t}$) is at the lower bound of $p = p_{min} = 0.01$.[67] Given our transformation function defined in (26), this can be done by setting $L_{0,\tilde{t}} = \ln(p_{min}) - \ln(1 - p_{min}) \approx -4.6$ and $L_{1,\tilde{t}} = 1$, thus fixing the location and scale of latent skills at all ages k_t to the age $\tilde{t}$ measure. Second, we cannot separately identify an age-varying measurement system (parameterized by the $L_{0,t}, L_{1,t}$ parameters) from age-varying production function primitives $\delta_{1,t}, \ldots, \delta_{5,t}$ and R_t. We solve this under-identification problem by assuming the measurement system is age-invariant as in Agostinelli and Wiswall

[66]An alternative, but equivalent, formulation that makes clear the role of measurement error is to write that a child of latent ability k answers a test item correctly if $L_0 + L_1 \ln k + \epsilon > 0$, and incorrectly otherwise. ϵ in this formulation is the stochastic measurement error. If ϵ takes on an i.i.d. extreme value distribution, then the probability of answering a test item correctly takes the familiar form given by (26).

[67]Since we have $NQ = 57$ questions, a child with a latent skill level of $k = k_{min}$ would not be expected to answer more than 1 question correctly).

(2022), and assume $L_{0,t} = -4.6$ and $L_{1,t} = 1$ for all t.

D.2 Identification

D.2.1 Systematically Missing Data

Although our model period is annual, gaps in the CDS data that make it impossible to use successive observations on child ability along with inputs to identify the production parameters directly. We observe an imperfect measure of child ability in 1997 (a score on a cognitive test), along with the factor utilization levels in that year, but we only observe the next measure of child ability five years later in 2002; for some households we also have a third observation in 2007. In between these dates, input decisions have been made and levels of child ability have been determined; these input decisions depend on wage and non-labor income draws in the intervening years.

One approach to accommodating the gaps in the time series of our panel data is to reduce the number of decision-making periods, collapsing the dynamic model to just a few periods (a single period for early and another for late childhood, say). Given the rapid changes in child development, even during a single year, we instead prefer to assume that the decision-period frequency is annual. In this case, we measure the initial child ability and input decisions at age t, and only observe subsequent child cognitive ability at age $t+5$. After repeated substitution in the "reduced" skill technology function (12), we have

$$\ln k_{h,t+5} = a_1(\{\ln R_s\}_{s=t}^{t+4}) + a_2(\{\mathbf{Z}_{h,s}\}_{s=t}^{t+4}) + a_3\phi_t \ln k_{h,t} + a_1(\{\varepsilon_{h,s}\}_{s=t}^{t+4}), \tag{27}$$

where

$$\begin{aligned}
a_1(\{X_s\}_{s=t}^{t+4}) &= X_{t+4} + \phi_{t+4}X_{t+3} + \phi_{t+4}\phi_{t+3}X_{t+2} + \phi_{t+4}\phi_{t+3}\phi_{t+2}X_{t+1} + a_3X_t, \\
a_2(\{\mathbf{Z}_{h,s}\}_{s=t}^{t+4}) &= \mathbf{Z}_{h,t+4}\boldsymbol{\delta}_{t+4} + \phi_{t+4}\mathbf{Z}_{h,t+3}\boldsymbol{\delta}_{t+3} + \phi_{t+4}\phi_{t+3}\mathbf{Z}_{h,t+2}\boldsymbol{\delta}_{t+2} \\
&\quad +\phi_{t+4}\phi_{t+3}\phi_{t+2}\mathbf{Z}_{h,t+1}\boldsymbol{\delta}_{t+1} + a_3\mathbf{Z}_{h,t}\boldsymbol{\delta}_t, \\
a_3 &= \phi_{t+4}\phi_{t+3}\phi_{t+2}\phi_{t+1}.
\end{aligned}$$

The main problem confronting us is the lack of information on the sequence of inputs $\mathbf{Z}_{h,t+1}, ..., \mathbf{Z}_{h,t+4}$. The input choice of household h at any time t depends on a subset of the state variables characterizing the household, which we denote by $\tilde{\Gamma}_{h,t} = (w_{h,1,t}\ w_{h,2,t}\ I_{h,t}\ k_t\ \beta_{c,t}\ \Upsilon_h)$, where Υ_h denotes the values of the preference parameters for the household. As we know, except for the binary CCT choice, the household's input decisions in period t do not directly depend on the value of $k_{h,t}$ or of previous values $k_{h,t-1}, k_{h,t-2}, ...$, conditional on $\tilde{\Gamma}_{h,t}$. The state variables $w_{h,1,t}$, $w_{h,2,t}$, $I_{h,t}$ are functions of exogenous shocks and observable household characteristics that evolve deterministically over time. Let the value of these covariates (e.g., parental ages and schooling levels) at any time t be given by $H_{h,t}$. As we have seen above, we have written the wage offer processes of the parents and the non-labor income process as parametric functions of $H_{h,t}$. For the moment, assume that the parameters of these three exogenous processes are known, and further assume further that the preference parameter distribution G is also known. Then, at time t, we can form the expected values

$$E(\mathbf{Z}_{h,s}|H_{h,s}),\ s=t+1,...,t+4,$$

where the expectation is taken with respect to the shocks to the wage and income processes at time s and the preference parameter distribution G. Then we can define the expectation of the function a_2 as follows:

$$\begin{aligned} & Ea_2(\{\mathbf{Z}_{h,s}\}_{s=t}^{t+4}|\{H_{h,s}\}_{s=t+1}^{t+4}) = E(\mathbf{Z}_{h,t+4}|H_{h,t+4})\boldsymbol{\delta}_{t+4} + \phi_{t+4}E(\mathbf{Z}_{h,t+3}|H_{h,t+3})\boldsymbol{\delta}_{t+3} \\ + \quad & \phi_{t+4}\phi_{t+3}E(\mathbf{Z}_{h,t+2}|H_{h,t+2})\boldsymbol{\delta}_{t+2} + \phi_{t+4}\phi_{t+3}\phi_{t+2}E(\mathbf{Z}_{h,t+1}|H_{h,t+1})\boldsymbol{\delta}_{t+1} + a_3\mathbf{Z}_{h,t}\boldsymbol{\delta}_t. \end{aligned}$$

Then (27) becomes

$$\begin{aligned} \ln k_{h,t+5} \quad = \quad & a_1(\{\ln R_s\}_{s=t}^{t+4}) + Ea_2(\{\mathbf{Z}_{h,s}\}_{s=t}^{t+4}|\{H_{h,s}\}_{s=t+1}^{t+4}) + a_3\phi_t \ln k_{h,t} \\ & + a_1(\{\varepsilon_{h,s}\}_{s=t}^{t+4}) + \left[a_2(\{\mathbf{Z}_{h,s}\}_{s=t}^{t+4}) - Ea_2(\{\mathbf{Z}_{h,s}\}_{s=t}^{t+4}|\{H_{h,s}\}_{s=t+1}^{t+4})\right], \end{aligned} \tag{28}$$

where the expression on the last line is the composite disturbance term, which is mean-independent of $\mathbf{Z}_{h,t}$ and $\{H_{h,s}\}_{s=t+1}^{t+4}$. In this case, OLS estimation of (28) yields consistent estimates of the parameters characterizing the function a_1 and the combination of parameters

$$\boldsymbol{\delta}_{t+4},\ \phi_{t+4}\boldsymbol{\delta}_{t+3},\ \phi_{t+4}\phi_{t+3}\boldsymbol{\delta}_{t+2},\ \phi_{t+4}\phi_{t+3}\phi_{t+2}\boldsymbol{\delta}_{t+1},\ \phi_{t+4}\cdots\phi_{t+1}\boldsymbol{\delta}_t,\ t=3,...,12.$$

We have a reasonably large number of children in each age group $t = 3, ..., 12$ in 1997. This means that as we vary the initial t, we have consistent estimates of $\boldsymbol{\delta}_7, \boldsymbol{\delta}_8, ..., \boldsymbol{\delta}_{16}$, among other combinations of parameters. The $\boldsymbol{\delta}_s,\ s = 7, ..., 16$, are parametric functions of s. Ignoring the dependence on parental education for the moment, they are functions of only two parameters for each input, since the production parameter associated with an input j ($j = 1, ..., 4$) at age t is

$$\delta_{j,t} = g(d_{j,0} + d_{j,1}t),$$

where $g(x)$ is a known, continuous and monotonic transformation function ($g(\cdot)$ is assumed to be an exponential function in our specification). Having consistent estimates for $\delta_{j,t}$, $t = 7, ..., 16$, allows us to consistently estimate $d_{j,0}$ and $d_{j,1}$ for each input j. The same argument applies with respect to the identification of the scalar parameters, ϕ_s, since these are specified as a two-parameter function of s. The $\ln R_s$ process is specified as a four-parameter function of s. Given the consistent estimates of the ϕ sequence, these four parameters are identified as well. In fact, all of the parameters in the production technology are over-identified under our functional form assumptions.

D.2.2 Household Utility and Child Discount Factor Parameters

In this subsection, we make the argument that under certain ideal conditions regarding data availability, all preference parameters can be nonparametrically identified. This is due to the simple structure of the choice problem and the fact that, given all of the state variables of the problem, including household preferences, the actions of the household are uniquely determined. Our argument is slightly complicated by the possibility that the household may be behaving "non-cooperatively" or "cooperatively" through the use of a CCT, and there is only limited information in the data that indicates which behavioral rule the household is utilizing, through the CDS data on conditional allowances. We will begin by assuming that all households are either never choosing to use a CCT, or always choosing to use a CCT.

The choices of household h in period t are summarized by the vector $\mathbf{a}_{h,t}$. The investment processes that determines the sequence of cognitive skills $\{k_{h,s}\}_{s=1}^{M+1}$ and child discount factors $\{\beta_{h,c,s}\}_{s=1}^{M+1}$ are the only endogenous dynamic processes in the model, so that the decision rules determining time allocations in the household in any period s, $s = 1, ..., M$, are very simple to characterize given knowledge of the state variables in each period s,. In particular, conditional on a binary CCT choice, none of the household's remaining decisions directly depend on the value of $k_{h,s}$ in any period. As we outline in Appendix B, there is a unique mapping between household choices $\mathbf{a}_{h,t}$ and the state variables of the problem in period t, $\Gamma_{h,t} = (w_{h,1,t}\ w_{h,2,t}\ I_{h,t}\ k_{h,t}\ \beta_{h,c,t}\ \beta_{h,p}\ \alpha)$, which includes exogenous state variables (i.e., wages, non-labor income, and the parents' discount factor), endogenous state variables (i.e., $k_{h,t}$ and $\beta_{h,c,t}$), and other household preference parameters denoted by the vector (α). We can denote this mapping as $\mathbf{a}_{h,t} = a^*(\Gamma_{h,t})$. If all actions $\mathbf{a}_{h,t}$ were observable in period t, as well as exogenous state variables, $w_{h,1,t}$, $w_{h,2,t}$, $I_{h,t}$ and $\beta_{h,p}$, and the child's discount factor $\beta_{h,c,t}$, then $k_{h,t}$ need not be observed (since period t decisions are not a function of $k_{h,t}$ conditional on the binary CCT choice), and a^* is invertible in terms of the unknown preference parameter vector α, with

$$\alpha = (a^*)^{-1}(\mathbf{a}_{h,t}|w_{h,1,t}\ w_{h,2,t}\ I_{h,t}\ \beta_{h,c,t}\ \beta_{h,p}).$$

Indeed, given sufficient data we could even allow for preference heterogeneity across households, and identify α_h non-parametrically for each household h. Given our fairly limited (and noisy) data, imposing instantaneous preference homogeneity (i.e., assuming $\alpha_h = \alpha$ for all h) is not required but greatly simplifies identification. The argument for the invertibility of this function follows Del Boca and Flinn (2012). Given this invertibility in terms of α, observations from only one period for all sample households can be used to back out the vector α.

In point of fact, the PSID and the CDS do not include information on all of the household's choices in any period t. Given the CDS data, we do have an observation on all of the time allocations, including hours worked of the parents, time parents spend with the child, and the child's time "investing" by herself. We do not have accurate direct information on the private consumption of the parents and the child, $c_{h,t}$ and $x_{h,t}$, respectively. In order to simplify the computation of the model solution and to enhance identification of the model, we have assumed that there are no capital markets available for borrowing or saving. This means that total expenditures in the period are equal to total income, so that

$$x_{h,t} + c_{h,t} + e_{h,t} = w_{h,1,t}h_{h,1,t} + w_{h,2,t}h_{h,2,t} + I_{h,t}.$$

Since we observe all of the arguments on the right-hand side, and since we only have measures of total child expenditures ($e_{h,t} + x_{h,t}$) from external CEX data (see Section 3), we do not observe the child's private "consumption" and "investment" components separately within each household h.

We do not use the noisy information on household expenditures provided in the PSID-CDS in estimating the model, which increases the need for parametric restrictions. We did not use this information because we thought that the reported values were far too low to be believable. In our opinion, this is due to respondents under-reporting indirect expenses associated with the child's presence in the household, such as expenditures on housing and food. In our accounting framework, such expenses would be considered investments in the child. Moreover, we are not explicitly considering the additional problem that the PSID information on the wages, work hours, and non-labor income of the parents is not always collected in the same year as is the information taken from the CDS, which includes the time allocations in investment. The first CDS interview is conducted in 1997, whereas the other information taken from the core survey of the PSID is available in 1996 and 1998. Our model of wages and non-labor income allows for life-cycle changes. And given wages, work hours, and non-labor income are reasonably stable over such a brief time period, the fact that the information is not collected at exactly the same point in time should not cause major problems in inferring the stable preference weights associated with the arguments of the parents' altruistic utility function.

By our assumptions regarding the time endowment of the parents and child and the types of activities to which time can be devoted, we can infer the quantity of leisure consumed by parents and the child as the time remaining after all other activities have been accounted for. If $x_{h,t}$ and $e_{h,t}$ were observed for all households, and if all parents in the population of intact households with children supplied positive amounts of time to the labor market in some period t (so that $h_{h,1,t} > 0$ and $h_{h,2,t} > 0$), then all preference parameters would be nonparametrically identified. Since we only observe $x_{h,t} + e_{h,t}$ and we also allow for corner solutions in labor supply, preferences are not nonparametrically identified. Instead, we assume homogeneous period-specific preferences but allow for heterogeneity in the discount factors, to highlight the main novel features of the model.

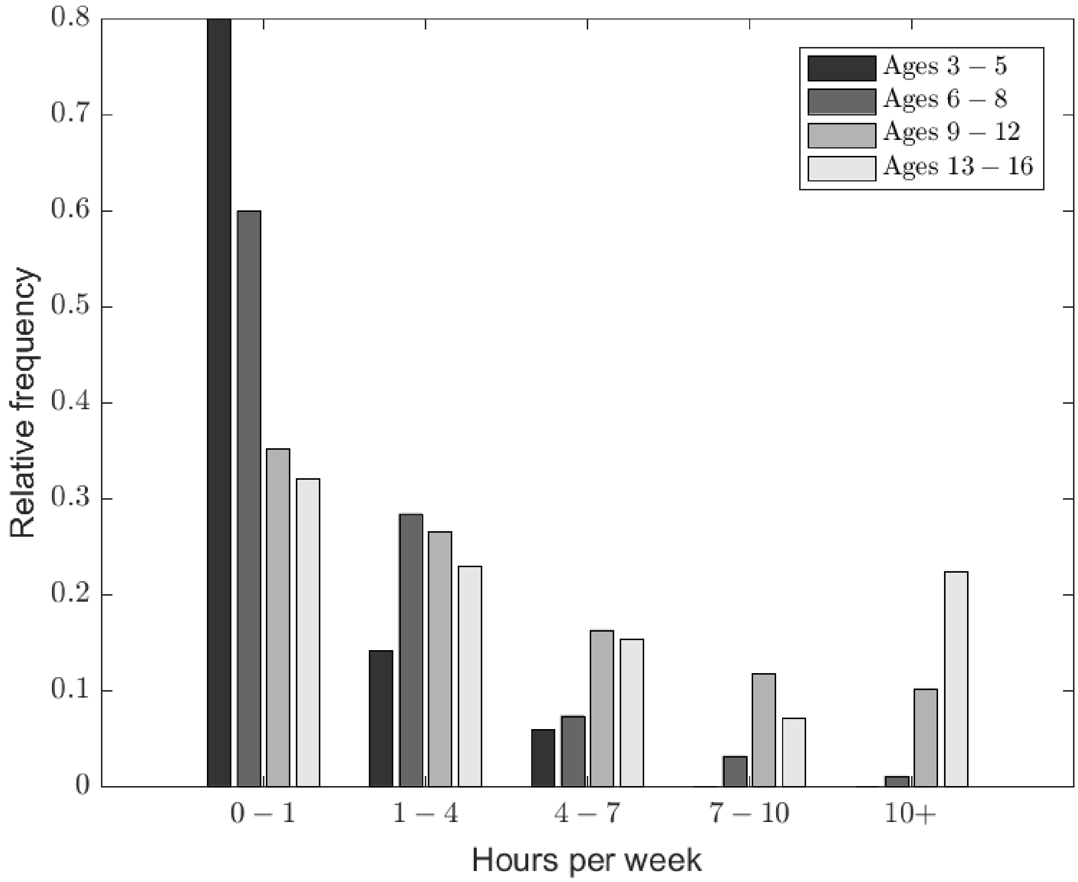
E Additional Tables and Figures

Table E.1: Data Correlations

Child Ages	**9-12**	**13-16**
Letter Word Score, Child time	0.261	0.156
	(0.069)	(0.071)
Letter Word Score, Mother's Educ.	0.245	0.265
	(0.068)	(0.068)
Letter Word Score, Father's Educ.	0.301	0.342
	(0.067)	(0.066)
Letter Word Score, HH Income	0.325	0.287
	(0.076)	(0.077)
Child time, Mother's Educ.	0.069	0.175
	(0.072)	(0.071)
Child time, Father's Educ.	0.102	0.285
	(0.071)	(0.069)
Child time, HH Income	0.141	0.294
	(0.082)	(0.079)

Source: PSID-CDS combined sample from 1997, 2002 and 2007 interviews and PSID core data between 1986 and 2010. To alleviate the missing data problem at young child ages, "HH income" is defined as the average total household income within each relevant child age bin. Standard Errors of the correlations are between brackets, and are defined as $SE_r = \sqrt{\frac{1-r^2}{n-2}}$.

Figure E.1: Distribution of Child Self-investment Time by Age

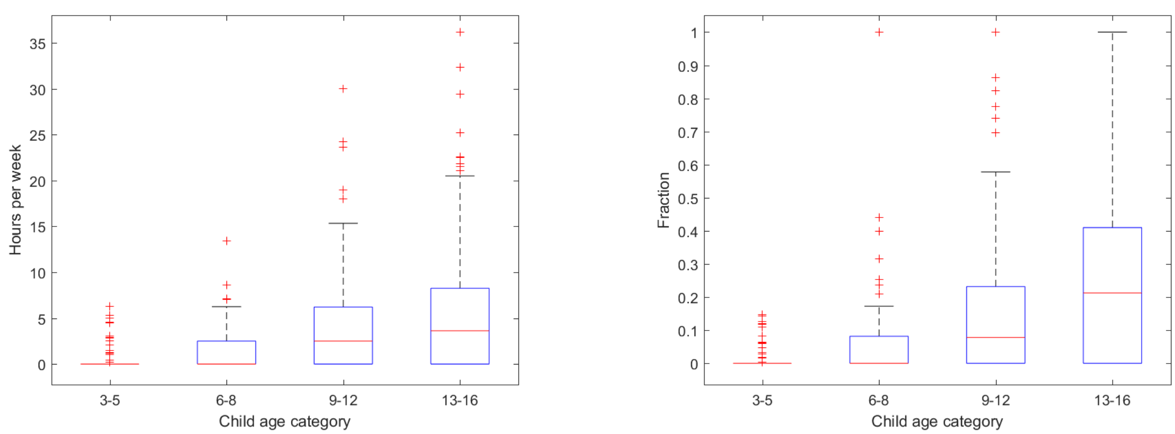


Source: PSID-CDS combined sample from 1997, 2002 and 2007 interviews.
Notes: Within each child age category, the vertical bars represent the fraction of households whose reported child self-investment time was between $0-1$ hours, $1-4$ hours, $4-7$ hours, $7-10$ hours, or more than 10 hours per week.

Figure E.2: Boxplots of Child Self-investment Time by Age

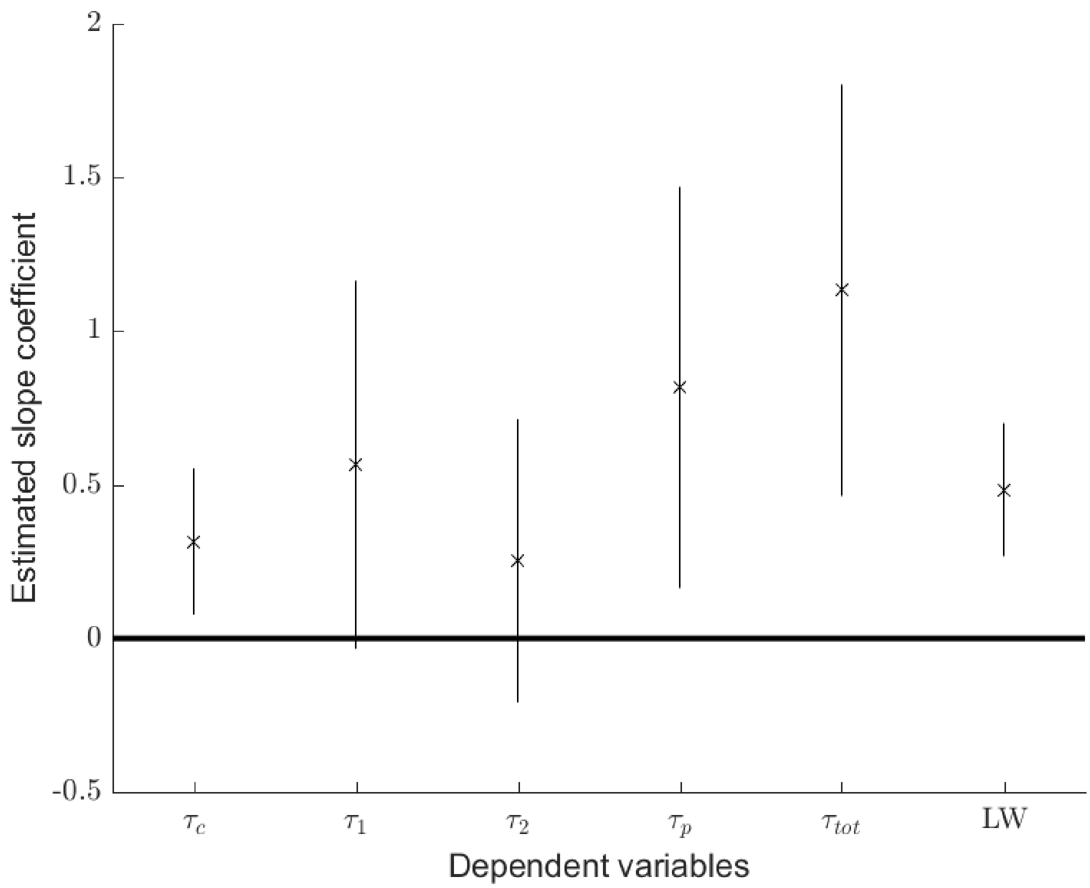
(a) Hours per week

(b) Fraction of total investment time



Source: PSID-CDS combined sample from 1997, 2002 and 2007 interviews.
Notes: The left panel plots the distribution of the reported weekly child study time for each child age category. The right panel shows child study time as a fraction of total investment time, defined as the sum of child study time and all active time with either or both of the parents.

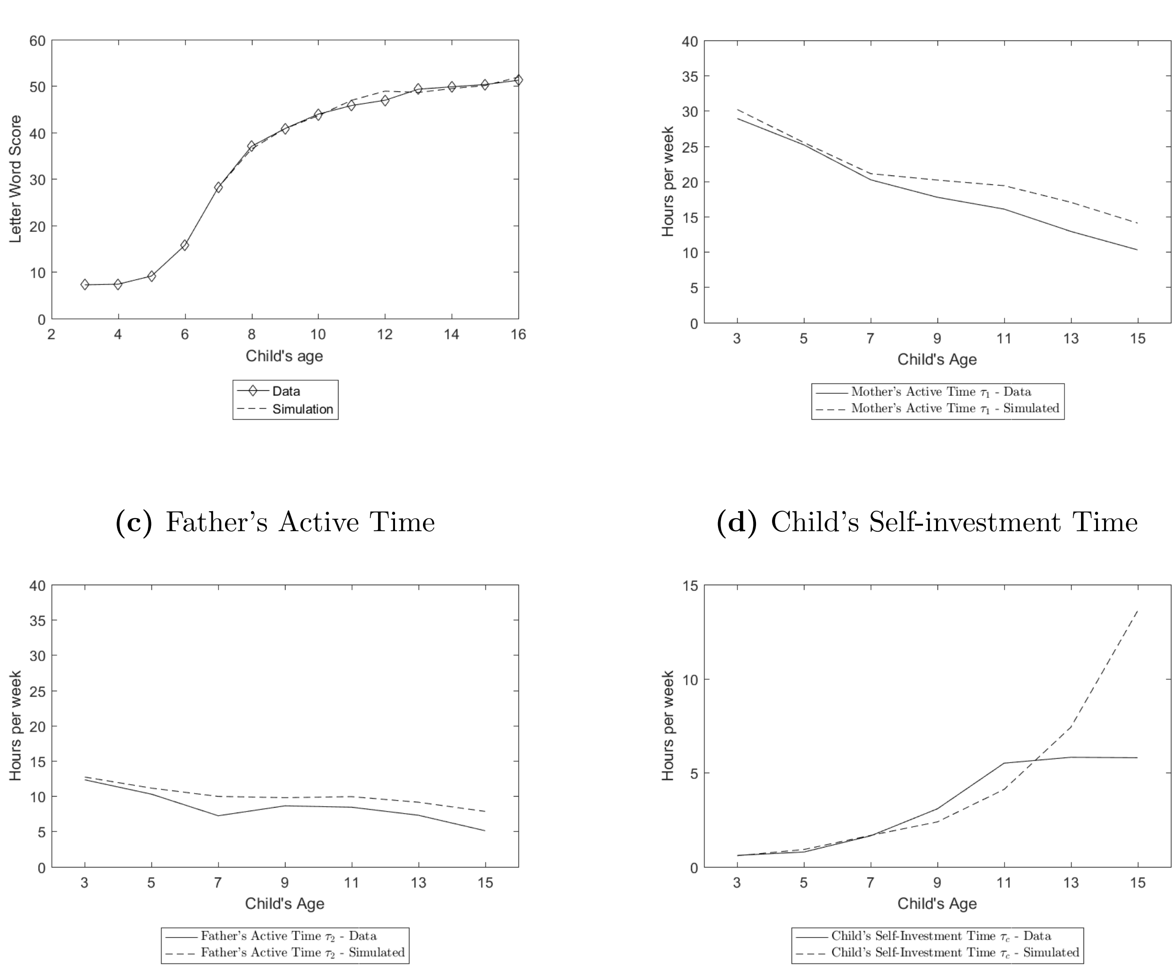
Figure E.3: The Effect of Parental Education on Productive Time Inputs and Test Scores



Source: PSID-CDS combined sample from 1997, 2002 and 2007 interviews and PSID core data between 1986 and 2010.

Notes: We regress various weekly time inputs and test scores on parental education (as proxied by the father's years of education). All regressions also control for maternal education, household income, and child age fixed effects. We plot the estimated slope coefficients on father's education and their corresponding 95% confidence intervals. The dependent variables are (from left to right): (1) the child's self-investment time, τ_c, (2) mother's active time, τ_1, (3) father's active time, τ_2, (4) total parental time, $\tau_p = \tau_1 + \tau_2$, (5) total investment time, $\tau_{tot} = \tau_c + \tau_p$ and (6) the child's raw Letter Word score, LW.

Figure E.4: Model Fit: Child Test Scores and Productive Time Inputs by Child Age



(a) Child's Letter Word Score

(b) Mother's Active Time

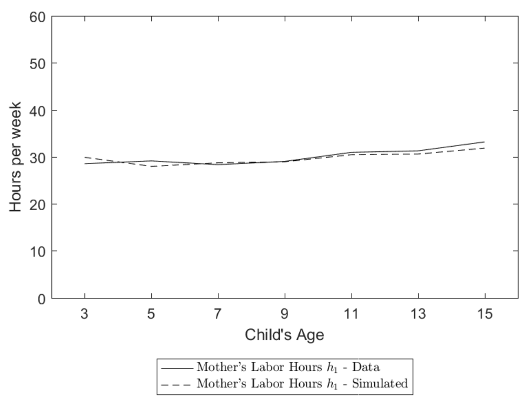
(c) Father's Active Time

(d) Child's Self-investment Time

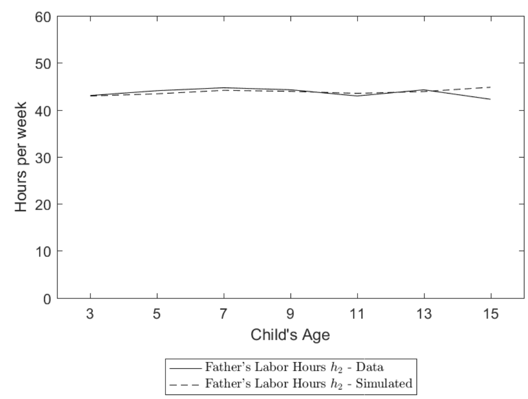
Notes: Data is actual data. Simulated is the model prediction at estimated parameters given above.
Source: PSID-CDS combined sample from 1997, 2002 and 2007 interviews and PSID core data between 1986 and 2010.

Figure E.5: Parental Labor Supply and LFP by Child Age

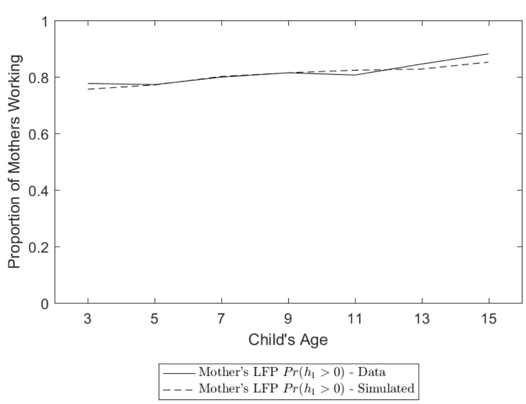
(a) Working Mother's Labor Supply



(b) Working Father's Labor Supply



(c) Mother's Labor Force Participation



(d) Father's Labor Force Participation

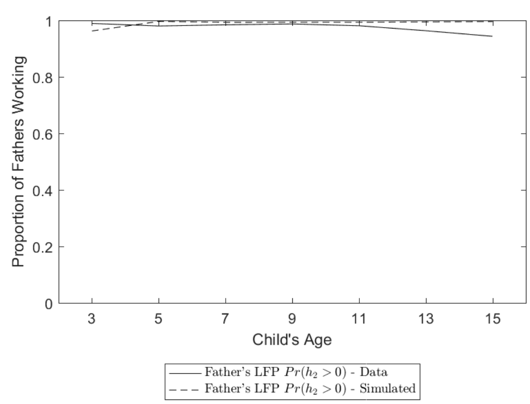
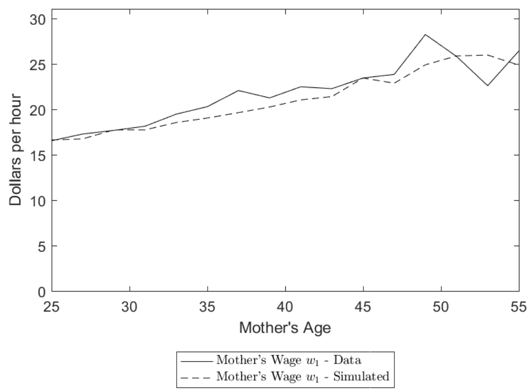
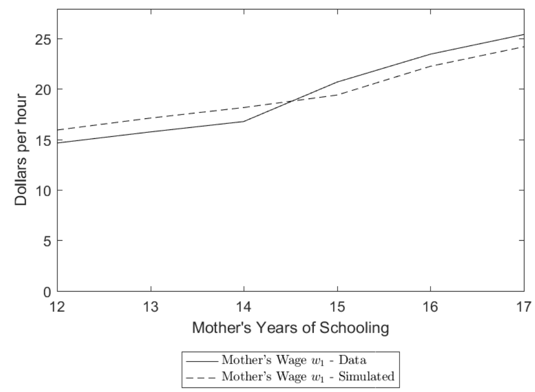


Figure E.6: Hourly Wages and Weekly Non-labor Income by Parental Age and Education

(a) Mother's Hourly Wage, by Age

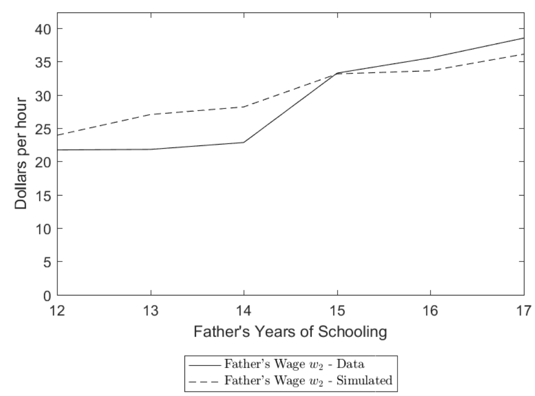


(b) Mother's Hourly Wage, by Education



(c) Father's Hourly Wage, by Age

(d) Father's Hourly Wage, by Education

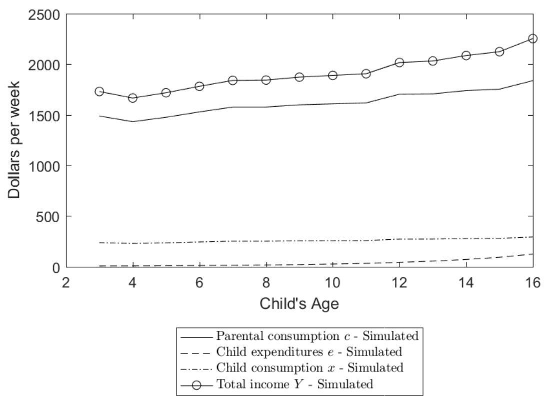


(e) Non-Labor Income, by Father's Age

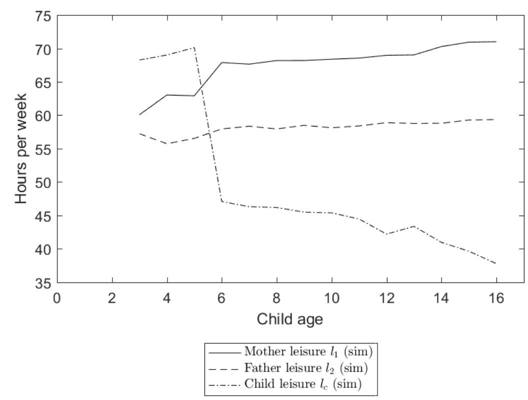
(f) Non-Labor Income, by Father's Education

Figure E.7: Expenditures, Leisure and CCT Use by Child Age

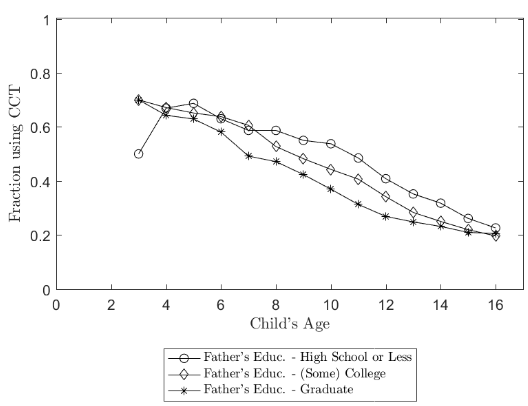
(a) Consumption, Expenditures and Income



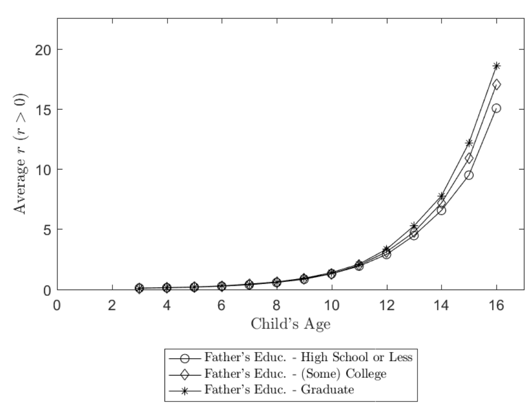
(b) Household Leisure Time



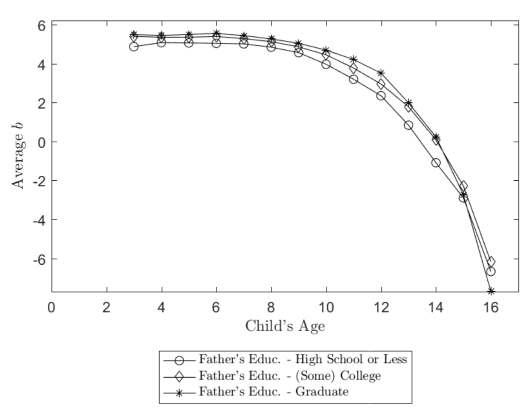
(c) Fraction of households using CCT



(d) Avg. Reward Elasticity r (if $CCT = 1$)



(e) Avg. Reward Intercept b



(f) Avg. Child Consumption x