Essays on Microeconometric Dynamic Public Finance

by

Sahber Ahmadi Renani

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Economics)

at the
University of Wisconsin–Madison
2023

Date of final oral examination: 20/1/2023

The dissertation is approved by the following members of the Final Oral Committee:

Matthew Wiswall, Professor, Economics
Naoki Aizawa, Assistant Professor, Economics
Corina Mommaerts, Assistant Professor, Economics
Anita Mukherjee, Assistant Professor, Risk and Insurance

Dedication

To my family, for their invaluable support.

Acknowledgments

I am grateful to my advisors Matthew Wiswall, Naoki Aizawa, and Corina Mommaerts for their estensive guidance. I thank all professors at the economics department of UW Madison, specially Jeffrey Smith, Christopher Taber, Jesse Gregory, John Kennan, and Job Boerma for their helpful comments. I am also thankful to Anson Zhou, Annie Lee, Yutang Chen, Macro Duarte, Yunhan Shin, John Stromme, Adibah Abdulhadi, and Alan Yang for their productive discourse.

Abstract

This dissertation studies two questions in dynamic public finance using microeconometric life-cycle models.

The first chapter explores the optimal design of the tax and transfer system for families. In this paper, I develop and estimate a rich microeconometric life-cycle model and solve for the optimal child-dependent tax function. The model features endogenous family labor supply, consumption-saving, and the development of children's cognitive ability in the family. First, I utilize the estimated model to evaluate the long-run effects of making the expanded child tax credits of the American Rescue Plan permanent. It has a modest negative effect on labor supply and increases children's test scores at age 18 by 2 percent of a standard deviation. The ex-ante welfare gain from this policy is 14 percent higher than an unconditional cash transfer that costs the same. Next, I solve for the optimal child-dependent taxes using flexible linear spline functions. The optimal tax policy of families with one or two children consists of a large guaranteed income and earning subsidies with a high-tax phase-out region along with child tax credits at the top of the income distribution. For families without children, the optimal policy is an EITC-type transfer to low-income families with a drastic increase in marginal taxes of middle and high-income families compared to the status quo. The optimal tax policy enhances the social welfare by increasing non-working time of mothers, lowering consumption inequality, increasing consumption smoothing, and improving children's abilities.

The second chapter study an important aspect of pension systems. In most countries that use a defined-benefit system, two functions govern the amount of pension benefit together: the first function summarizes the history of earnings into one variable (the history of earrings function), and the second function calculates the benefits based on that variable (the pension benefit function).

History dependence of a pension system affects the timing of retirement and the level of consumption insurance of a retiree by governing how her pension benefit is influenced by the profile of lifetime labor market shocks. This is particularly important for individuals with low attachment to the workforce. Although there is a sizable literature on the optimal design of the pension benefit function, the design of the history of earnings function is an understudied area of research. In this paper, I address this issue by exploring how different ways to summarize the history of earnings affect workers. I present some new stylized facts regarding the lifetime earnings of workers over the life cycle and show how different ways to summarise this lifetime history of earnings would affect workers' pension benefits. Next, I estimate a dynamic model of labor supply, saving, and retirement with different labor market shocks and study a counterfactual policy that changes the current US pension policy which uses the top 35 years of earnings to account for the lifetime earnings, common in other OECD countries.

Contents

List of Figures

List of Tables

Chapter 1

Children and Optimal Taxation of Families over the Life Cycle

1.1 Introduction

How much tax a family pays in the US depends on how many dependent children they have. Child-dependency of tax and transfer systems is a common feature across many countries[1]. There are numerous open questions regarding the design of this part of the tax system. How much should the amount of tax credits be for each additional child? Should middle and high-income families receive these benefits, or should they be only limited to low-income families? Should families with small children receive more tax credits than families with older dependent children? Debates around reforming the child tax credit (CTC), the earned income tax credit (EITC), and the dependent exemption show disagreement among policy-makers and researchers on answers to these questions. The goal of this paper is to shed light on these questions by exploring the design of the optimal tax as a function of the number of children and income of families. I do so by emphasizing two main motivations for child-dependent taxation:

[1]For instance, Canada Child Benefit (CCB) provides families up to $ 6,496 (in Canadian $) per year per child under the age of 6, and up to $ 5,481 per year per child aged 6 to 17. In UK, the new Universal Credit system offers £ 282.50 monthly allowance to a parent of a child lower than 16 years old. In Germany, the child tax exemption (Kinderfreibetrag) can make the total child tax exemption amount as high as € 7,356.

First, the number and ages of children are correlated with unobservables such as marginal utility of expenditure and female labor supply elasticity. Conditioning taxation on these variables helps the social planner to redistribute more with less distortion of labor supply. To distinguish between the two different types of tagging (Akerlof (1978)) that conditioning taxes on children present, I call this motivation for child-dependent taxes *static tagging*.

Second, consider the life cycle of a family who will have a child at some point in their life. This usually happens when they are young and have lower wages and assets, especially for couples without higher education. The borrowing constraint limits the ability of these parents to smooth consumption and labor over their life. The social planner can help families to relax this constraint by reallocating resources from the childless period of their life to when they have children by using child-dependent taxes. This *dynamic tagging* motivation does not rely on the heterogeneity of households, but rather on the heterogeneous states of a household in different parts of their life[2].

A family's income also shapes children's abilities (Dahl and Lochner (2012)) that are major determinants of their future education, labor market, and social life outcomes (Heckman et al. (2006)). Families might under-invest in their children relative to the socially optimal level due to borrowing constraints or low permanent incomes. The government has the ability to provide insurance against this risk of families by redistributing income across the life cycle of a family (dynamic tagging) and between families (static tagging).

In this paper, I develop and estimate a life-cycle model that takes into account the above-mentioned mechanisms. My model features endogenous family labor supply, consumption-saving, and development of children's ability in the family. It also includes heterogeneity in altruism among families and uncertainties in wages, fertility, and the child-development process over the life cycle. I utilize data from the Panel Study of Income Dynamics (PSID) and its Child Development Supplement (CDS) to estimate the model using the method of simulated moments. I make use of indirect inference type moments to clearly identify the cognitive ability production function and the preference parameters regarding child investments. The estimated model is able to match the key features of households' assets, spousal labor supply, and cognitive test scores of children of

[2]The same motivation is present in age-dependent taxation as in Weinzierl (2011).

families of different sizes over their life cycle.

By comparing different elasticises, I show there are sizable possible benefits from child-dependent taxes due to static tagging. The labor force participation elasticity of females with a child is 64 % more than females without children. Also, given the higher marginal utility of expenditure of families with children, redistribution to these families increases the social welfare more when compared to the families without children. I also find that dynamic tagging is a major consideration in child-dependent taxation. In response to a permanent increase in wages, families increase their total expenditure much more when they have children compared to the other stages of their life. Furthermore, the labor force participation of females and males are also more responsive when their families have a child present in their home.

I use the estimated model to evaluate the effects of a tax policy currently under discussion among policymakers: permanently keeping the expanded child tax credit part of the American Rescue Plan. Compare to the economy with the previous level of child tax credits, I find that this policy increases welfare by 4.17 % in consumption equivalence. Families who have 2 children in life gain the most from this policy, but the welfare gain is positive for families of all sizes. This policy has a modest effect on the labor supply. It decreases female labor force participation by 2.47 % and hours worked by men by 0.65 %. The labor supply effects are close to those in Corinth et al. (2021) which uses the common estimates of elasticities in the literature to predict the behavioral responses of this policy. The expanded child tax credits increases the mean of children's final Letter-Word score by 0.14 % (0.02 standard deviations) and decrease its standard deviation by 1.67 %. To compare the welfare gain of this policy with other policies of the same costs, I evaluate the effects of an unconditional cash transfer to all families that costs the same as the expanded child-tax credits of the American Rescue Plan. I find that the welfare gain of the unconditional cash transfer is more homogeneous across family types, but the ex-ante welfare gain is 0.54 percentage points lower in consumption equivalence.

Next, by approximating the tax schedule with linear splines functions, I solve for optimal child-dependent taxes. My results signify the importance of child dependency in a tax system as the tax treatment of families with and without children are quite different. Under a utilitarian social planner, the optimal tax policy of families with one or two children consists of both a large guaranteed income (around $ 7,000 for a family with 1 child) and high earning subsidies (around 1.9 substitution rate) until the family's income reaches around $ 20,000. After that, the transfers will be taxed at a very high rate (around 50 %). The transfers are still positive and large for middle-income families with children with a break-even income of around $ 75,000. For families without children, the optimal policy is an EITC-type transfer with around 1.5 substitution rate by the government to low-income families, accompanied by a drastic increase in the marginal taxes of middle and high-income families compared to the status quo. Although the average taxes of high-income families with children also rise compared to the status quo, the increase is much smaller than the increase for families without children. This suggests that high-income families should also benefit from child tax credits. A couple with $ 150,000 in income who has one child will pay around $ 38,000 in taxes, around $ 13,000 less than a zero-child couple with the same income.

The welfare decomposition shows that most of the welfare gain from optimal child-dependent taxes comes from lower hours of work of mothers (the group with the highest disutility of work); as their labor force participation will go down by around 31 %. With 20 % decrease in the standard deviation of total expenditure, another main source of welfare gain is the large decrease in inequality in total expenditure of families and more consumption smoothing. Social welfare will also be increased by better allocation of investment in cognitive ability both across children and over the life of a child as the standard deviation of final test scores decreases by 5 % . The welfare of families who never have any children over their life will be around 11 % lower, while families who have 2 children over their life will gain around 12 %. The welfare of families who have one child in life increases by 0.2 %. The ex-ante welfare gain is 11.96 % in consumption equivalence.

Under the optimal tax policy, the utilitarian social planner provides a large transfer to low and middle-income families with children to compensate for the extra costs associated with having children. This rise in redistribution increases social welfare, but it is not without costs. First, the horizontal equality principle will be violated as in any tagging policy. Also, note that the optimal policy does not represent a Pareto improvement. While families with children benefit from changing the tax system to the optimal tax policy, the couples who never have any child over their life experience significant welfare losses[3]. Second, while the EITC-type tax subsidy embedded in the tax system encourages work, especially on the extensive margin, the income effect of the large transfer causes a substantial decrease in the labor force participation of secondary earners. Since the disutility of work of females (especially mothers of small children) is large, the social planer is willing to accept this cost for the low-wage mothers, as the increase in non-working time of mothers will also increase welfare.

Similar to Saez (2002), with the inclusion of an extensive margin of labor supply, guaranteed income becomes less preferable and an income subsidy program becomes more important to encourage labor force participation at the bottom of the income distribution. This can be seen in the optimal tax policy for families without children. However, as the government wants to redistribute to families with children to compensate for their cost of child-rearing, it is more costly to do so if the mother participates in the labor market. Given the none-separability of female labor force participation and consumption in the model (as in Heckman (1974)), if a mother works, the marginal utility of expenditure of her family would be higher, as they have to buy formal child care and previously home-produced goods. Hence, if the mother does not work, it requires fewer transfers from the government to provide the family with the same level of utility. When the wage of a mother increases, the cost of her non-employment also increases. Given the high elasticity of labor force participation of mothers, the transfers have to be be taxed away to encourage labor supply. However, for dual-earner families with children, the higher marginal utility of expenditure along with the higher cost of children induces lower taxes on top, compared to families without children.

Since part of the non-working time of mothers is time spent with children, society may accept a large number of non-working mothers especially if they have small children (similar to long maternal leave policies). However, society may dislike this policy since it increases the number

[3]This is not a problem if society accepts to compensate families for the costs of childbearing. Based on an online survey, Saez and Stantcheva (2016) finds that ”Redistribution based on marginal utility is socially acceptable if there are objective reasons for a person to have higher needs, such as having a medical condition requiring high expenses, or a large family with many dependents.”. The fact that child-dependant taxation is common across many countries suggests that societies are willing to compensate for the costs of children at least partly. In the optimal tax problem, it is feasible to allow the social planner to only take into account heterogeneities among families which she considers worthy of different treatment (like child development and expenditure costs) and assume the rest of the utility function (like disutility of work) the same for all families (similar approach as a paternalistic social planner in behavioral public finance).

of non-working individuals on a large scale. I incorporate this concern by solving for the optimal taxation problem from the viewpoint of a social planner who does not value the utility of non-working time of individuals. Compared to the optimal tax policy with the utilitarian social welfare function, the transfers offered to low and middle-income families will be lower, but still sizable. However, the social planner will completely remove the income guarantee for families with zero income. The decrease in the female labor force participation rate is still noticeable. However, the decrease is 8 percentage points and 4 percentage points less for families with one child and two children, respectively.

There are several limitations to the results presented in this paper. Most importantly, inclusion of child-care tax credits can equip the social planner with a more work-friendly substitute to child tax credits and lower the need for large income transfers (Ho and Pavoni (2020)). The result presented here should be seen as the optimal taxation problem of a government with no access to child-care tax credits. The extent of changes in the results with inclusion of this instrument is left for future work[4]. In order to rigorously study the dynamics of households over their life cycle while keeping the model tractable, I also abstract from two other motivations that have been discussed in the literature. First, I ignore fertility responses to changes in taxes. The resulting bias from this omission is small in my counterfactual exercises around the current policy, but can potentially be sizable in the optimal tax problem given a large amount of transfers to families with children (i.e. Kurnaz (2021) and Zhou (2021)). Second, I am not considering the dynastic effects of child-dependent taxes. As Caucutt and Lochner (2012), Cunha (2013), and Lee and Seshadri (2019) report, the fiscal externality of child-related transfers in increasing the tax base of next generations can be large.

Related Literature: This paper has two main contributions. First, this paper bridges between two literatures: a large literature that studies family labor supply and consumption smoothing over the life cycle and how taxes and transfers affect families (i.e. ?? (Blu), Blundell et al. (2016), Blundell et al. (2018), Low et al. (2018), **?**, and Guner et al. (2020)); and a growing literature that investigates the development of children's abilities inside households and the effects of child-related

[4]Note that since wages are exogenous in the model, the cost of exclusion of mothers from the labor market is underestimated (Adda et al. (2017) and Blundell et al. (2016)). Taking into account on-the-job human capital accumulations of mothers lowers the optimal amount of guaranteed income and makes providing work incentives in form of wage subsidies and child tax credits more important.

transfers on the child-development process (i.e. Keane and Wolpin (2001), Bernal (2008), Cuhna et al. (2010), Boca et al. (2014), Agostinelli and Wiswall (2016), Verriest (2019), and Mullins (2019)). Considering both of these strands of literature together, this paper estimates a life-cycle model that takes into account family labor supply, consumption smoothing, and child development.

Second, the optimal dynamic taxation literature (i.e. Erosa and Gervais (2002), Conesa et al. (2009), Weinzierl (2011), Farhi and Werning (2013), Golosov et al. (2016), Heathcote et al. (2017)) ignores the effects of children on workers and the gains from the dependence of taxes on the number of children[5]. This paper adds to this literature by introducing dynamic gains from child-dependent taxes and solving for the optimal child-dependent tax problem in a rich mircoeconometric life-cycle model that takes into account the impact of children on labor supply and consumption decisions of individuals by utilizing flexible linear spline functions. Although several other papers also explore either optimal taxation problem in a child-development environment (i.e. Cigno and Pettini (2002), Cigno et al. (2003), and Mullins (2019)) or optimal taxation problem conditional on family size (i.e. Cremer et al. (2003) and Kurnaz (2021)), they do not model the dynamic of household's decisions over different stages of their life cycle[6].

This paper also relates to several other area of the literature. The effect of borrowing constraints and the ability of parents to insure against labor market shocks on children's outcomes has also been studied in OLG environments (i.e. Cunha (2013) and Caucutt and Lochner (2012)) and designed-based settings (i.e. Abbott (2017) and Carneiro and Ginja (2016)). Child equivalent scales and consumption-saving decisions of households with children over the life cycle have also been studied by Banks et al. (1992), Banks et al. (1998), and Bick and Choi (2013). However, they do not consider the child-development process and joint family labor supply decisions. A growing literature (i.e. Blundell and Shephard (2011), Aizawa (2017), **?**, Aizawa et al. (2019), and Parodi (2019)) utilizes structural microeconometrics models to solve various optimal policy problems. I

[5]Karabarbounis (2016) explores optimal taxation based on different tags including filing status, Wu and Krueger (2021) considers the effect of family labor supply decisions on optimal progressively of taxation, and Malkov (2021) studies optimal taxation of singles and couples. None study optimal taxation conditional on the number of children. For a survey of the optimal dynamic taxation literature, see Stantcheva (2020).

[6]Optimal taxation conditional on family size has also been explored in the context of the intergenerational insurance problem (i.e. Farhi and Werning (2007), Hosseini et al. (2013), and Stantcheva (2015)). This literature focuses less on the insurance problem over the life cycle, which is the question studied in this paper.

contribute to this approach to the optimal policy problem by extending the linear spline functions method as in Gayle and Shephard (2019) into a dynamic environment.

The remainder of this paper is organized as follows. In Section 2.5, I develop the empirical model and characterize the solution. Section 2.3 describes the data. In Section 1.4, I show the estimation methodology, identification argument, and the goodness of fit. In Section 1.5, I discuss the implications that the data and the estimated model have on the design of optimal child-dependent taxes. In Section 1.6, I compare the effects of some counterfactual tax polices. In Section 1.7, I formulate the optimal taxation problem and present the results. Section 2.8 concludes.

1.2 Empirical Model

1.2.1 Life Cycle and Demographics

The model starts at period 1, with a continuum of families. A family consists of a couple (a male (m) and a female (f)) both at age 22 with no children. The decision-making of the family follows a unitary model of the household and there is no possibility of divorce in the model. Time is discrete and each time period (t) represents one year. Spouses retire at age 62 ($t_r = 41$) and die at age 76 ($t_d = 55$). Each spouse is either college-educated or not ($edu^m, edu^f \in \{1, 0\}$). Families are indexed by i (which is excluded here for brevity), spouses are indexed by j, and (possible) children are indexed by k (in order of birth).

Each family can have up to 2 children. n_t denotes the number of children currently present in the household. Women have a fertility window between the ages of 22 and 37 ($t \leq T_{Fer} = 16$). They give birth to a child with the probability $Pr_t^b(n_t, edu^f, edu^m)$ that depends on their age, level of education, and current number of children. Children live with their families for their first 18 years of life. They leave the household at the end of age 17 ($0 \leq z_t^k \leq z_L = 17$).

1.2.2 Decisions

In each period before retirement, households make time and budget decisions. Households choose how to allocate their time between work (h_t^f, h_t^m) and leisure (l_t^f, l_t^m), where there are 9 discrete possibilities. Both members of the couple choose whether to not work, work part time, or work full time ($h_t^i \in \{0, h_P, h_F\}$).

The number of options among which a household can allocate their budget depends on the number of children in the household. All families, regardless of child status, decide how much to consume c_t and how much assets to leave for the next period a_{t+1}. Families with children also make a decision on how much to invest in the cognitive ability of each of their children ($e^k_{z^k_t}, k \leq n_t, z^k_t \in \{0, \cdots, z_L\}$). The cognitive ability of children ($\theta^k_{z^k_t}$) is dependent on these investments and develops according to a production function that will be explained in Section 1.2.6.

"Investment in the cognitive ability of children" is defined as the part of total expenditure of the family that increases the cognitive ability of the child. "Consumption expenditure" is defined as part of it that directly increase the utility of parents but has no effect on cognitive ability of the child. Although the same variable $e^k_{z^k_t}$ is used for all the ages, the model allows that the type of goods that count as $e^k_{z^k_t}$ changes by the age of the child. Since the coverage of $e^k_{z^k_t}$ in the data is poor, it is treated as unobservable. The final cognitive ability of the child $\theta^k_{z^L+1}$ is finalized a year after the last year of her presence in the household at age 18 ($Z_L + 1$). During retirement, a family's budget is only allocated to consumption or saving.

1.2.3 Preferences

The per-period utility of a family without any children at $t < t_r$ is

$$u_0 = \frac{1}{1-\sigma}\left[\mu^{h_f}(h_t^f)\frac{c_t}{\rho^0(n_t^{Le})}\right]^{(1-\sigma)} - \mu^{h_m}(h_t^m) + I_0. \tag{1.1}$$

The first term in u_0 shows a household's utility from consumption and disutility from female labor supply. σ stands for the coefficient of relative risk aversion of consumption and ρ^0 represents the consumption equivalent scale of a couple without children. I am taking into account the transfers of parents to children when they leave the household by allowing n_t^{Le} to affect this equivalent scale. μ^{h_f} shows the disutility cost of female labor supply and is assumed to be non-separable with the utility of consumption. One reason for this assumption is that when a wife participates in the labor market, it causes the household to substitute some of home-produced goods with market goods, thus it affects the marginal utility of consumption (c_t)[7]. In the second term, μ^{h_m} represents the disutility of work hours of the male[8]. I_0 is the utility from the final cognitive ability of a child that has left the home one year ago and will be explained after the discussion of the utility functions of families with children.

Children affect the utility function of parents through their effects on the marginal utility of consumption, disutility of labor supply of women, and altruism of parents toward their abilities. The per-period utility of a family for $t < t_r$ for families with one and two children are

$$u_1 = \frac{1}{1-\sigma}\Big[\frac{\mathbb{L}_1(h_t^f, z_t^k)\{c_t \cdot (e_t^k)^{\delta_1(z_t^k)}\}}{\rho^1(z_t^k, n_t^{Le})}\Big]^{(1-\sigma)} - \mu^{h_m}(h_t^m) + \mu^\theta[\ln(\theta^k_{z_t^k})] + I_1, \tag{1.2}$$

$$u_2 = \frac{1}{1-\sigma}\Big[\frac{\mathbb{L}_2(h_t^f, z_t^1, z_t^2)\{c_t \cdot (e_t^1 + e_t^2)^{\delta_2(z_t^1, z_t^2)}\}}{\rho^2(z_t^1, z_t^2)}\Big]^{(1-\sigma)} - \mu^{h_m}(h_t^m) + \mu^\theta \sum_{k=1}^{2}\Big[\ln(\theta^k_{z_t^k})\Big]. \tag{1.3}$$

where $\delta_1(z_t^k)$ and $\delta_2(z_t^1, z_t^2)$ are the shares of investment in children in utility of one-child and two-children families which depend on the public good natures of these investments and can vary by the age of the children. ρ^1 and ρ^2 are the consumption equivalent scales of families with one child and two children, respectively. These consumption equivalent scales depend on the age of children to capture the extra costs of children during their presence in the household. $\mathbb{L}_1(h_t^f, z_t^k) = \big(1 - \mu^{LP,K} \cdot \mathbb{I}(h_t^f > 0, z_t^k \leq 6)\big) \cdot \mu^{h_f}$ and $\mathbb{L}_2(h_t^f, z_t^1, z_t^2) = \big(1 - \mu^{LP,K} \cdot \mathbb{I}(h_t^f > 0, \min(z_t^1, z_t^2) \leq 6)\big) \cdot \mu^{h_f}$, where $\mu^{LP,K}$ represents the effect of the presence of children on the disutility from labor force participation of mothers[9].

μ^θ shows the altruism coefficient for children's cognitive ability. Parents value the final cognitive ability of a child by $\chi^\theta \mu^\theta(\ln(\theta^k_{z_L+1}))$, where χ^θ is the altruism scale coefficient of the final cognitive ability of the child. These final cognitive abilities are realized one year after the child leaves the home. $I_0 = \mathbb{I}\{n_{t-1} = 1\}\chi^\theta\mu^\theta(\ln(\theta^1_{z_L+1}))$ and $I_1 = \mathbb{I}\{n_{t-1} = 2\}\chi^\theta\mu^\theta(\ln(\theta^{-k}_{z_L+1}))$ in the utility functions of no

[7]See Heckman (1974) and Blundell et al. (1994) for the discussion of non-separability between consumption and female labor supply.

[8]μ^{h_m} and μ^{h_f} are non-parametric functions. If $h_m = 0$, then $\mu^{h_m} = 0$, while if $h_f = 0$, then $\mu^{h_f} = 1$.

[9]Presence of children can directly affect the disutility of labor force participation of women by increasing the utility of staying at home due to attachment of mothers to their children and by changing the value of home production of women (i.e. child-care).

and one child families show the utility of parents from the final cognitive ability of children one year after they leave the home[10].

There is no bequests in the model. Defining $\rho_r(n_t^{Le})$ to be the equivalent scale of consumption of a retiree couple, the utility of a family in each period t during retirement is

$$u_r(c_t) = \frac{1}{1-\sigma}[\frac{c_t}{\rho_r(n_t^{Le})}]^{(1-\sigma)}. \tag{1.4}$$

The only source of preference heterogeneity in the model is on the altruism coefficient $\mu^\theta \sim N(\overline{\mu^\theta}, \sigma^2_{\mu^\theta})$. The rest of the parameters are common across families.

1.2.4 Constraints

At each $t < t_r$, families face two time constraints

$$\begin{aligned} l_t^m + h_t^m &= \bar{L}, \quad h_t^m \in \{0, h_P, h_F\},\ l_t^m > 0 \\ l_t^f + h_t^f &= \bar{L}, \quad h_t^f \in \{0, h_P, h_F\},\ l_t^m > 0 \end{aligned} \tag{1.5}$$

and a budget constraint

$$a_{t+1} = R\Big(a_t + y_t - T_t - TSS_t + SNAP_t - c_t - \mathbb{I}\{n_t > 0\}\sum_{k=1}^{n_t} e^k_{z^k_t}\Big), \quad a_{t+1} \geq 0. \tag{1.6}$$

Where $\bar{L}$ in the time constraints is the total hours available for each spouse during the year. The sum of income of the female ($y_t^f = h_t^f \omega_t^f$) and the male ($y_t^m = h_t^m \omega_t^m$) constitute the total family income $y_t = y_t^m + y_t^f$, where ω_t^j is the wage of spouse j. T_t is the total amount of tax paid by the household that is determined by the tax function $\mathbb{T}(y_t, n_t)$ that takes the total family income and the number of children present in the household as inputs. TSS_t is the sum of social security tax payments by both spouses, which for each of them is determined according to the social security tax function $\mathbb{TSS}(y_t^j)$. $SNAP_t$ represents the amount of transfer from the Supplemental Nutrition Assistance

[10]Note that in the utility function of one-child families, $k = 1$ if this is their first child and $k = 2$ if this is their second child and the first child has left the home.

Program (SNAP) program (formerly known as the Food Stamp Program). I assume that all eligible families receive food stamp and it is identical to an income transfer[11], and is determined by the food stamp function $\mathbb{SNAP}(y_t, n_t)$, which depends on the total income and the number of children[12]. R is the absolute rate of return to assets and I assume that households cannot borrow against their future income and transfers. (1.6) also assumes that households can not borrow against their future income and government transfers[13].

The budget constraint during retirement is

$$a_{t+1} = R\Big(a_t + PPB - c_t\Big), \quad a_{t+1} \geq 0, \tag{1.7}$$

where PPB is the sum of public pension benefits of the couple during retirement[14].

1.2.5 Wage Processes

For spouse j at period t, his or her wage follows

$$\ln(\omega_t^j) = \beta_0^j + \beta_1^j t + \beta_2^j t^2 + \beta_3^j(edu_t^j) + \beta_4^j(t) \times (edu_t^j) + \xi_t^{\omega,j}, \tag{1.8}$$

$$\xi_t^{\omega,j} = \xi_{t-1}^{\omega,j} + \iota_t^{\omega,j}, \; if \quad t > 1, \tag{1.9}$$

[11] For evidence, see Hoynes and Schanzenbach (2009).

[12] Although the main goal of this paper is to study the labor income tax function, there are two reasons for including other government programs. First, to better understand the budget of low-income families, only considering the tax function is not enough. As the marginal taxes on low and middle income families are low (beside the child-dependent parts), the social security tax has a non-trivial effect on both their budget and their labor supply decisions. I also include SNAP since it is the biggest income floor program in the US, and the only one that is also available to no-child families. Other income floor programs (like TANF) are smaller in size and given that their eligibility rules are history dependent, are computationally less tractable. Second, changing the tax schedule also changes the distribution of pre-tax income due to behavioral responses of families. Inclusion of these programs helps to take into account the fiscal externally that arises from a change in the tax function and confirms that the government budget has not changed under the optimal tax schedule.

[13] The main motivations for saving in this model are for precautionary reasons, extra costs of children, and retirement. This no-borrowing constraint has bite in reality if the amount of transfers to families with children are large and households can not borrow against those transfers when they do not have children. On the other hand, as there is a large number of households with debt in the data, the no-borrowing constraint underestimates the amount of private insurance available to households and overestimates the need for the government-provided consumption issuance.

[14] Given that I am not modeling the individuals' social security payment history, I approximate PPB by the average pension benefit of the couple's education group.

where $\xi_t^{w,j}, j \in \{m, f\}$ are permanent wages of spouses that follow AR1 processes and $\iota_t^{w,j}$s are i.i.d shocks to permanent wages that follow a joint normal distribution

$$\begin{bmatrix} \iota_t^{\omega,f} \\ \iota_t^{\omega,m} \end{bmatrix} \sim \mathbb{N}(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\iota_{\omega,f}} & \rho_{\iota_{\omega,f,m}} \\ \rho_{\iota_{\omega,f,m}} & \sigma^2_{\iota_{\omega,m}} \end{bmatrix}). \tag{1.10}$$

Initial permanent wages also follow a joint normal distribution. The deterministic part of each wage process depends on age, and education and individuals with different education levels can have different wage trajectories over the life cycle.

1.2.6 Technology of Cognitive-Ability Formation

Each child is born with innate cognitive ability $\ln(\theta_0^k) \sim N(\overline{\theta_0}, \sigma^2_{\theta_0})$[15]. After birth, the cognitive ability of a child develops over her first 18 years of life. It evolves from her innate ability (initial endowment) and grows depending on parental investments and developmental shocks

$$\ln \theta^k_{z^k_{t+1}} = \alpha^z(z^k_t) + \alpha^\theta(z^k_t) \ln \theta^k_{z^k_t} + \alpha^e(z^k_t) \ln e^k_{z^k_t} + \xi^{\theta,k}_{z^k_t} + \epsilon^{\theta,k}_{z^k_t}, \tag{1.11}$$

$$\xi^{\theta,k}_{z^k_t} = \xi^{\theta,k}_{z^k_{t-1}} + \iota^{\theta,k}_{z^k_t} \quad \textit{if} \quad z^k_t > 0, \quad \xi^{\theta,k}_0 = 0, \quad \iota^{\theta,k}_{z^k_t} \sim \mathbb{N}(0, \sigma^2_{\iota_\theta}), \tag{1.12}$$

$$\epsilon^{\theta,k}_{z^k_t} \sim \mathbb{N}(0, \sigma^2_{\epsilon_\theta}). \tag{1.13}$$

$\alpha^z(z^k_t)$, $\alpha^\theta(z^k_t)$, and $\alpha^e(z^k_t)$ are (age-of-child-dependent) coefficients of total factor productivity (TFP), current cognitive ability, and investment in cognitive ability, respectively[16]. $\xi^{\theta,k}_{z^k_t}$ is the permanent cognitive ability, $\iota^{\theta,k}_{z^k_t}$ is the i.i.d shock to permanent cognitive ability, and $\epsilon^{\theta,k}_{z^k_t}$ is the i.i.d

[15] In the current model, the innate ability of children do not depend on any characteristic of parents (like education or income). However, parental characteristics can affect the distribution of innate ability through the transfer of genetic endowments and maternal well-being during the fetal period. Conditioning the distribution of innate ability to some characteristic of parents will be incorporated in future work.

[16] I assume that a child cognitive ability follows a Cobb-Douglas production function. One limitation of this production function is its restriction on the elasticity of substitution. Considering a production function with a flexible substitution pattern is not computationally feasible in my model, since then I would have to track the cognitive ability of children over the life cycle of parents. Agostinelli and Wiswall (2016) estimate a trans-log production function of cognitive ability of children. The coefficient on the interaction of cognitive ability and investment is not statistically significant in any of their specifications. While the definition of investment in children is different in their model than mine, this result suggests that the lack of flexibility in substituting patterns with a Cobb-Douglas production function is less of the concern.

transitory shock to cognitive ability. Persistent shocks represent the shocks in the environment of the child that persist over time (like home, school, and neighborhood) as well as non-cognitive ability[17] of the child that are not modeled here and assumed to evolve stochastically[18]. The transitory shocks reflects the shocks in the life of the child that may only affect her in a given year (like teacher quality or medical issues).

1.2.7 Household Problem

In this section, I present some of the value functions of the household at different stages of life. The rest of them are provided in Appendix 1.B.1. I define $\Omega = \{edu^f, edu^m, \mu^\theta\}$ to be the vector of permanent heterogeneity which consists of the education levels of both spouses and the altruism coefficient for cognitive ability of possible children. Also in the value functions, β represents the subjective discount factor.

Problem of Families without Children

Families without children simply decides on consumption and hours of work of spouses, given their constraints. While the female's age is in the fertility window, state variables are $\{a_t, \mathsf{W}_t\}$, corresponding to a family's assets and wage shocks $\mathsf{W}_t = \{\xi_t^{w,f}, \xi_t^{w,m}, \iota_t^{w,f}, \iota_t^{w,m}\}$. In these periods, in addition to uncertainty about future innovations on permanent wages, a family faces uncertainty about having a child in the next period, and conditional on having a child, uncertainty about her initial cognitive ability and initial permanent and transitory cognitive ability shocks. The value function of a family without any children at $t < T_{Fer}$ is

$$V_t^{0C,F}(a_t, \mathsf{W}_t; \Omega) = \max_{c_t, h_t^f, h_t^m} \Big\{ u_0(c_t, h_t^f, h_t^m; \Omega) + \beta \Big[(1 - Pr_t^b(edu^f, edu^m, 0)) \mathbb{E}\Big(V_{t+1}^{0C,F}(a_{t+1}, \mathsf{W}_{t+1}; \Omega) \Big)$$
$$+ Pr_t^b(edu^f, edu^m, 0) \mathbb{E}\Big(V_{t+1}^{1C,F}(a_{t+1}, \mathsf{W}_{t+1}, 0, \Theta_0^1; \Omega) \Big) \Big] \Big\}, \tag{1.14}$$

[17] More precise answer is all the other traits of the child that the measure of cognitive ability used in this paper (Letter-Word exam) is less influenced by them, but affect the developmental process of the child. The composition of these traits can change over time.

[18] In this paper, parental investments only affect the development of the cognitive ability of children. Since the effect of parental investments on non-cognitive abilities are less prone to fading-out over time (Cuhna et al. (2010)), considering non-cognitive abilities may increase the optimal level of transfers to families with children. Inclusion of non-cognitive ability is left to future work.

subject to (1.6), (1.5), and (1.8). The first expectation on the right-hand side is over persistent wage shocks, and the second expectation is over a potential child's innate cognitive ability and persistent and transitory shocks to it in addition to the persistent wage shocks. $V_{t+1}^{1C,F}(.)$ is the value function of families with 1 child which will be covered in Section 1.2.7. The value function of no-child families after the fertility window has an additional state variable n_t^{Le} (due to its effect on utility of consumption) and is presented in Appendix 1.B.1.

Problem of Families with One Child

Families with a single child also decide on the investment in their child in comparison to no-child families. Defining $\Theta_{z_t^k}^k$ to be the state variables related to cognitive ability (cognitive ability, persistent cognitive ability, innovation on persistent cognitive ability, and transitory shock on cognitive ability) of child k at age z_t^k, $\Theta_{z_t^k}^k = \{\theta_{z_t^k}^k, \xi_{z_t^k}^{\theta,k}, \iota_{z_t^k}^{\theta,k}, \epsilon_{z_t^k}^{\theta,k}\}$. In addition to innovations to permanent wages, families also face uncertainty about permanent and transitory shocks to the cognitive ability of their child. At $t < T_{Fer}$, the value function of one-child families during this time period is[19]:

$$\begin{aligned} V_t^{1C,F}(a_t, \mathsf{W}_t, z_t^1, \Theta_{z_t^1}^1; \Omega) = \max_{c_t, e_{z_t^1}^1, h_t^f, h_t^m} \Big\{ u_1(c_t, e_{z_t^1}^1, h_t^f, h_t^m, \theta_{z_t^1}^1, z_t^1; \Omega) \\ + \beta \Big[(1 - Pr_t^b(edu^f, edu^m, 1)) \mathbb{E} V_{t+1}^{1C,F}(a_{t+1}, \mathsf{W}_{t+1}, z_{t+1}^1, \Theta_{z_{t+1}^1}^1; \Omega) \\ + Pr_t^b(edu^f, edu^m, 1) \mathbb{E} V_{t+1}^{2C}(a_{t+1}, \mathsf{W}_{t+1}, z_{t+1}^1, 0, \Theta_{z_{t+1}^1}^1, \Theta_0^2; \Omega) \Big] \Big\}, \end{aligned} \tag{1.15}$$

subject to (1.5), (1.6), (1.8), (1.11), and $z_{t+1}^k = z_{t+1}^k + 1$. $V_{t+1}^{2C}(.)$ is the value function of families with 2 children at ages of z_{t+1}^1 and 0 and will be explained in Section 1.2.7. The first expectation on the right-hand side is over persistent wage shocks and the second expectation is over persistent wage shocks, the second child's innate cognitive ability, and persistent and transitory shocks to abilities of both children.

The separability of utility from cognitive ability of the child and other decisions allows us to write any of the value functions of one-child families into the sum of two parts. Here, I show how I

[19] k is either 1 or 2 depending on whether this is the family's first or second child currently present in the household. However, since $T_{Fer} < z_L$, when $t < T_{Fer}$, k can only be 1.

use this method to simplify the problem of families with a child after the fertility window. Note that since it is possible that this family was a two-children family before, there is an additional state variable n_t^{Le}.

$$V_t^{1C}(a_t, \mathsf{W}_t, z_t^k, n_t^{Le}, \Theta_{z_t^k}^k; \Omega) = \tilde{V}_t^{1C}(a_t, \mathsf{W}_t, z_t^k, n_t^{Le}; \Omega) + \hat{V}_t^{1C}(z_t^k, \Theta_{z_t^k}^k; \Omega), \tag{1.16}$$

where

$$\tilde{V}_t^{1C}(a_t, \mathsf{W}_t, z_t^k, n_t^{Le}; \Omega) = \max_{c_t, e^1_{z_t^1}, h_t^f, h_t^m} \Big\{ \frac{1}{1-\sigma} \Big[\big(1 - \mu^{LP}(z_t^k)\mathbb{I}(h_t^m > 0)\big) \frac{\big\{ c_t^{1-\delta_1(z_t^k)} (e_t^k)^{\delta_1(z_t^k)} \big\}}{\rho(1, n_t^{Le})} \Big]^{(1-\sigma)}$$
$$- \mu^h(h_t^f, h_t^m) + \beta \Big[\eta^k_{z^k_{t+1}} \Big(\alpha^z_{z_t^k} z_t^k + \alpha^e_{z_t^k} \ln e^k_{z_t^k} \Big) + \mathbb{E}\tilde{V}_{t+1}^{1C}(a_{t+1}, \mathsf{W}_{t+1}, z_{t+1}^k, n_t^{Le}; \Omega) \Big] \Big\}, \tag{1.17}$$

and

$$\hat{V}_t^1(\theta^k_{z_t^k}, \epsilon_t^k, z_t^k; \Omega) = \eta^k_{z_t^k} \ln \theta^k_{z_t^k} + \beta \eta^k_{z^k_{t+1}} (\xi^{\theta,k}_{z_t^k} + \epsilon^{\theta,k}_{z_t^k}). \tag{1.18}$$

$\eta^k_{z_t^k}$ is the altruism coefficient adjusted for dynamic complementary of abilities and has two parts. The first part shows the current utility from current cognitive ability of the child. The second part represents all the future utilities that will be gained from future cognitive abilities of the child as the current cognitive ability is also an input for generating future cognitive abilities. $\eta^k_{z_t^k}$ is defined recursively

$$\eta^k_{z_L+1} = \chi^\theta \mu^\theta$$
$$\eta^k_{z_t^k} = \mu^\theta + \beta \alpha^\theta_{z_t^k} \eta^k_{z^k_{t+1}}, \; \textit{if } z_t^k \leq z_L. \tag{1.19}$$

Using first-order conditions and applying the envelope theorem to solve the budget allocation problem, we can write the Euler equation as

$$\frac{\partial}{\partial c_t} U_1(c_t, e_t^k, \cdots) = \beta \mathbb{E}[\frac{\partial}{\partial a_{t+1}} V_{t+1}^{1C}(.)] = \frac{\partial}{\partial e_t} U_1(c_t, e_t^k, \cdots) + \beta \mathbb{E}[\frac{\partial}{\partial e^k_{z_t^k}} \theta^k_{z^k_{t+1}} \cdot \frac{\partial}{\partial \theta^k_{z^k_{t+1}}} V_{t+1}^{1C}(.)]. \tag{1.20}$$

The first equality is the familiar life-cycle Euler equation. The second equality arises in a child-development environment and equalizes the marginal utility cost of forgoing 1$ of consumption to the marginal benefit of investing it on the child. It will increase the utility of consumption by $\frac{\partial}{\partial e_t} U_1(c_t, e^k_t, \cdots)$ and the next period cognitive ability by its marginal productivity $\frac{\partial}{\partial e^k_{z^k_t}} \theta^k_{z^k_{t+1}}$, which has the marginal value of $\beta \frac{\partial}{\partial \theta^k_{z^k_{t+1}}} V^{1C}_{t+1}(.)$[20]. Using (1.16), (1.18), and (1.19), the last part of (1.20) can be simplified to

$$\beta \mathbb{E}[\frac{\partial}{\partial e^k_{z^k_t}} \theta^k_{z^k_{t+1}} \cdot \frac{\partial}{\partial \theta^k_{z^k_{t+1}}} V^{1C}_{t+1}(.)] = \beta \alpha^e_{z^k_t} \eta^k_{z^k_t} \frac{1}{e^k_{z^k_t}}. \tag{1.21}$$

The marginal benefit of additional investment in the child depends on the subjective discount rate, the coefficient of investment in the production function, the current age, the altruism coefficient and scale, and all the future coefficients of current investment in the production function. Notice that current cognitive ability and its related shocks do not affect the budget allocation problem which greatly simplifies the numerical solution. The intuition of this is that higher cognitive ability increases the marginal productivity of the investment, but also decreases the marginal utility of the gain from that investment. Due to the Cobb-Douglas production function and logarithmic utility function for cognitive ability, these two effects cancel each other out.

Problem of Families with Two Children

Families with two children additionally make a decision on investment in their second child. Unless the first (oldest) child is about to leave the household ($z^1_t = z_L$), the value function of two-children families is

$$\begin{aligned} V^{2C}_t(a_t, \mathsf{W}_t, z^1_t, z^2_t, \Theta^1_{z^1_t}, \Theta^2_{z^2_t}; \Omega) = \max_{c_t, e^1_{z^1_t}, e^2_{z^2_t}, h^f_t, h^m_t} \Big\{ u_1(c_t, e^1_{z^k_1}, e^2_{z^k_2}, h^f_t, h^m_t, \theta^1_{z^1_t}, \theta^2_{z^2_t}, z^1_t, z^2_t; \Omega) \\ + \beta \mathbb{E}\Big[V^{2C}_{t+1}(a_{t+1}, \mathsf{W}_{t+1}, z^1_{t+1}, z^2_{t+1}, \Theta^1_{z^1_{t+1}}, \Theta^2_{z^2_{t+1}}; \Omega) \Big] \Big\}, \end{aligned} \tag{1.22}$$

[20]This include next year's utility from it and future utilities from the increase in cognitive abilities because of the higher marginal productivity of investments, as can be seen in (1.19).

subject to (1.5), (1.6), (1.8), (1.11), and $z^k_{t+1} = z^k_{t+1} + 1$. The expectation on the right-hand is over persistent wage shocks and persistent and transitory shocks to the abilities of both children.

Note that we can use the same method as in the one-child family case to simplify the dynamic problem of two-children families.

$$\frac{\partial}{\partial c_t} U_2(c_t, .) = \beta R \mathbb{E}[\frac{\partial}{\partial a_{t+1}} V^{2C}_{t+1}] = \beta \alpha^e_{z^1_t} \eta^1_{z^1_t} \frac{1}{e^1_{z^1_t}} = \beta \alpha^e_{z^2_t} \eta^2_{z^2_t} \frac{1}{e^2_{z^2_t}}. \tag{1.23}$$

This equation also gives us the optimal relative investments between the two children $\frac{\alpha^e_{z^2_t} \eta^2_{z^2_t}}{\alpha^e_{z^1_t} \eta^1_{z^1_t}}$. Given the assumption that parents' altruism is identical across their children, the optimal relative investments only depends on their ages.

Problem of Families During Retirement

The value of households during retirement ($t \geq t^r$) is

$$V^r_t(a_t, n^{Le}_t; \Omega) = \max_{c_t} \left\{ u_r(c_t; n^{Le}_t) + \beta V^r_{t+1}(a_{t+1}, n^{Le}_t; \Omega) \right\} \tag{1.24}$$

subject to (1.7). Since there is no bequest motive, the terminal value at $t = t_d$ is zero.

1.2.8 Model Solution

I use backward induction to numerically solve the model. For each set of state variables and each time allocation, after checking the borrowing constraint, I find the optimal budget allocation by solving Euler equations like (1.20) using a nested fixed-point algorithm. Then, I solve for the optimal time allocation of each set of state variables given the previous results using grid search. For a detailed explanation of the algorithm, see Appendix 1.C.1.

1.3 Data

1.3.1 Data Sources

To estimate the empirical model, I use data from the Panel Study of Income Dynamics (PSID) and its Child Development Supplement (CDS). The PSID is the longest running longitudinal household survey in the world. It began as a nationally representative sample of 5,000 families in the United States.

The CDS is a component of the PSID. The CDS provide information on different forms of parental investment, neighborhood and home conditions, health and insurance coverage, time diaries, and various measures of the cognitive and non-cognitive skills of the children. The first CDS-I study in 1997 includes around 3500 children. The children in CDS-I were 0 to 12 years old. They were followed over two waves of CDS-II in 2002-3, and later in CDS-III in 2007. CDS IV in 2014 starts with a new sample of children. I utilize all of these four waves.

1.3.2 Creating the Sample

I create the sample of families from the PSID in the following way. I only consider families who will have at most two children throughout their life. I use the couples who are married after 1970. The start of each family is from when the father is at least 22 years old until he is 75 years old. For the couples who get divorced, I drop the observations after the divorce.

1.3.3 Summary statistics

Table 1.1 reports the descriptive summary of the final sample. The variables in the CDS columns only count observations in which I observe the Letter-Word exam score of the child. All monetary variables are in 2015 dollars and all time variables are in weekly units.

Table 1.1: Descriptive Statistics of PSID and PSID-CDS

	PSID			CDS		
	mean	Median	sd	mean	Median	sd
Age of Female	34.53	33	(9.92)	38.70	39	(6.76)
at Birth of the First Child	26.32	26	(4.88)	27.51	27	(4.79)
at Birth of the Second Child	29.13	29	(4.77)	30.26	30	(4.50)
Age of Male	36.20	34	(10.01)	40.16	40	(7.09)
at Birth of the First Child	28.00	28	(5.07)	28.92	28	(5.23)
at Birth of the Second Child	30.83	31	(4.98)	31.83	31	(5.08)
Number of Children	0.86	1	(0.85)	1.48	2	(0.74)
throughout Life	1.44	2	(0.75)	1.80	2	(0.40)
Years of Schooling of Female	14.12	14	(2.24)	14.44	14	(2.07)
Years of Schooling of Male	13.85	14	(2.35)	14.13	14	(2.24)
Hourly Wage of Female	20.56	17	(23.91)	23.00	19	(17.57)
Hourly Wage of Male	27.92	22	(39.86)	34.29	24	(49.04)
Weekly Work Hours of Female	26.16	30	(23.22)	27.18	33	(17.39)
Weekly Work Hours of Male	41.67	40	(20.72)	42.50	41	(13.16)
Income of Female	26204.12	21026	(29743.43)	32561.88	26350	(34449.97)
Income of Male	57311.15	46542	(66316.75)	75995.21	53227	(136585.75)
Income of Family	83515.27	70523	(76822.45)	108557.09	83499	(141394.54)
Assets	275170.76	72528	(1245588.45)	419299.97	171468	(908844.72)
Age of First Child	10.13	8	(8.75)	11.62	11	(5.13)
Age of Second Child	9.64	8	(8.22)	9.05	9	(5.04)
Letter-Word Score of First Child	36.81	42	(15.82)	36.81	42	(15.82)
Letter-Word Score of Second Child	39.00	44	(14.45)	39.00	44	(14.45)
Observations	54205			1185		

The CDS has a smaller sample size than the PSID. Moreover, the couples in the CDS are more educated and richer. The mean ages of females at the birth of her first and second child are 26.32 and 29.13 in the PSID and 27.51 and 30.26 in the CDS. The average total number of children of families throughout their life is 1.44 in PSID and 1.80 in CDS. I discuss the differences in ages, wages, and assets of families conditional on the number of children and their implications for the design of optimal tax in Section 1.5. For descriptive statistics conditional on the number of children, see Appendix 1.A.

1.4 Taking the model to the Data

1.4.1 Empirical Issues

Measurement Model of Cognitive Ability

I assume the cognitive ability of the child is observable to the parents, but not to the econometrician. To measure the distribution of latent cognitive ability of the children, I use the Letter-Word score. Following Boca et al. (2014), I assume that the probability of answering each question depends on a child's cognitive ability and follows a 2 parameters Rasch model, in which the difficulty and the discrimination parameters are normalized to 1: $p(\theta) = \frac{\theta}{1+\theta}$. Hence, for each student the number of correct answers out of 57 questions of the exam follows a Binomial distribution $S \sim Bin(57, \theta)$[21].

Equivalent Scales

I assume the following functional forms for equivalent scales

$$\begin{cases} \rho^0(n_t^{Le}) & = 1.5 + \mathbb{I}\{n_t^{Le} = 1\}\gamma_1 + 2\mathbb{I}\{n_t^{Le} = 2\}\gamma_2, \\ \rho^1(z_t^k, n_t^{Le}) & = 1.5 + \rho^*(z_t^k) + \mathbb{I}\{n_t^{Le} = 1\}\gamma_2, \\ \rho^2(z_t^1, z_t^2, n_t^{Le}) & = 1.5 + \rho^*(z_t^1) + \rho^*(z_t^2). \\ \rho^r(n_t^{Le}) & = \rho^{*,r} + \mathbb{I}\{n_t^{Le} = 1\}\gamma_1 + 2\mathbb{I}\{n_t^{Le} = 2\}\gamma_2. \end{cases}$$

1.5 is the equivalent scale of families without children[22] and $\rho^*(z_t^k)$ represents the extra costs

[21] I am assuming that the difficulty and discrimination coefficients of all the questions are the same. While there is no prior on how heterogeneous the discrimination coefficients of these questions are, the difficulty of questions are higher for further questions by design. This assumption is made purely to reduce the computational cost of estimating more parameters. While this assumption can overestimate the marginal benefit of inputs at low ages, the extent of the bias is unknown. Estimation of an item response model along the ability production function is left to future work.

[22] Based on the "OECD modified scales" `https://www.oecd.org/els/soc/OECD-Note-EquivalenceScales.pdf`.

associates with a child. $\rho^*(z_t^k)$ is equal to $(0.182, 0.287, 0.328, 0.361)$ given ages between 0-2, 3-5, 6-10, 11-18 from Banks et al. (1998). γ_1 and γ_2 represent the households preference for the consumption of children who left the households for families who have one and two children throughout their life, respectively. $\rho^{*,r}$ represents the equivalent scale of a retired couple.

Functional Form of Age-Dependent Parameters

In the model presented in Section 2.5, all the consumption equivalence scales and production function parameters are age-of-child dependent. Without a functional form assumption on how these parameters change by age of the child, estimation of these parameters requires a large number of observations and is computational burdensome. Thus, I approximate these functions using piece-wise linear functions that are constant within three stages of a child's life: before school, elementary school, and high school. I use linear interpolation for ages that are at the beginning and end of these parts.

The kink points in the coefficients of the cognitive ability production function are chosen in a way to allow for different parameters in different developmental and schooling stages of children. The fixed parts of the production function are on ages 0-4, 8-10, and 14-17. Note that unlike the functional form assumptions made in Boca et al. (2014), I am not restricting the production function coefficients to be monotone given ages.

The literature on equivalence scales finds large variation in these parameters given children's ages (Blundell and Lewbel (1991) and Banks et al. (1992)). In this model, this is especially important due to the child-development process. Parts of the household's expenditure that are productive at different developmental stages of the child are different, so do as the parents' enjoyment from consumption of those expenditure. I allow the rivalry of goods bought as unobservable investment in children to vary by the age of children in the same way marginal benefit of this investment changes by the age of children by choosing the same functional form for the share of investment in children in the utility of consumption and the marginal benefit of investment in children[23].

[23]In models with unobservable investment in children and an age-of-child dependent production function, incorporating the age-of-child dependency of marginal utility of consumption is essential to match the saving decisions of households. Alternately, if the marginal utility of consumption were not age-of-child dependent, in the ages of children

Note that given the flexible functional form for the age-dependant parameters of the cognitive ability production function, the marginal benefit of these investments (shown in (1.21)) can have a complicated shape dependent on the age of the child. I approximate this function using the same functional form as the production function coefficients, but I choose the location of the kink points so that the approximation error (evaluated by sum of the residuals) is the lowest. These kink points will determine the functional form of the share of investment in children in the utility of the consumption of parents. I treat this approximation as part of the model, as parents approximate the marginal benefits of each part of their expectation given the age of their child.

While I estimate the share of investment in the utility of consumption for one-child families, I assume the following functional form for two-children families

$$\delta^2(z_t^1 + z_t^2) = \frac{1}{2}[\delta^1(z_t^1) + \delta^1(z_t^2)] + \frac{1}{2}\max[\delta^1(z_t^1), \delta^1(z_t^2)]. \tag{1.25}$$

1.4.2 Estimation

A few parameters are chosen based on the previous literature. The yearly interest rate is opted as $R = 1.03$ (same as Blundell et al. (2018)), risk aversion is set to $\sigma = 3$ (between the estimates for non-college and college educated in Cagetti (2003) and in the middle of common parameterization in the literature for non-separable utility functions (Conesa et al. (2009) and Low and Pistaferri (2015)), and the subjective discount factor is chosen to the $\beta = 0.9597$ (mean of estimates in Cagetti (2003) based on median of assets in PSID). For the rest of the parameters, I use a two-step estimation method. First, I estimate the parameters that can be more clearly identified from the data without use of the complete model. I estimate the tax schedule, fertility process, wage processes, distribution of education level of parents, and initial assets directly from the data[24]. Second, I estimate the

where marginal benefit of investment is higher, the model would have predicted that both the investment and the total expenditure of families are higher. This is not the case for the total expenditure of the types of goods that improve child development are different in different ages. Since we follow both small and teenage children, this is very likely the case.

[24]For most of these parameters, this is because they can be consistently estimated without the use of the structural model given the exogeneity assumptions. The only exemption is wage processes, as their estimation without the use of the structural model suffers from the selection bias due to the large fraction of females who choose not to work. However, when I estimate the wage process of women utilizing the structural model and SMM, I do not find a significant change in the results. Hence, I use the reduced form approach to decrease the computational burden of estimation of more parameters.

rest of the parameters (preference parameters and children's cognitive ability's distribution and production function) using a simulated method of moments (SMM).

1.4.3 Directly Estimated Parameters

Tax Function

The tax brackets and marginal taxes have changed significantly between 1980 and 2015. Hence, the individuals in my data have been subject to 35 years of different tax rules. Since I am not separating cohorts, I cannot take the tax functions directly from tax rules. So instead, I approximate the tax functions using linear splines with fixed knots. For the details of approximation of the tax function, see Appendix 1.D.2.

Other Government Transfers

To calculate the social security payment, I use the rules as they were in the year 2000. For the food stamp program, I assume it functions as an income-floor program for low income households and is no different than cash. I employ the formula presented in Wilde (2001) and approximate it based on rules and deductions of food stamp programs across states using data provided in Low and Pistaferri (2015).

Wage Processes

I treat the difference between the the wage in the wage process of the model and the observed wage as measurement error uncorrelated with permanent wage shocks (but correlated across spouses). I estimate the distribution of wage shocks via GMM using variance, auto-correlation and correlation across spouses of the unexplained error as moment conditions. The details of the identification can be found in Section 1.D.1. The parameters of the deterministic part of the wage functions are reported in Table 1.2.

Table 1.2: Parameters of The Deterministic Part of Wage Equations

	log(*Wage Female*)	log(*Wage Male*)
Age	0.0516^{***}	0.0619^{***}
	(12.38)	(13.96)
Age^2	-0.000554^{***}	-0.000672^{***}
	(-9.94)	(-11.38)
College Education	0.134^{**}	-0.0732
	(2.86)	(-1.41)
Age $\times$ College Education	0.00781^{***}	0.0123^{***}
	(5.27)	(7.85)
Constant	1.649^{***}	1.716^{***}
	(22.52)	(21.30)
Observations	24176	24176
Adjusted R^2	0.192	0.185

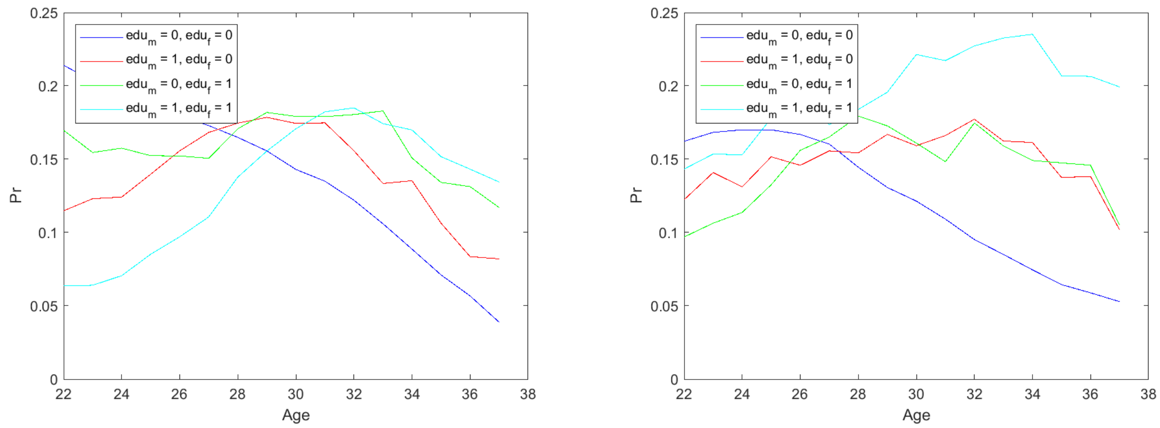
t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The estimated parameters of the distribution of permanent wage shocks are $\sigma^2_{\iota_{w,f}} = 0.02697$, $\rho_{\iota_{w,f,m}} = 0.00358$, and $\sigma^2_{\iota_{w,m}} = 0.01638$.

Fertility Process

I estimate the fertility process using a bin estimator. Figure 1.1 shows the probability of having the first and the second child conditional on education group of spouses and average their ages. Families with a college-educated wife have their first and second child later in life compared to families where the wife is not college-educated.



(A) First Child (B) Second Child

Figure 1.1: Probability of Fertility

Initial Distributions

For estimation of the initial distribution of assets, I use the total family wealth (including home equity) data from the PSID[25]. I approximate the initial distribution of assets using the log-normal distribution $f(log(a_0^i) \mid a_0^i > 0) = \mathbb{N}(\overline{a_0}, \sigma_{a_0}^2) = \mathbb{N}(9.36, 1.76)$, assuming that the probability of having positive initial assets (versus zero) follows a Bernoulli distribution with the probability $pr(a_0 > 0) = 0.81$.

I assume an individual is college educated if he or she has more then 16 years of schooling in the data. Table 1.3 shows the joint distribution of education of parents.

Table 1.3: Distribution of Education of Spouses

	$edu^m = 0$	$edu^m = 1$
$edu^f = 0$	57.04 %	7.36 %
$edu^f = 1$	13.97 %	21.61 %

To estimate the initial distribution of permanent wages, I use the unexplained error of the wage equations in the initial period: $\sigma^2_{\xi_0^{w,m}} = 0.118$, $\sigma^2_{\xi_0^{w,f}} = 0.048$, and $\rho_{\xi_0^{w,m,f}} = .129$.

[25]It is the sum of seven asset types (farms and businesses, checking/savings accounts, real estate, stocks, vehicles, other assets, annuity/IRA, other debt, and home equity (value of home minus mortgage)).

1.4.4 Parameters Estimated Using Method of Simulated Moments

To estimate the rest of the parameters, I use the simulated method of moments. I simulate the life-cycle behavior of 10,000 families. Then, I build a minimum distance estimator to estimate the remaining 27 parameters of the model.

$$\Gamma^* = argmin_{\Gamma}(M_D - M_S(\Gamma))W(M_D - M_S(\Gamma)), \tag{1.26}$$

where M_D and $M_S(\Gamma)$ are the data and simulated moments, respectively, which include 59 moments provided in Appendix 1.D.3. W is the weighing matrix which uses the inverse of the variance of empirical moments calculated from the bootstrap method.

1.4.5 Identification

In this section, I explain the identification of parameters estimated via SMM, focusing more on the parameters related to the child-development process.

Parameters Related to Consumption, Saving, and Labor Supply

The parameters estimated via SMM are estimated together, however, some moments are more informative about some parameters. For instance, the median of asset holdings before the retirement age disciplines the equivalence scale during the retirement of spouses, differences in the level of asset holdings of families with different number of children at different ages are reflected in age-and-number-of-children-dependent shares of investment in children in the utility of consumption, and the level of asset of households after the departure of their children are informative about the effect of children outside the household on equivalent scales. Different aggregate moments of the distribution of labor supply of spouses pins down disutility of part-time and full-time works, while hours of work of mothers of small children provide information on the non-separable disutility cost of working mothers.

Parameters Related to the Child-Development Process

Letter-Word exam scores are stochastically determined by the cognitive ability of children that are in turn shaped by parent's investments along other inputs. Hence, the distribution of scores across ages and how scores vary in cross-sectional variation in income are informative about altruism coefficients. The identification of altruism in preferences hinge on the assumption that the production function parameters do not depend on any unobservable variable, but they can be conditional on the observable variables (here it only depends on a child's age). The mean of exam scores in different ages is reflected at the mean of the distribution of the altruism coefficient[26], the mean of scores in old ages compare to younger age of the children helps to identify the altruism scale coefficient of final cognitive ability, and variation in scores of children in response to changes in family income is informative about standard deviation of altruism coefficient.

For identification of the production function, I use data from children that I observe more than once in my sample. Given that in the first 3 waves of the CDS, the exams were administrated to the same children if their age was lower than 18, I can observe children at most 3 times with a 5 year gap between observations. The cross-moments of scores from these children and past ages, scores, and family incomes are informative about parameters of production function. To best use the data, I utilize an indirect inference method. I replace the unobserved child-specific expenditure in (1.11) with the observed family income and run the following regression using both the model-generated data and the actual data:

$$\ln(1+S_{t+5}^{i,k}) = \beta_0+\beta_1 z_t^{i,k}+\beta_2\ln(1+S_t^{i,k})+\beta_3 z_t^{i,k}\ln(1+S_t^{i,k})+\beta_4\ln(1+y_t^i)+\beta_5 z_t^{i,k}\ln(1+y_t^i)+\beta_6 n_t^i+\epsilon_t^{i,k} \tag{1.27}$$

Although there is no closed-form solution for the share of total expenditure allocated to unobserved investment in children, this regression helps to identify the production function parameters conditioning on other inputs. I use regressions on a sub-sample of these variables and correlation

[26]Note that the altruism coefficient and equivalent scales together govern moments of children's scores and assets. However, the moments of children's scores are determined by the altruism coefficient relative to equivalent scale (from (1.21)) and the moments of assets of families with children are shaped by the level of these parameters (which determine the total expenditure) .

between these variables as additional moments to identify production function parameters. Mean of children's scores at different ages is also most informative of the age-of-child-dependent TFP coefficients.

The distribution of scores at age 3 (the first time we observe children's scores) helps to identify the distribution of innate cognitive ability of children (given the extrapolation of age-of-child-dependent parameters of production function to ages below 3)[27]. The cross-sectional variation and the time-variation of the unexplained part of the regression in (1.27) are reflected in the standard deviation of transitory and permanent cognitive ability shocks. I use the standard deviation and autocorrolation of these unexplained shocks as informative moments for these parameters.

1.4.6 Estimated Parameters

The estimated parameters are reported in 1.4. For men, the disutility from working full time is twice the disutility of part-time work. For women, the fixed cost disutility of labor force participation is large and having a less than 6-year-old child further increases this cost. Families who have two-children in life value transfers to the children after they leave the household more than families with one child in life. The variation in the altruism coefficient among families is small relative to its mean. Families value the final cognitive ability of their children three times more than their previous abilities. I discuss how parameters of the child development production function and the share of child investment in utility of parents vary by age of the child in Section 1.5.2.

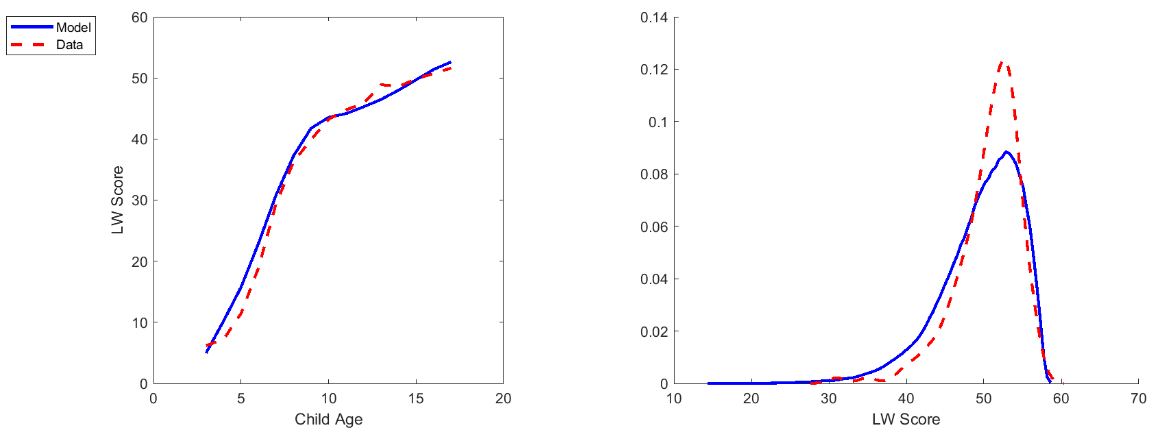
[27] Although the letter-word exam was also administered on children lower than 3 years old, the number of observations for children below 3 is not enough for estimation with reasonable precision.

Table 1.4: Estimated Parameters via SMM

$\mu^{h_m}(h_P)$ $(\times 10^8)$	Disutility of Part Time Work of Male	.11433
$\mu^{h_m}(h_F)$ $(\times 10^8)$	Disutility of Full Time Work of Male	.26241
$\mu^{h_f}(h_P)$	Disutility of Part Time Work of Male	.90622
$\mu^{h_f}(h_F)$	Disutility of Full Time Work of Female	.64351
$\mu^{LP,K}$	Added Disutility of Work of Female with a Small Child	.03775
$\delta_1(0 \leq z \leq 6)$	Shares of Investment in Children in Utility of Consumption	.07998
$\delta_1(6 < z \leq 12)$ $(\times 10^6)$		.0042
$\delta_1(12 \leq z \leq 17)$ $(\times 10^6)$		7.0498
γ^1 $(\times 10^6)$	preference for Transfer to the left child in 1 child family	.26942
γ^2	Preference for Transfer to the left child in 2 child family	.33194
μ^r	Equivalent Scale of a Retired Couple	.69562
$\bar{\mu^\theta}$ $(\times 10^8)$	Mean of Altruism Cognitive Ability	.37505
σ_{μ^θ} $(\times 10^8)$	Standard Deviation of Altruism Cognitive Ability	.00214
χ^θ	Altruism Scale of the Final Cognitive Ability	3.02507
$\alpha^z(0 \leq z \leq 6)$	Coefficients of Total Factor Productivity	.00003
$\alpha^z(6 < z \leq 12)$		.00981
$\alpha^z(12 \leq z \leq 17)$		.02158
$\alpha^\theta(0 \leq z \leq 6)$	Coefficients of Current Cognitive Ability	.63217
$\alpha^\theta(6 < z \leq 12)$		.32327
$\alpha^\theta(12 \leq z \leq 17)$		1.10268
$\alpha^e(0 \leq z \leq 6)$	Coefficients of Child Investment	.0034
$\alpha^e(6 < z \leq 12)$		.08842
$\alpha^e(12 \leq z \leq 17)$		.02972
σ_{ι_θ}	Standard Deviation of Shocks to Permanent Cognitive Ability	.04532
σ_{ϵ_θ}	Standard Deviation of Transitory Cognitive Ability Shocks	.00312
$\bar{\theta_0}$	Mean of Innate Cognitive Ability	-10.30102
σ_{θ_0}	Standard Deviation of Innate Cognitive Ability	1.59501

1.4.7 Model Fit

The estimated model performs well in reproducing the important features of the data relevant to the design of optimal child-dependent taxes. The targeted moments, including the indirect inference regressions of the child development process, mimic the data well. For the fit of targeted moments, see Appendix 1.D.3. In this section, I show unconditional and conditional moments over the life cycle of couples and children. The moments presented in this section are not estimation targets, however the model replicates most of them accurately. Figure 1.2 shows the scores of children over their life cycle and the histogram of scores of children ages 15-17.

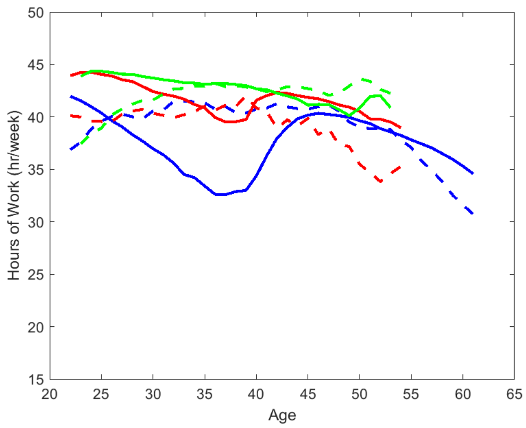


(A) Mean of Scores over time Time (B) Scores in Ages 15-17

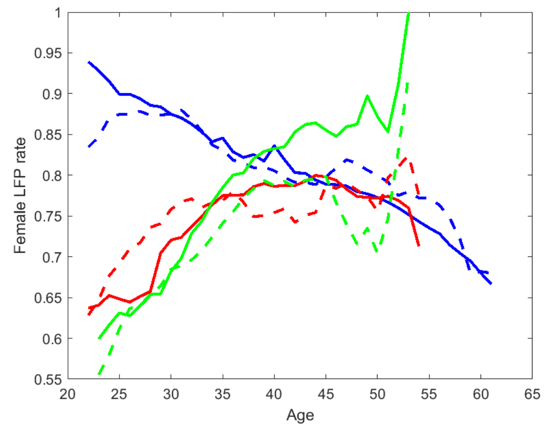
Figure 1.2: Model Fit of Aggregate Moments of Children

Figure 1.3 shows the model fit of mean work hours for working men, labor supply participation rate of women, and median family assets conditional on the number of children present in the household. To smooth the data, a 3-year moving average method is used.

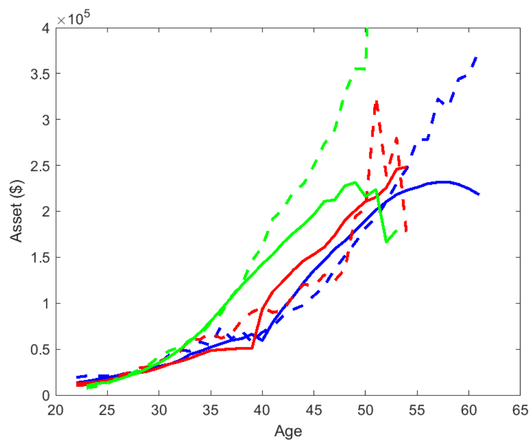
The model is able to match the decisions of families of different sizes over their life cycle well. However, there are two main caveats. First, the model underestimates the hours of work of men from no-child families in their 30s. When the female's wage is much higher than her spouse's wage, the income effect in the model becomes large enough to force men out of labor force participation. Assuming non-separability between disutility of work of spouses solves this issue. Second, while the model fits asset holdings of no-child and one-child families well, it underestimates asset holdings of families with 2 children whose parents are older than 40 years old. The literature finds higher motivation for saving among high educated individuals by estimating higher risk aversion and discount factor for them (Cagetti (2003)). In this current version of the model, there is no preference heterogeneity on the parameters that mainly derive the saving decision. Since parents with two children above age 40 are mostly high educated couples and this share increases more and more with age, the model underestimates the saving of this group.



(A) Mean of Hours of Work of Male



(B) Female Labor Force Participation Rate



(C) Median of Assets

Figure 1.3: Model Fit of Aggregate Moments Conditional on the Number of Children over the Life Cycle

1.5 Implications of the Data and the Model for the Design of Optimal Tax

In this section, I discuss how the data and estimated model can guide the design of optimal child-dependent taxes.

1.5.1 Tagging Based on Differences in Labor Supply Elasticises and Marginal Utilities of Expenditure

Table 1.5 reports the Marshalian labor force participation elasticities. Own elasticities are the percentage change in labor force participation of a spouse given a 1 % anticipated change in his or her wages letting all the decisions of the household adjust. The cross elasticities are the responses to the increase in one's spouse's wages in different scenarios[28].

The column "All" in Table 1.5 shows the elasticities with respect to the increase in wages in all years, regardless of if a couple has a child or not. The other columns show the elasticities with respect to the increase in wages only at times when the family has zero, one, and two children, respectively. All elasticities are in the range of Marshalian labor supply elasticities of married males and females in the literature[29].

Table 1.5: Marshalian Labor Force Participation Elasticities

		All	$n_t = 0$	$n_t = 1$	$n_t = 2$
Female	Own	0.56	0.53	0.87	0.63
	Cross	-0.44	-0.25	-0.39	-0.27
Male	Own	0.05	0.15	0.06	0.07
	Cross	-0.15	-0.17	-0.07	-0.08

Note: See text for detailed description.

[28]Given the 1 % change in wages, the Marshalian elasticities are defined as $me = [mean(\mathbb{I}\{\widehat{h_t^{i,j}} > 0\}) - mean(\mathbb{I}\{h_t^{i,j} > 0\})]/mean(\mathbb{I}\{h_t^{i,j} > 0\})$, where the hat on a variable represents its value after the change in wages.

[29]See Keane (2011) for the survey treatment of elasticities of labor supply.

As shown in Table 1.5, the male labor force participation response to changes in wages is small. Female elasticity of labor force participation is highest for mothers of one child, and the second highest elasticity is for mothers of two children. This suggests that labor supply of women who currently have children are more responsive to an increase in marginal taxes than women without children. The cross-elasticities are negative as expected (due to income-effects), and largest for women with 1 child.

To understands dynamic tagging and how responses of workers change over their life cycle, Table 1.6 shows the effects of a 1 % increase in wages at different parts of the life cycle of a family. The labor force participation of females and males are more responsive when these families have a child present in their home. In response to a 1% increase in wages of the wife or the husband in all years, the families with one child in life increase their total expenditure more when they have children. The total expenditure of two-children families increases much more when both of their children are living with them than when they do not have any children at home. This suggests higher marginal value of income in those years.

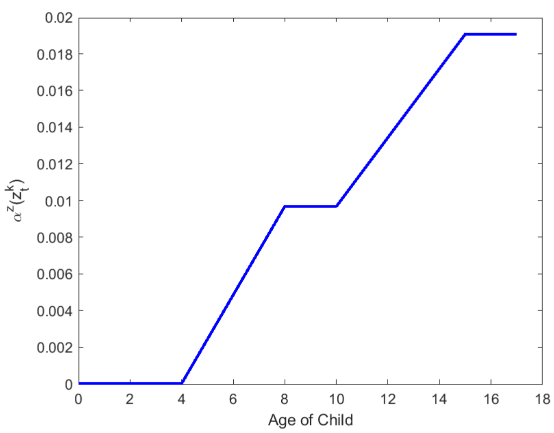
Table 1.6: Effects of 1 % Increase in Wages

1 % increase in wages of female	$n^{Life}=0$	$n^{Life}=1$			$n^{Life}=2$			
	All	All	$n^k=0$	$n^k=1$	All	$n^k=0$	$n^k=1$	$n^k=2$
$E+C$	0.35	0.36	0.31	0.39	0.44	0.32	0.46	0.48
LP^f	0.45	0.52	0.47	0.57	0.58	0.50	0.69	0.59
LP^m	-0.35	-0.22	-0.27	-0.12	-0.17	-0.14	-0.11	-0.10

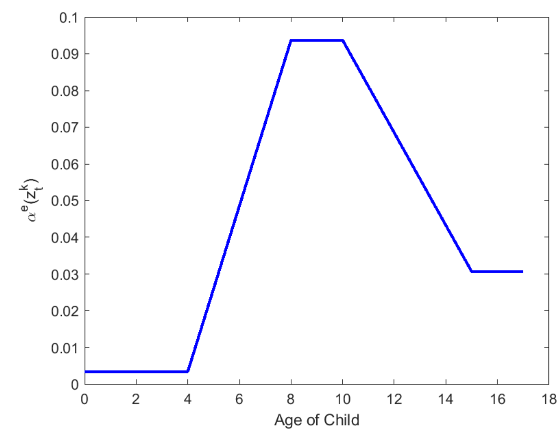
1 % increase in wages of male	$n^{Life}=0$	$n^{Life}=1$			$n^{Life}=2$			
	All	All	$n^k=0$	$n^k=1$	All	$n^k=0$	$n^k=1$	$n^k=2$
$E+C$	0.32	0.38	0.34	0.45	0.44	0.38	0.45	0.55
LP^f	-0.26	-0.41	-0.38	-0.45	-0.41	-0.36	-0.54	-0.41
LP^m	0.16	0.07	0.09	0.05	0.05	0.05	0.04	0.04

1.5.2 Tagging for the Purpose of Child Development

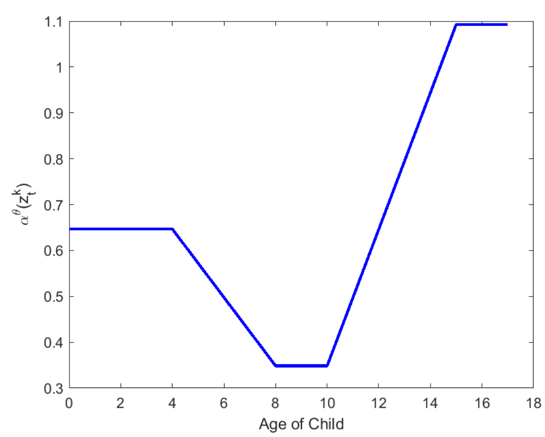
The first three figures in Figure 1.4 shows coefficients of the cognitive ability production function for TFP, investment, and past ability over the age of the child. The investment is most valuable when the child is in middle school and least valuable when the child is younger than school-age. The effect of past skill is highest at the latest stage of child development (when the increase in ability is mostly derived from past ability) and lowest in middle school.



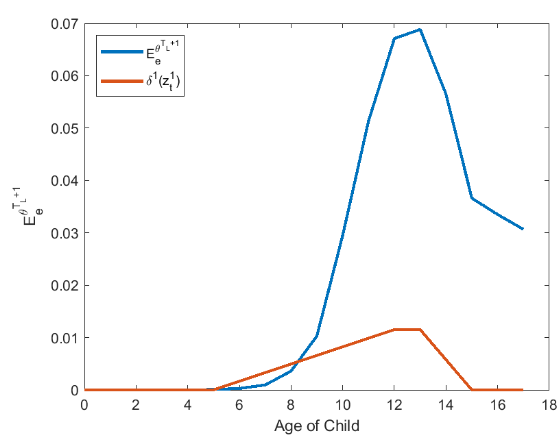
(A) $\alpha^z(z_t^k)$



(B) $\alpha^e(z_t^k)$



(C) $\alpha^\theta(z_t^k)$



(D) $Elasticity_e^{\theta^{z_L+1}}$ and $\delta^1(z_t^1)$

Figure 1.4: Production Function Coefficients and Marginal Benefit of 1 dollar of investment in Child

The last figure in Figure 1.4 shows the elasticity of the final ability of the child with respect to investment at different ages. These elasticities are calculated using $\alpha^e(z^k)\prod_{s=z^k+1}^{Z_L}\alpha^\theta(s)$. A 1 %

additional investment increases a child's next year ability by $\alpha^e(z^k)$, and this increased ability will increase future abilities until the final year by $\alpha^\theta(s)$ each year due to the dynamic complementary of abilities. As can be seen from the last figure in Figure 1.4 this elasticity starts to rise at the start of elementary school and is at its peak between middle school and high school, when the marginal productivity of investment is still high and the marginal productivity of ability is rising. Note that the public good nature of child investment also changes with the age of the child. At the year that the elasticity of final ability with respect to the child investment peaks, more goods from a household's total expenditure count as child investment.

The effect of government transfers to families on final scores of children does not necessary follows the same pattern as the last figure in Figure 1.4. Families use the transfers to consume, save, buy leisure, or invest in cognitive ability of their child. If families are not facing borrowing constraints, the year of transfer does not matter. Parents decide how much to invest on their children and optimally divide that throughout the life cycle of their child, given the production function of cognitive ability. With a borrowing constraint, the effect of transfers on final scores of children depend on the year of transfer.

The first 3 parts of Table 1.7 compares the effects of \$ 5000 unconditional transfer for each child when child is 0-5 years old, 6-11 years old, and 12-17 years old. I show the aggregate mean of consumption (C) , child investment (E), total expenditure $(C+E)$, labor force participation of women LP_F, hours of work of men H_m, family income Y, and final score of children (S_L), and aggregate standard deviation in total expenditure and final score of children for all families and conditional on number of children. The standard deviation of final test scores decreases more when the payments are to the families with 6-11 years old, compared to a policy that transfers to families with 0-5 years old or 12-17 years old children. However, the effect on mean of final scores are very similar in all three cases and around 0.1 % (1.5 % of the standard deviation final test scores). The last part of Table 1.7 shows the effect of transfer of the same amount of money at age 54, when all children have left their household. The effects on investment and mean and standard deviation of final test score are all smaller than previous cases, pointing out to borrowing constraint.

Table 1.7: Effects of Transfers to Families Given The Age a Child

	Mean							Sd		
n_t	C	E	$C+E$	LP_f	H_m	Y	S_L	$C+E$	S_L	
\$ 5000 Yearly Transfer to Families for each 0-5 Years Old Child										
0	.49	-	.49	-.58	-.21	-.4	0	.02	0	0
1	1.58	1.69	1.59	-2.98	-.32	-1.04	.02	1.13	.18	0
2	1.75	3.04	1.97	-2.93	-.34	-1.07	.1	1.51	-1.53	0
All	.93	2.63	1.01	-1.7	-.27	-.73	.09	.96	-1.31	8.64
\$ 5000 Yearly Transfer to Families for each 6-11 Years Old Child										
0	.48	-	.48	-.6	-.22	-.42	0	.05	0	0
1	1.59	1.9	1.61	-1.19	-.25	-.65	.04	.51	-.26	0
2	2.21	3.14	2.37	-2.07	-.33	-.97	.11	.84	-1.79	0
All	1	2.76	1.09	-1.08	-.26	-.62	.1	.77	-1.59	2.69
\$ 5000 Yearly Transfer to Families for each 12-17 Years Old Child										
0	.47	-	.47	-.56	-.2	-.42	0	.06	0	0
1	-1.01	1.69	-.8	-.89	-.24	-.57	.05	-1.66	-.45	0
2	-.55	2.59	-.01	-1.19	-.26	-.69	.1	-1.63	-1.62	0
All	0	2.32	.12	-.78	-.23	-.53	.1	-.85	-1.47	.95
\$ 30000 Transfer to Families who had 1 child at Age 54										
0	-.02	-	-.02	-.1	-.25	-.23	0	-.04	0	
1	.44	1.39	.52	-.06	-.18	-.22	.02	.39	-.19	
2	.71	2.16	.95	-.08	-.27	-.33	.08	.62	-.62	
All	.19	1.91	.28	-.09	-.24	-.25	.07	.37	-.57	

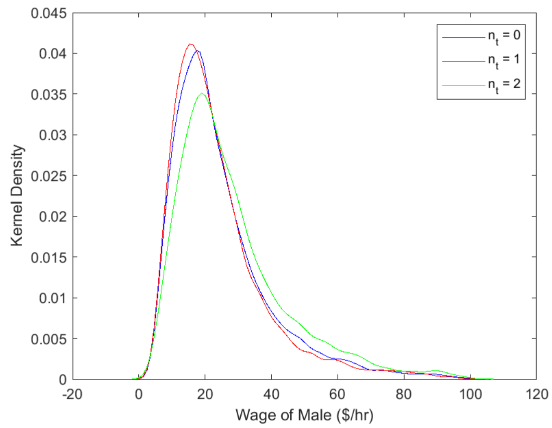
Notes: Capital letters show the aggregate moments: consumption (C), child investment (E), total expenditure $(C+E)$, labor force participation of women LP_F, hours of work of men H_m, family income Y, and final score of children (S_L) Outcomes are reported in % changes compared to the status quo.

1.5.3 Tagging Based on Differences in the Distributions of Wages and Assets

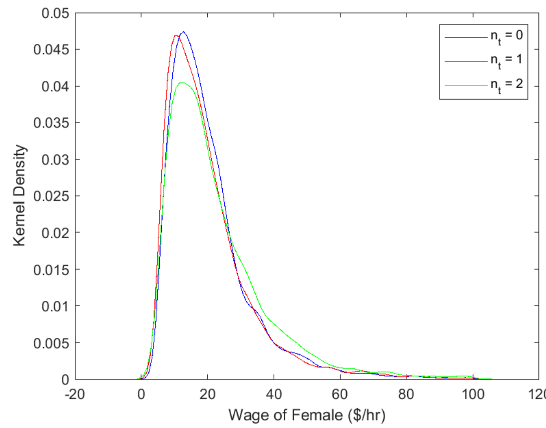
Akerlof (1978) and Cremer et al. (2010) argue that the difference in the distributions of productivity across groups gives motivation for tagging as the social planner can redistribute more to low-productivity workers in a group with less high-productivity workers to mimic their behavior. In this section, I argue that this can not be a motivation for child-dependent tagging.

In Figure 1.5, I show the distribution of wages and assets across couples with different numbers of children[30]. The two-children families have higher wages and assets than all other family types and no-child families have slightly higher wages and assets than families with one child. To see the reason behind the difference in these distributions, note that higher educated couples tend to have their first and especially their second child later than lower educated couples (Figure 1.1). As a result, families with children consist of both young lower educated couples (with low wage rates and asset levels) and middle-aged higher educated couples (with higher wages and higher assets). This leads to large variations in wages and assets of families with children.

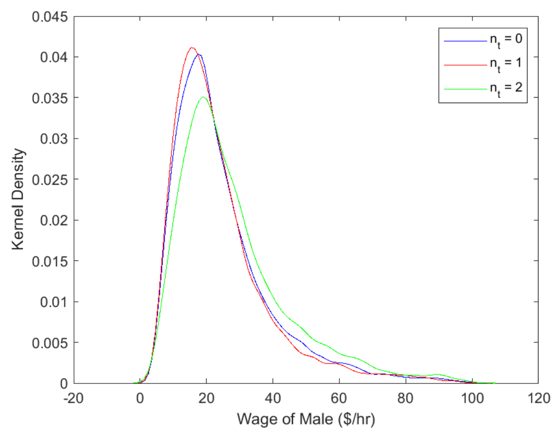
[30]Since the data on assets of families are unbalanced and scarce for couples, I show the distribution of assets using the simulation data.



(A) Wage of male in data



(B) Wage of female in data



(C) Asset in Model

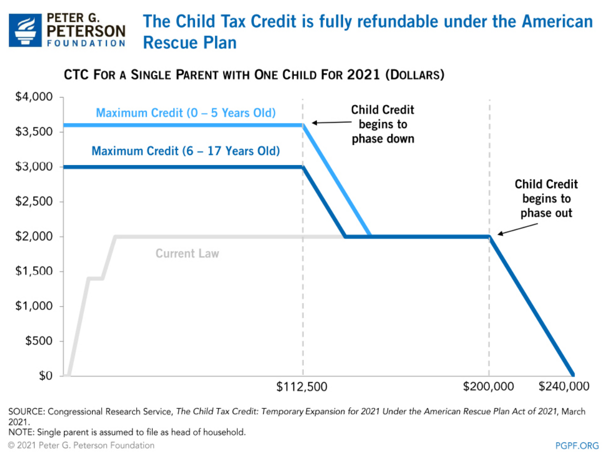
Figure 1.5: Histogram of Wages of Spouses and Asset of Family Across Family Types

Although child-dependent taxes are successful in conditioning taxes on the marginal utility of expenditure and labor supply elasticises, they are not effective in conditioning taxes on the mean and inequality in assets and wages since the young and old are pooled together. This suggests that introducing age-dependency in addition to child-dependency further improves the social welfare.

1.6 Counterfactual Tax Policy: Expansion of Child Tax Credit

In this section, I use my estimated model to evaluate the effects of permanently keeping the expanded child tax credit established as part of the American Rescue Plan (ARP). In 2021, the ARP only temporarily expanded the child tax credit; currently, there is debate over whether to make these tax credits permanent.[31]. This legislation eliminated the phase-in part and the cap on the refundable portion. It also increased the credit amount from $2,000 to $3,600 per child age 5 and under and $3,000 per older child between ages 6 and 17. Under ARP, the CTC begins to phase down when the income of married couples exceeds $150,000 ($112,500 for single taxpayers filing as head of household) by $ 50 for each additional $ 1000 of income until the credit equals $2,000 per child (the old amount of CTC). For the second time, the CTC starts to phase down for incomes above $ 400,000 ($200,000 for single taxpayers filing as head of household).

Figure 1.6: Comparing of Child tax Credit under the American Rescue Plan and the Existing Law



Notes: For married couples, the CTC will begin to phase down at incomes above $ 150,000 and $4 00,000. Figure from The Peter G. Peterson Foundation.

In Table 1.8, I show aggregate moments for all families and conditional on number of children. In part 2 of Table 1.8, I report the percentage change in these variables in an economy with the 2022 tax plan and the ARP child tax credit instead of the old one. The consumption equivalent variation for all families, and for each type of family, is presented in the last column.

[31]In October 2021, 448 economists wrote an open letter to Congressional leaders supporting this policy `https://www.cnbc.com/2021/09/16/over-400-economists-letter-favor-extending-300-dollar-child-tax-credits.html`.

Table 1.8: Effect of Permanently keeping American Rescue Plan Child Tax Credits

	Mean							Sd		
n_t	C	E	$C+E$	LP_f	H_m	Y	S_L	$C+E$	S_L	CEV
2022 Tax Plan with Old Child Tax Credit										
0	77394	-	77394	0.75	40.88	87823	-	32745	-	
1	87167	13326	92993	0.71	42.63	89770	53.82	34035	5.25	
2	78179	12262	93973	0.74	43.54	95968	53.45	38416	5.74	
All	79287	7195	83104	0.74	39.88	90225	53.51	34898	5.67	
2022 Tax Plan with American Rescue Plan Child Tax Credit (% change)										
0	1.08	-	1.08	-1.22	-0.38	-0.50	-	0.12	-	0.21
1	1.75	2.84	1.81	-3.52	-0.87	-1.56	0.05	-1.42	-0.25	2.28
2	2.34	4.40	2.69	-4.23	-0.93	-1.86	0.16	-0.92	-1.87	4.07
All	1.43	3.98	1.55	-2.47	-0.65	-1.10	0.14	0.04	-1.67	4.17
2022 Tax Plan with Unconditional Transfer of Cost of American Rescue Plan (% change)										
0	0.90	-	0.90	-1.30	-0.62	-1.27	-	0.08	-	3.25
1	1.20	3.08	1.34	-2.47	-0.55	-1.25	0.11	-0.37	-1.58	3.22
2	1.30	3.39	1.66	-1.97	-0.39	-1.91	0.10	0.15	-1.28	3.00
All	1.03	3.30	1.15	-1.73	-0.54	-1.18	0.10	0.25	-1.32	3.63

Notes: Capital letters show the aggregate moments: consumption (C), child investment (E), total expenditure $(C+E)$, labor force participation of women LP_F, hours of work of men H_m, family income Y, and final score of children (S_L). Outcomes are reported in % changes compared to the status quo. Unconditional CEV is defined in (1.28). Conditional CEVs are calculated by taking the average of CEVs of the families in that group. See text for detailed description

Mean of consumption of all types of families will go up under the 2022 tax plan with the ARP child tax credit. For one-child and two-children families, the mean of consumption will be increased by 1.75 % and 2.34 %, respectively. Even though no-child families are not receiving any tax credit, their consumption will also be increased by 1.08 percentage due to less need for saving for when they have children as government provides more insurance in those times. The standard deviation of consumption will go down among one-child and two-children families, but will increase among no-child families, therefore overall it will rise slightly. The around 4 percent increase in the mean of investment in children will increase the average score of children and lower inequality. The mean of the final score of children will go up by 0.14 % and the standard deviation will go down by 1.67 %. This change in tax policy has a modest effect on labor supply. It will decrease labor force participation of females by 2.47 % and hours worked by men by 0.65 %. The negative effect on labor supply is highest among two-children families as the female labor force participation will be lowered by 4.23 %.

To compare the welfare gain of a change in a policy, the last column in Table 1.8 reports consumption equivalence variations (CEV). CEV is defined as the per-period extra consumption that a couple is willing to accept in the veil of ignorance to forgo the new enlivenment. Defining the allocations of the status quo by zeros and the new economies by hats on households' variables:

$$\begin{aligned} W^0((1+CEV)c^0, h^{m,0}, h^{f,0}, e^0) &= (1+CEV)^{1-\sigma} W^{NS,0}(c^0, h^{f,0}, e^0) + W^{S,0}(h^{m,0}, e^0) \\ &= \widehat{W}(\widehat{c}, \widehat{h^m}, \widehat{h^f}, \widehat{e}), \end{aligned} \tag{1.28}$$

where W represents the expected life-time utility before realization of any shock. W^{NS} and W^{NS} show the non-separable and separable with consumption parts of the expected welfare. I also define the CEV conditional on the total number of children in life as the average of CEVs of the families in that group[32]. The CEVs reported in this paper are larger than the common values in the optimal dynamic tax literature, but in line with the value of social insurance programs in Low et al. (2010) which also features a discrete choice labor supply model and non-separable disutility of leisure.

This change in taxes will increase welfare by 4.17 percentage in terms of consumption equivalence. The increase in welfare is 4.07 % for families with 2 children in life and 2.28 % for families with 1 child in life. Due to stochastic fertility, families who end up with no child in life also benefit from this policy by 0.21 %.

Note that I evaluated the effect of this counterfactual policy under a non-balanced budget. This policy would cost $ 2320.93 per capita per year according to the estimated model. Next, I report the result of a counterfactual which gives the same amount of money unconditionally to all families in

[32]CEV of a family is defined same as (1.28), but uses the ex-post allocations and the life-time welfare of a family.

all years. In part 3 of Table 1.8, I report the effects of such policy. The unconditional transfer of the same amount has a more homogeneous effect across family types, but the welfare gain of such a policy is 0.54 % lower.

1.7 Optimal Child-Dependent Taxation

In this section, I solve for the optimal child-dependent taxation problem in the environment presented in Section 2.5. I use the linear spline method to approximate tax functions. I set the number $N = 5$ and location of knots $\{d_1, \cdots, d_N\} = \{0, 20000, 50000, 90000, 150000\}$ in tax functions for families with no and with one child[33]. The optimal tax problem consists of finding the amount of tax payment of families at these nodes $T = \Big\{\{T^{0k}(d_0), \cdots, T^{0k}(d_N)\}, \{T^{1k}(d_0), \cdots, T^{1k}(d_N)\}\Big\}$. The tax payment of the families with income between these knots will be determined by linear interpolation, where extrapolation will be necessary for incomes above the last knot (the first knot is always set to zero income).

The numerical problem of the social planner is to choose the 2N parameters of the tax function $\mathbb{T}(y, n)$ in order to maximize the social welfare function

$$W = \sum_{i=1}^{n_{sim}} \sum_{t=1}^{t_d - 1} \beta^{t-1} U_t^i(.|\mathbb{T}(y_t, n_t)) \tag{1.29}$$

subject to the government budget constraint

$$T_{tot} + TSS_{tot} - FS_{tot} = T_{tot}^{SQ} + TSS_{tot}^{SQ} - FS_{tot}^{SQ}, \tag{1.30}$$

which shows that the present value of total government expenditure (sum of total tax, total social security tax, and total food stamp benefits) has to be the same as the total expenditure in the status quo. $T_{tot} = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \sum_{t=0}^{T_r} \frac{1}{R} \mathbb{T}(y_t^i, n_t^i)$ is the total amount of taxes collected by the government. Note that TSS_{tot} and FS_{tot} only change due to changes in the distribution of income in the optimal tax environment. n_{sim} is the number of simulated families and $U_t^i(.|\mathbb{T}(y_t, n_t))$ is the instantaneous

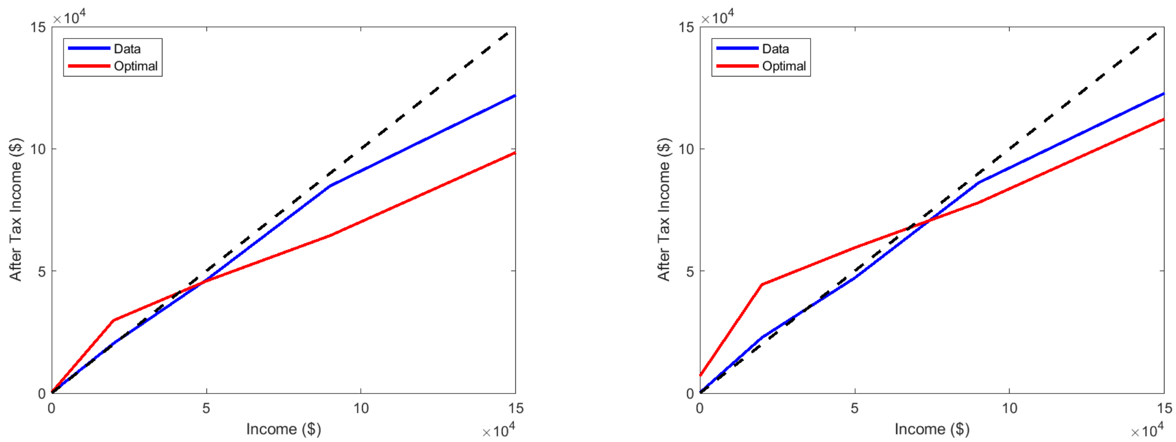
[33]Given the tax credit is the same for each additional child, there is no need to specify tax functions for two-children families.

utility of family i at t which is determined according to (1.25) and (1.4) where the decision and state variables are the result of the optimal dynamic problem of households under the tax policy $\mathbb{T}(y, n)$.

To solve this constrained maximization problem, I utilize a Nelder-Mead algorithm to find $N-1$ of tax parameters (all except the $T^{1k}(d_0)$). The last parameter is determined from the government budget balance in (1.42) and found with the help of a fixed point algorithm. I use the fact that in the absence of behavioral responses $T^{1k}(d_0)$ only affects families with income lower than d_1 to improve the speed of convergence of this fixed point problem. For the detailed explanation of the algorithm, see Appendix 1.C.2.

1.7.1 Optimal Child-Dependant Taxation Results

Figure 1.7 shows the optimal tax schedule for families with zero and one child. The black dashed line represents $45°$ line where the after-tax and before-tax incomes are the same. The optimal tax policy of families with one or two children consists of both a large guaranteed income (around $ 7,000 for a family with 1 child) and high earning subsidies (around 1.9 substitution rate) until the family's income reaches around $ 20,000. After that, the transfers will be taxed at a very high rate (around 50 %). The transfers are still positive and large for middle-income families with children with a break-even income of around $ 75,000. For families without children, the optimal policy is an EITC-type transfer with around 1.5 substitution rate by the government to low-income families, accompanied by a drastic increase in the marginal taxes of middle and high-income families compared to the status quo. Although the average taxes of high-income families with children also rise compared to the status quo, the increase is much smaller than the increase for families without children. This suggests that high-income families should also benefit from child tax credits. A couple with $ 150,000 in income who has one child will pay around $ 38,000 in taxes, around $ 13,000 less than a zero-child couple with the same income.



(A) Families with no children

(B) Families with 1 child]

Figure 1.7: After Tax Income versus Before Tax Income Under Optimal Child-Dependent Taxes

Table 1.9: Aggregate Outcomes under Optimal Child-Dependant Taxes

	Mean							Sd		
n_t	C	E	$C+E$	LP_f	H_m	Y	S_L	$C+E$	S_L	CEV
Aggregate Moments										
0	65484	-	65484	.6	41.83	80059	0	24663	0	-
1	75643	12129	80945	.43	42.77	77759	53.87	24859	4.72	-
2	72620	11234	87089	.44	41.52	82686	53.84	28593	4.78	-
All	68561	6580	72052	.52	41.99	80124	53.84	26990	4.77	-
% change Compare to Status Quo										
0	-11.02	-	-11.02	-22.19	1.53	-9.25	0	-20.38	0	-10.97
1	-8.35	-16.32	-8.91	-40.3	.49	-13.74	-.12	-26.51	-.58	.2
2	-4.86	-3.74	-4.68	-41.87	-3.85	-15.42	.36	-23.58	-5.79	11.16
All	-9.4	-7.53	-9.32	-31.07	-.07	-11.9	.29	-19.82	-5.11	11.96

Notes: Capital letters show the aggregate moments: consumption (C), child investment (E), total expenditure $(C+E)$, labor force participation of women LP_F, hours of work of men H_m, family income Y, and final score of children (S_L). Outcomes are reported in % changes compared to the status quo. Conditional CEVs are calculated based on the number of children in life.

Table 1.9 shows the aggregate moments under the optimal taxes outlined in Figure 1.7 and the percentage changes compared to the status quo[34]. The labor supply of females with children is the most affected outcome under the optimal tax policy. Labor supply participation of females decreases by 30 % overall and by 40 % for families with children. The mean total expenditure will also go down by 9.4 %. However, the inequality in total expenditure and the final test scores of children will become substantially lower. The standard deviation of total expenditure will go down by around 20 % and the standard deviation of the final score will be 5 % less compared to status quo. Families who never have any children in their life will lose welfare by around 11 % and families with 2 children in their life will gain around 12 %. The welfare of families who have 1 child in life increases by 0.2 %. The ex-ante welfare gain is 11.96 % in consumption equivalence.

To further understand the mechanisms that drive these increases in social welfare under optimal taxation, I decompose the welfare gain from optimal policy into changes in allocations of expenditure, hours of work, and child investment in Table 1.10. The consumption utility of child investment is counted in the utility of expenditures and the welfare from child investment only accounts for the welfare from the effect of child investments on children's score. I further decompose each of them into changes in (1) level, (2) distribution among families, and (3) distribution over time[35].

Table 1.10: Decomposition of Optimal Child-Dependent Tax

	Level	Distribution across Families	Distribution over Life	Total
Total Expenditure	-9.43	5.31	0.63	-3.98
Hours of Work of Females	11.84	6.25	-5.78	11.89
Hours of Work of Males	0.03	0.31	-0.01	0.17
Investment on Children (only the effect on ability)	-0.025	0.49	0.14	0.37
Total				11.89

Note: See text for detailed description.

Table 1.10 shows that most of the welfare gain from optimal taxation comes from less hours of work by females. Another large source of welfare gain is lower inequality in expenditure of families

[34] To keep the economy under the optimal tax comparable with the status quo economy, the status quo results are simulated using the current taxes approximated with the same functional form used in the optimal tax exercise.

[35] Given the non-separability of inputs in the utility function, the order of decomposition affects the results. Here, in each decomposition, I only change one of the inputs, the rest are the same as the status quo.

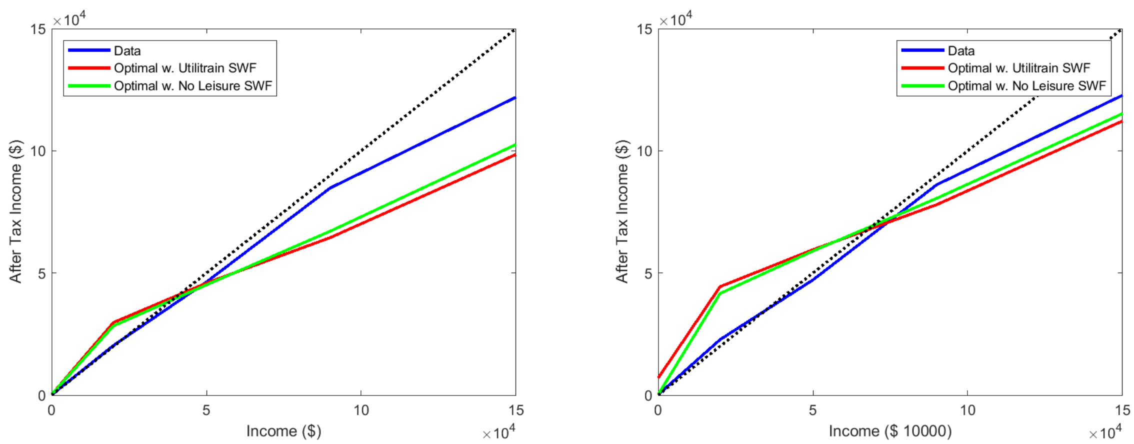
(5.31 %). Changes in the scores of children as the result of the change in investment in them causes 0.37 % of welfare gain. Although there is a welfare loss (-0.25 %) due to the lower mean level of investment in children, the social welfare from the effect of investment in scores increases by better allocation of investment across children (0.49 %) and over the life of children (0.14 %).

Under the optimal tax policy, the utilitarian social planner provides a large transfer to low and middle-income families with children to compensate for the extra costs associated with having children. This rise in redistribution increases social welfare, but it is not without costs. First, the horizontal equality principle will be violated as in any tagging policy. Also, note that the optimal policy does not represent a Pareto improvement. While families with children benefit from changing the tax system to the optimal tax policy, the couples who never have any child over their life experience significant welfare losses. Second, while the EITC type tax subsidy embedded in the tax system encourages work, especially in the extensive margin, the income effect of the large transfer causes a substantial decrease in the labor force participation of the secondary earner. Since the disutility of work of females (especially mothers of small children) is large, the social planer is willing to accept this cost, as the increase in non-working time of mothers will also increase the welfare. Notice that since wages are exogenous in the model, the cost of exclusion of mothers from labor market is underestimated.

As in Saez (2001), with the inclusion of extensive margin of labor supply, guaranteed income becomes less preferable and an income subsidy program becomes more important to encourage labor force participation at the bottom of the income distribution. However, it is more costly to redistribute to families with children to compensate for their higher cost of child-rearing, if the mother participates in the labor market. Given the none-separability of female labor force participation and consumption in the model (as in Heckman (1974)), if a mother works, the marginal utility of expenditure of her family would be higher. This can be interpreted as the cost of buying formal child care and other home-produced goods. Hence, if the mother does not work, it requires fewer transfers from the government to provide the family with the same level of utility. As for dual-earner families with children, the higher marginal utility of expenditure along with the higher cost of children induces lower taxes on top.

1.7.2 Optimal Taxation with A Social Planner Who Does Not Value Leisure

While in my model I do not differentiate between the part of non-working time that is spent on leisure and the part that is used for home production, part of the non-working time of mothers is the time spent with children and society may accept a large number of non-working mothers especially if they have small children (similar to long maternal leave policies). However, the society may dislike this policy since it increases the number of non-working individuals in a large scale. I incorporate this concern by solving for the optimal taxation problem from the viewpoint of a social planner who does not value the utility of non-working time of individuals. Table 1.8 reports the optimal tax policy with a social welfare function that sets the disutility of work of everyone to the disutility from the full time work, regardless of the work status of the individual. Compared to the optimal tax policy with the utilitarian social welfare function, the transfers offered to low and middle-income families will be lower, but still sizable. However, the social planer will completely remove the income guarantee for families with zero income.



(A) Families with no children

(B) Families with 1 child

Figure 1.8: After Tax Income versus Before Tax Income with A Social Planner Who Does Not Value Leisure

In Table 1.11, I report the percentage change in aggregate outcomes relative to the status quo. The decrease in female labor force participation rate is still noticeable. However, the decrease is 8 percentage points and 4 percentage points less for families with one child and two children, respectively.

Table 1.11: Aggregate Outcomes of Optimal Child-Dependant Taxes under A Social Planner Who Does Not Value Leisure

	Mean							Sd		
n_t	C	E	$C+E$	LP_f	H_m	Y	S_L	$C+E$	S_L	CEV
Utilitarian Social Planner (% change Compare to Status Quo)										
0	-9.74	-	-9.74	-28.19	.94	-7.04	0	-11.65	0	-6.61
1	-7.26	-12.94	-7.66	-44.36	.9	-10.38	-.04	-15.1	-1.4	1.97
2	-4.75	-3.53	-4.55	-41.95	-1.97	-12	.29	-13.64	-4.81	9.33
All	-8.37	-6.37	-8.27	-35.21	.2	-9.1	.24	-11.25	-4.37	11.73
The Social Planner Who Does Not Value Leisure (% change Compare to Status Quo)										
0	-7.64	-	-7.64	-23.39	.97	-4.98	0	-8.25	0	-6.33
1	-5.83	-9.56	-6.1	-36.55	.71	-8.91	-.01	-12.02	-1.45	1.85
2	-3.35	-.99	-2.96	-33.78	-1.05	-10.63	.31	-12.02	-4.98	9.09
All	-6.53	-3.57	-6.39	-28.92	.38	-7.36	.26	-8.57	-4.52	11.19

Notes: Capital letters show the aggregate moments: consumption (C), child investment (E), total expenditure $(C+E)$, labor force participation of women LP_F, hours of work of men H_m, family income Y, and final score of children (S_L). Outcomes are reported in % changes compared to the status quo. Conditional CEVs are calculated based on the number of children in life.

1.7.3 Effects of Restrictions on the Tax Function

There are several changes in the tax function that can influence these results. First, in this optimal tax exercise, the taxation is not conditioned on the age of the families' children. By large transfers to families with children, the social planner decreases the labor supply of all mothers, regardless of the age of their children. A tax policy that treats families with small children differently than families with adult children might suggest lower transfers for the second group. Second, inclusion of child-care tax credits can equip the social planner with a more work-friendly substitute to child tax credits. Third, considering different tax credits for the first and second child might also results in lower transfers for the second child of a family as the labor supply cost of the second child is lower. Forth, the joint taxation restriction in the model creates a high positive marginal tax on the secondary earner of families with children, which reduces the female labor force participation. Separate taxation and lowering marginal taxes on the secondary earners can incentivize women to work more[36]. Fifth, the one main problem that child-dependent taxes cannot solve is the fact the families of the same size consist of both young and old couples with large variations in wages and assets. This suggests that there is a potential for welfare gain by introducing age-dependency into the tax system. The extent of welfare gain from relaxing these restrictions on the tax function is left for future work.

1.8 Conclusion

The paper explores the design of optimal tax as a function of the number of children and income of families. First, I estimate a microeconometric life-cycle model that takes into account static tagging and dynamic tagging motivations for child-dependent taxation. The model features endogenous family labor supply, consumption-saving, and the development of children's cognitive ability in the family. It also includes heterogeneity in altruism among families and uncertainties in wages, fertility, and the child-development process over the life cycle.

Next, I use my estimated model to evaluate the effects of a tax policy currently under discussion among policy makers: permanently keeping the expanded child tax credit part of the American Rescue Plan. It has a modest negative effect on labor supply and increases children's final scores by 0.02 standard deviations. The ex-ante welfare gain from this policy is 14 % higher than an unconditional cash transfer that costs the same.

At the end, I solve for optimal child-dependant taxes by approximating the tax schedule with linear splines functions. My results signify the importance of the child dependency in a tax system as the tax treatment of families with and without children are quite different. Under a utilitarian

[36]Kleven et al. (2009) and Gayle and Shephard (2019) find that introducing negative jointness in taxation of couples can increase welfare. In Kleven et al. (2009), as income of primary-earner increases, the secondary earner contributes less to the utility of family, so the need to redistribute from the better-off dual-earner families to lone-earner families is reduced. In a model with endogenous marriage market and a home production function requiring time from both spouses, Gayle and Shephard (2019) find small welfare gain for negatives jointness.

social planner, the optimal tax policy of families with one or two children consists of both a large guaranteed income (around \$ 70,00 for a family with 1 child) and high earning subsidies (around 1.9 substitution rate) until the family's income reaches around \$ 20,000. After that, the transfers will be taxed at a very high rate (around 50 %). The transfers are still positive and large for middle-income families with children with a break-even income of around \$ 75,000.

For families without children, the optimal policy is an EITC type transfer with around 1.5 substitution rate by the government to low-income families, accompanied by a drastic increase in the marginal taxes of middle and high-income families compared to the status quo. Although the average taxes of high-income families with children also rise compared to the status quo, the increase is much smaller than the increase for families without children. This suggests that high-income families should also benefit from child tax credits. A couple with \$ 150,000 income who has one child will pay around \$ 38,000 in taxes, around \$ 13,000 less than a zero-child couple with the same income.

Chapter 1 Appendices

1.A Conditional Summary statistics

Table 1.12: Descriptive Statistics Conditional on Concurrent Number of Children

	No Child			One Child			Two Children		
	mean	Median	sd	mean	Median	sd	mean	Median	sd
Age of Female	37.19	35	(12.21)	30.49	30	(7.20)	34.07	34	(6.33)
Age of Male	38.89	37	(12.29)	32.10	31	(7.33)	35.74	36	(6.44)
Years of Schooling of Female	14.24	14	(2.25)	13.87	14	(2.17)	14.15	14	(2.27)
Years of Schooling of Male	13.97	14	(2.38)	13.62	13	(2.22)	13.87	14	(2.40)
Hourly Wage of Female	20.67	17	(17.41)	19.34	16	(19.10)	21.46	17	(34.02)
Hourly Wage of Male	28.16	22	(53.79)	24.61	20	(20.97)	30.47	24	(30.23)
Weekly Work Hours of Female	29.55	35	(21.12)	23.77	25	(24.14)	23.78	25	(24.50)
Weekly Work Hours of Male	40.19	40	(20.33)	41.46	40	(21.21)	43.83	42	(20.67)
Income of Female	30553.79	26433	(29866.27)	21869.79	16396	(25299.42)	24188.52	16658	(32222.22)
Income of Male	55721.28	44043	(73148.71)	50126.90	42884	(45030.71)	65710.04	53916	(71599.13)
Income of Family	86275.08	72691	(83099.40)	71996.69	62269	(56919.50)	89898.56	75702	(82232.45)
Assets	343842.41	87000	(1637112.85)	156082.96	41972	(477457.58)	252939.21	79349	(862763.71)
Letter-Word Score of First Child	.	.	(.)	33.21	40	(17.09)	38.03	43	(15.03)
Letter-Word Score of Second Child	.	.	(.)	.	.	(.)	35.61	41	(14.83)
Observations	23595			14094			16343		

Table 1.13: Descriptive Statistics Conditional on Total Number of Children throughout Life

	No Child			One Child			Two Children		
	mean	Median	sd	mean	Median	sd	mean	Median	sd
Age of Female	34.99	33	(10.32)	33.57	32	(9.63)	34.79	33	(9.89)
Age of Male	36.51	34	(10.28)	35.21	33	(9.87)	36.51	35	(9.97)
Years of Schooling of Female	14.12	14	(2.36)	13.88	14	(2.18)	14.21	14	(2.23)
Years of Schooling of Male	13.64	14	(2.54)	13.64	14	(2.24)	13.99	14	(2.33)
Hourly Wage of Female	20.56	17	(18.03)	19.39	16	(19.14)	21.02	17	(26.73)
Hourly Wage of Male	25.18	20	(23.69)	25.26	21	(21.92)	29.59	23	(47.44)
Weekly Work Hours of Female	30.05	35	(25.32)	27.24	32	(25.01)	24.77	28	(21.77)
Weekly Work Hours of Male	40.07	40	(23.59)	40.81	40	(22.70)	42.41	41	(19.05)
Income of Female	29982.65	25052	(31385.35)	25269.73	20916	(27488.62)	25634.15	20024	(30101.12)
Income of Male	49834.84	39983	(56858.23)	50515.22	42169	(48288.14)	61811.59	50059	(73758.75)
Income of Family	79817.49	67412	(70462.47)	75784.96	64957	(62285.50)	87445.74	73465	(82949.43)
Assets	209721.52	48887	(602201.90)	191822.29	52468	(573302.13)	328196.28	88761	(1545809.11)
Letter-Word Score of First Child	.	.	(.)	34.45	41	(16.55)	37.61	43	(15.50)
Letter-Word Score of Second Child	.	.	(.)	.	.	(.)	39.00	44	(14.45)
Observations	8755			13022			32428		

In Table 1.12, I provide descriptive statistics conditional on the number of children present in the household. In Table 1.13, I show the same statistics conditional on the total number of children throughout the life a family. The sample of one-child families consist of younger couples than two-children and no-child families. Couples who have two children in life make up the majority of my sample and have higher wages and assets than families with zero children or one child.

1.B Dynamic Solution

1.B.1 All the Value Functions

Problem of Families without Child

Families without children simply make decisions on consumption and hours of work of spouses given the constraints. State variables are $\{a_t, \xi_t^{w,f}, \xi_t^{w,m}, \iota_t^{w,f}, \iota_t^{w,m}\}$, corresponding to assets, persistent wages of the wife and the husband, and innovations to persistent wages of the wife and the husband. While the female's age is in the fertility window (Ages 22-35), in addition to uncertainty about future innovations on permanent wages, a family faces uncertainty about having a child in the next period, and conditional on having a child uncertainty about her initial cognitive ability and initial permanent and transitory cognitive ability shocks. The value function of a family without child at $t < T_{Fer}$ is

$$
\begin{aligned}
&V_t^{0C,F}(a_t, \xi_t^{w,f}, \xi_t^{w,m}, \iota_t^{w,f}, \iota_t^{w,m}; \Omega) = \max_{c_t, h_t^f, h_t^m} \Big\{ u_0(c_t, h_t^f, h_t^m; \Omega) \\
&+\beta\Big[(1 - Pr_t^b(edu^f, 0))\mathbb{E}_{\{\iota_{t+1}^{w,f}, \iota_{t+1}^{w,m}\}}\Big(V_{t+1}^{0C,F}(a_{t+1}, \xi_{t+1}^{w,f}, \xi_{t+1}^{w,m}, \iota_{t+1}^{w,f}, \iota_{t+1}^{w,m}; \Omega)\Big) \\
&+Pr_t^b(edu^f, 0)\mathbb{E}_{\{\iota_{t+1}^{w,f}, \iota_{t+1}^{w,m}, \theta_0^1, \xi_0^{\theta,1}, \epsilon_0^{\theta,1}\}}\Big(V_{t+1}^{1C,F}(a_{t+1}, \xi_{t+1}^{w,f}, \xi_{t+1}^{w,m}, \iota_{t+1}^{w,f}, \iota_{t+1}^{w,m}, \theta_0^1, \xi_0^{\theta,1}, \iota_0^{\theta,1}, \epsilon_0^{\theta,1}, 0; \Omega)\Big)\Big]\Big\},
\end{aligned}
\tag{1.31}
$$

subject to (1.6), (1.5), and (1.8). The first expectation on the right-hand side is over persistent wage shocks, and the second expectation is over the child's initial cognitive ability and the persistent and transitory shocks to it in addition to the wage shocks. $V_{t+1}^{1C,F}(.)$ is the value function of families with 1 child and its last input is the age of the new-born child 0.

When $t = T_{Fer}$, the value function is the same, except that the two value function on the right-hand side are V_{t+1}^{0C} and V_{t+1}^{1C} instead of $V_{t+1}^{0C,F}$ and $V_{t+1}^{1C,F}$, because there is no possibility of fertility in the next period.

After the fertility window of the wife ends, there is no possibility of having a child next period. Beside the first year after the child of the one-child family has left the home, the value function of no-child families at $t > T_{Fer}$ is

$$
\begin{aligned}
&V_t^{0C}(a_t, \xi_t^{w,f}, \xi_t^{w,m}, \iota_t^{w,f}, \iota_t^{w,m}; \Omega) = \max_{c_t, h_t^f, h_t^m} \Big\{ u_0(c_t, h_t^f, h_t^m; \Omega) \\
&+\beta\Big[\mathbb{E}_{\{\iota_{t+1}^{w,f}, \iota_{t+1}^{w,m}\}}\Big(V_{t+1}^{0C}(a_{t+1}, \xi_{t+1}^{w,f}, \xi_{t+1}^{w,m}, \iota_{t+1}^{w,f}, \iota_{t+1}^{w,m}; \Omega)\Big)\Big]\Big\},
\end{aligned}
\tag{1.32}
$$

subject to (1.5), (1.6), and (1.8). At the first period after the child of a one-child family turns 18 and leaves the home, the family receive utility based on her final cognitive ability, so the value function of families at those periods has additional state variable of final cognitive ability of the child

$$V_t^{0C,I}(a_t, \xi_t^{w,f}, \xi_t^{w,m}, \iota_t^{w,f}, \iota_t^{w,m}, \theta_{z_L+1}^k; \Omega) = \max_{c_t, h_t^f, h_t^m} \Big\{ u_0(c_t, h_t^f, h_t^m, \theta_{z_L+1}^k; \Omega)$$

$$+\beta \Big[\mathbb{E}_{\{\iota_{t+1}^{w,f}, \iota_{t+1}^{w,m}\}} \Big(V_{t+1}^{0C}(a_{t+1}, \xi_{t+1}^{w,f}, \xi_{t+1}^{w,m}, \iota_{t+1}^{w,f}, \iota_{t+1}^{w,m}; \Omega) \Big) \Big] \Big\}, \tag{1.33}$$

subject to (1.5), (1.6), and (1.8).

Problem of Families with One Child

Families with a child have to decide on consumption, investment in their children and hours of work of spouses given the constraints. State variables are $\{a_t, \xi_t^{w,f}, \xi_t^{w,m}, \iota_t^{w,f}, \iota_t^{w,m}, \theta_{z_t^k}^k, \xi_{z_t^k}^{\theta,k}, \iota_{z_t^k}^{\theta,k}, \epsilon_{z_t^k}^{\theta,k}, z_t^k\}$, corresponding to assets, persistent wages of the wife and the husband, innovation to persistent wages of the wife and the husband, child's cognitive ability, persistent cognitive ability, innovation on persistent cognitive ability, transitory shock in cognitive ability, and age. k is either 1 or 2 depending on whether this is the family's first or second child currently present in the household. Note that since $T_{Fer} < z_L$, when $t < T_{Fer}$, k can only be 1 and there is no possibility of child leaving the home next period. Families face uncertainty about future's innovations on permanent wages, and child's cognitive ability permanent and transitory shocks.

At $t < T_{Fer}$, value function of one-child families during this time period is

$$V_t^{1C,F}(a_t, \xi_t^{w,f}, \xi_t^{w,m}, \iota_t^{w,f}, \iota_t^{w,m}, \theta_{z_t^1}^1, \xi_{z_t^1}^{\theta,1}, \iota_{z_t^1}^{\theta,1}, \epsilon_{z_t^1}^{\theta,1}, z_t^1; \Omega) = \max_{c_t, e_{z_t^1}^1, h_t^f, h_t^m} \Big\{ u_1(c_t, e_{z_t^1}^1, h_t^f, h_t^m, \theta_{z_t^1}^1, z_t^1; \Omega)$$

$$+\beta \Big[(1 - Pr_t^b(edu^f, 1)) \mathbb{E}_{\{\iota_{t+1}^{w,f}, \iota_{t+1}^{w,m}, \iota_{z_{t+1}^1}^{\theta,1}, \epsilon_{z_{t+1}^1}^{\theta,1}\}} V_{t+1}^{1C,F}(a_{t+1}, \xi_{t+1}^{w,f}, \xi_{t+1}^{w,m}, \iota_{t+1}^{w,f}, \iota_{t+1}^{w,m}, \theta_{z_{t+1}^1}^1, \xi_{z_{t+1}^1}^{\theta,1}, \iota_{z_{t+1}^1}^{\theta,1}, \epsilon_{z_{t+1}^1}^{\theta,1}, z_{t+1}^1; \Omega)$$

$$+Pr_t^b(edu^f, 1) \mathbb{E}_{\{\iota_{t+1}^{w,f}, \iota_{t+1}^{w,m}, \iota_{z_{t+1}^1}^{\theta,1}, \epsilon_{z_{t+1}^1}^{\theta,1}, \theta_0^2, \iota_0^{\theta,2}, \epsilon_0^{\theta,2}\}} V_{t+1}^{2C}(a_{t+1}, \xi_{t+1}^{w,f}, \xi_{t+1}^{w,m}, \iota_{t+1}^{w,f}, \iota_{t+1}^{w,m},$$

$$\theta_{z_{t+1}^1}^1, \xi_{z_{t+1}^1}^{\theta,1}, \iota_{z_{t+1}^1}^{\theta,1}, \epsilon_{z_{t+1}^1}^{\theta,1}, z_{t+1}^1, \theta_0^2, \xi_0^{\theta,2}, \iota_0^{\theta,2}, \epsilon_0^{\theta,2}, 0; \Omega) \Big] \Big\}, \tag{1.34}$$

subject to (1.5), (1.6), (1.8), (1.11), and $z_{t+1}^k = z_{t+1}^k + 1$. $V_{t+1}^{2C}(.)$ is the value function of families with 2 children and its last input is the age of the second new-born child 0. When $t = T_{Fer}$, the value function is the same, except that the first value function on the right-hand side is V_{t+1}^{1C} instead of $V_{t+1}^{1C,F}$.

At $t > T_{Fer}$, the value function of one-child families during this time period beside 2 exceptions is

$$V_t^{1C}(a_t, \xi_t^{w,f}, \xi_t^{w,m}, \iota_t^{w,f}, \iota_t^{w,m}, \theta_{z_t^k}^k, \xi_{z_t^k}^{\theta,k}, \iota_{z_t^k}^{\theta,k}, \epsilon_{z_t^k}^{\theta,k}, z_t^k; \Omega) = \max_{c_t, e_{z_t^k}^k, h_t^f, h_t^m} \Big\{ u_1(c_t, e_{z_t^k}^k, h_t^f, h_t^m, \theta_{z_t^k}^k, z_t^k; \Omega)$$

$$+\beta \mathbb{E}_{\{\iota_{t+1}^{w,f}, \iota_{t+1}^{w,m}, \iota_{z_{t+1}^k}^{\theta,k}, \epsilon_{z_{t+1}^k}^{\theta,k}\}} \Big[V_{t+1}^{1C}(a_{t+1}, \xi_{t+1}^{w,f}, \xi_{t+1}^{w,m}, \iota_{t+1}^{w,f}, \iota_{t+1}^{w,m}, \theta_{z_{t+1}^k}^k, \xi_{z_{t+1}^k}^{\theta,k}, \iota_{z_{t+1}^k}^{\theta,k}, \epsilon_{z_{t+1}^k}^{\theta,k}, z_{t+1}^k; \Omega) \Big] \Big\}, \quad (1.35)$$

subject to (1.5), (1.6), (1.8),(1.11), and $z_{t+1}^k = z_{t+1}^k + 1$. The first expectation is at the first period after the child of a two-child family turns 18 and leaves the home, the family receive utility based on her final cognitive ability. The value function of families at those periods $V_t^{1C,I}(.)$ are the same except an additional state variable of final cognitive ability of the left child $\theta_{z_L+1}^2$ in $V_t^{1C}(.)$ and $u_1(.)$. The second exemption is in the last year that the child is present in the household ($z_t^k = z_L$). In that case, the value function is

$$V_t^{1C,L}(a_t, \xi_t^{w,f}, \xi_t^{w,m}, \iota_t^{w,f}, \iota_t^{w,m}, \theta_{z_L}^k, \xi_{z_L}^{\theta,k}, \iota_{z_L}^{\theta,k}, \epsilon_{z_L}^{\theta,k}, z_L; \Omega) = \max_{c_t, e_{z_L}^k, h_t^f, h_t^m} \Big\{ u_1(c_t, e_{z_L}^k, h_t^f, h_t^m, \theta_{z_L}^k, z_L; \Omega)$$

$$+\beta \mathbb{E}_{\{\iota_{t+1}^{w,f}, \iota_{t+1}^{w,m}\}} \Big[V_t^{0C,I}(a_t, \xi_{t+1}^{w,f}, \xi_{t+1}^{w,m}, \iota_{t+1}^{w,f}, \iota_{t+1}^{w,m}, \theta_{z_L+1}^k; \Omega) \Big] \Big\}, \quad (1.36)$$

subject to (1.5), (1.6), (1.8), and (1.11).

Problem of Families with Two Children

Families with two children make decisions on how to allocate their budget on consumption, investment in their first and second children and how much each of spouses should work given the constraints. State variables are $\{a_t, \xi_t^{w,f}, \xi_t^{w,m}, \iota_t^{w,f}, \iota_t^{w,m}, \theta_{z_t^1}^1, \xi_{z_t^1}^{\theta,1}, \iota_{z_t^1}^{\theta,1}, \epsilon_{z_t^1}^{\theta,1}, z_t^1, \theta_{z_t^2}^2, \xi_{z_t^2}^{\theta,2}, \iota_{z_t^2}^{\theta,2}, \epsilon_{z_t^2}^{\theta,2}, z_t^2\}$, corresponding to assets, persistent wages of the wife and the husband, innovation to persistent wages of the wife and the husband, child's cognitive ability, persistent cognitive ability, innovation on persistent cognitive ability, transitory shock on cognitive ability, and age of first and second child, respectively. Unless the age of the first (oldest) child is z_L, the value function of two-children families is

$$V_t^{2C}(a_t,\xi_t^{w,f},\xi_t^{w,m},\iota_t^{w,f},\iota_t^{w,m},\theta^1_{z_t^1},\xi^{\theta,1}_{z_t^1},\iota^{\theta,1}_{z_t^1},\epsilon^{\theta,1}_{z_t^1},z_t^1,\theta^2_{z_t^2},\xi^{\theta,2}_{z_t^2},\iota^{\theta,2}_{z_t^2},\epsilon^{\theta,2}_{z_t^2},z_t^2;\Omega) = \max_{c_t,e^1_{z_t^1},e^2_{z_t^2},h_t^f,h_t^m}\Big\{$$
$$u_1(c_t,e^1_{z_1^k},e^2_{z_2^k},h_t^f,h_t^m,\theta^1_{z_t^1},\theta^2_{z_t^2},z_t^1,z_t^2;\Omega) + \beta\mathbb{E}_{\{\iota^{w,f}_{t+1},\iota^{w,m}_{t+1},\iota^{\theta,1}_{z^1_{t+1}},\epsilon^{\theta,1}_{z^1_{t+1}},\iota^{\theta,2}_{z^2_{t+1}},\epsilon^{\theta,2}_{z^2_{t+1}}\}}\Big[$$
$$V_{t+1}^{2C}(a_{t+1},\xi_{t+1}^{w,f},\xi_{t+1}^{w,m},\iota_{t+1}^{w,f},\iota_{t+1}^{w,m},\theta^1_{z^1_{t+1}},\xi^{\theta,1}_{z^1_{t+1}},\iota^{\theta,1}_{z^1_{t+1}},\epsilon^{\theta,1}_{z^1_{t+1}},z^1_{t+1},\theta^2_{z^2_{t+1}},\xi^{\theta,2}_{z^2_{t+1}},\iota^{\theta,2}_{z^2_{t+1}},\epsilon^{\theta,2}_{z^2_{t+1}},z^2_{t+1};\Omega)\Big]\Big\}, \tag{1.37}$$

subject to (1.5), (1.6), (1.8), (1.11), and $z^k_{t+1} = z^k_{t+1} + 1$. If $z^1_t = z_L$, this is the last year that there are two children are present in the household. Therefore, the value function $V^{2C,L}_{t+1}(.)$ is the same, except that at the right-hand side $V^{2C}_{t+1}(.)$ is replaced by $V^{1C,I}_{t+1}(.)$.

$$V_t^{2C,L}(a_t,\xi_t^{w,f},\xi_t^{w,m},\iota_t^{w,f},\iota_t^{w,m},\theta^1_{z_t^1},\xi^{\theta,1}_{z_t^1},\iota^{\theta,1}_{z_t^1},\epsilon^{\theta,1}_{z_t^1},z_t^1,\theta^2_{z_t^2},\xi^{\theta,2}_{z_t^2},\iota^{\theta,2}_{z_t^2},\epsilon^{\theta,2}_{z_t^2},z_t^2;\Omega) = \max_{c_t,e^1_{z_t^1},e^2_{z_t^2},h_t^f,h_t^m}\Big\{$$
$$u_1(c_t,e^1_{z_1^k},e^2_{z_2^k},h_t^f,h_t^m,\theta^1_{z_t^1},\theta^2_{z_t^2},z_t^1,z_t^2;\Omega)$$
$$+\beta\mathbb{E}_{\{\iota^{w,f}_{t+1},\iota^{w,m}_{t+1},\iota^{\theta,k}_{z^k_{t+1}},\epsilon^{\theta,k}_{z^k_{t+1}}\}}\Big[V_{t+1}^{1C}(a_{t+1},\xi_{t+1}^{w,f},\xi_{t+1}^{w,m},\iota_{t+1}^{w,f},\iota_{t+1}^{w,m},\theta^k_{z^k_{t+1}},\xi^{\theta,k}_{z^k_{t+1}},\iota^{\theta,k}_{z^k_{t+1}},\epsilon^{\theta,k}_{z^k_{t+1}},z^k_{t+1};\Omega)\Big]\Big\}, \tag{1.38}$$

1.C Numerical Algorithms

1.C.1 Dynamic Solution

In this section, I present the algorithm used to solve to the dynamic optimization problem of one-child family after fertility window.

At each t, we have

1. Grids for:
   - Permanent exogenous states variables (Ω)
     - education level of parents: college educated or not (2×2 points)
     - altruism coefficient for the child: approximate normal distribution with 2 points (one standard deviation above and below mean)

- Temporary exogenous states variables (excluding child's skill shocks)
  - permanent wage shock of parents: approximate joint normal distribution using Gauss Hermite method with 3 points.
- Endogenous state variables (excluding child's skill).
  - asset: 14 points (between 0 $ (borrowing constraint) and 700 k $, where $2/3$ of grids are in the first $1/5$ with equal length).
  - permanent wages of parents: 8×8 points (with equal space).
  - possible ages of the child given ages of parents.
  - number of children who left the household.

2. Grid of Hours of Work of each parent
3. Tolerance (acceptable norm) for budget allocation problem
   - asset: 1000 $
   - investment in children: 500 $
4. Expected value of
   - Next period expected value function (excluding utility from child's skill) $\mathbb{E}\Big(V_t^{1C}(a_{t+1}, \mathsf{W}_{t+1}, z_{t+1}^1; \Omega)\Big)$.
   - Next period expected marginal value of asset $\mathbb{E}\Big(V_{a_{t+1},t}^{1C}(a_{t+1}, \mathsf{W}_{t+1}, z_{t+1}^1; \Omega)\Big)$

Then, given the grids and the expected value and marginal value functions

1. Loop over all the state variables.
2. Loop over all joint time allocation possibilities.
3. Solve the budget using "Solve FOCs algorithm" (explained later) and next period asset, consumption, and child investment .
4. Calculate the value function (excluding utility from child's skill).

5. Find hours of work of parents by finding the highest value function given choice of time allocation.

6. Calculate the marginal value of asset using envelope conditions.

   - $V^{1C}_{a_t,t}(a_t, \mathsf{W}_t, z^1_t; \Omega)\Big) = MU_c(c_t)$

7. Calculate the expectation of value function and marginal value of assets over the next period's temporary exogenous state variables and save them.

Solve FOCs algorithm

For each period t and each state s

1. Inputs:

   - Work hours of parents.
   - Wages of parents (Calculated using time period and state variables)
   - Grid of asset
   - The vector of expected marginal value of the next period assets defined over the grid of assets
   - Net wealth of the family

2. Process:

   (a) Check if the asset borrowing constraint holds:

   - Assume saving is binding and solve for the optimal child-specific expenditure using " Fixed point algorithm to find the optimal child-specific expenditure given a saving" and find the next period skill according to this child-specific expenditure
   - Calculate the RHS of the Euler equation of assets using liner interpolation on assets and LHS by using the derivative of utility of consumption. If the LHS is bigger than RHS, then the saving and child-specific expenditure are found

$$MU_{c_t} = \beta R \mathbb{E}\Big(V^{t+1}_{a_{t+1}}(a_{t+1}, \mathsf{W}_{t+1}, z^1_{t+1}; \Omega)\Big) \tag{1.39}$$

(b) If not, guess an amount of saving, find the child-specific expenditure using " Fixed point algorithm to find the optimal child-specific expenditure given a saving"

(c) Calculate the LHS and RHS of the Euler equation.

(d) It the | $RHS - LHS$ | is less than the error term, then we are done.

(e) If not, if $(RHS - LHS)$ is bigger than zero, then decrease the saving, otherwise increase it.

(f) Repeat step (b)-(d) until the norm is less than the error term.

Fixed point algorithm to find the optimal child-specific expenditure given a saving

The input is the same as "Solve FOCs algorithm" but with the value of assets (or total expenditure). This algorithm finds how total expenditure is divided between consumption and child investment.

1. Guess a value of child-specific expenditure and calculate the next period skill of the child

2. Calculate the LHS of the Euler equation of skill of the child from marginal utility of consumption and the RHS from the expected marginal value of child-specific expenditure.

$$MU_{c_t} = \beta \alpha^e_{z^k_t} \eta^k_{z^k_t} \frac{1}{e^k_{z^k_t}} \tag{1.40}$$

3. If the | $RHS - LHS$ | is less than the error term then we are done.

4. If not, if $(RHS - LHS)$ is bigger than zero, then decrease the saving, otherwise increase it.

5. Repeat step 1-5 until the norm is less than the error term.

1.C.2 Numerical Algorithm for Optimal Taxation

In this section, I present a simplified version of the algorithm used to solve to the optimal taxation problem. I use linear spline method to approximate tax functions. I set the number $N = 5$ and location of knots $\{d_1, \cdots, d_N\} = \{0, 20000, 50000, 90000, 150000\}$ in tax functions for families with no children and with one child (given the tax credit is the same for each additional child, no need to specify tax functions for two-children families). The optimal tax problem now consists of finding the amount of tax at these nodes $T = \left\{ \{T^{0k}(d_0), \cdots, T^{0k}(d_N)\}, \{T^{1k}(d_0), \cdots, T^{1k}(d_N)\} \right\}$ for each family type. The tax amount for the rest of the incomes will be determined by linear interpolation between these points and extrapolation for income levels more than the last knot (the first knot is always set to zero income).

The numerical problem of the social planner is to choose the 2N parameters of the tax function $\mathbb{T}(y, n)$ in order to maximize the social welfare function

$$W = \sum_{i=1}^{n_{sim}} \sum_{t=1}^{t_d - 1} \beta^{t-1} U_t^i(.|\mathbb{T}(y_t, n_t)) \tag{1.41}$$

subject to the government budget constraint

$$T_{tot} + TSS_{tot} - FS_{tot} = T_{tot}^{SQ} + TSS_{tot}^{SQ} - FS_{tot}^{SQ}, \tag{1.42}$$

which shows that the present value of total government expenditure (sum of total tax, total social security tax, and total food stamp) has to be the same as the total expenditure in the status quo. $T_{tot} = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \sum_{t=0}^{T_r} \frac{1}{R} \mathbb{T}(y_t^i, n_t^i)$ is the total amount of taxes collected by the government. Note that TSS_{tot} and FS_{tot} only change due to changes in distribution of income in the optimal tax environment. n_{sim} is the number of simulated families and $U_t^i(.|\mathbb{T}(y_t, n_t))$ is the instantaneous utility of family i at t which is determined according to (1.25) and (1.4) where the decision and state variables are the result of optimal dynamic problem of households under the tax policy $\mathbb{T}(y, n)$.

To solve this constrained maximization problem, I use the Nelder-Mead algorithm to find $N - 1$ of tax parameters (all except the $T^{1k}(d_0)$). The last parameter is determined from government budget balance in (1.42) and found using a fixed point algorithm. The only constraint on the parameters (beside the general government budget constraint) is that the average taxes should below one $T^F(d_i) \leq d_i,\ F \in \{0k, 1k\}$.

Define $T_{iter}^{-1} = \left\{ \{T^{0k}(d_0), \cdots, T^{0k}(d_N)\}, \{T^{1k}(d_1), \cdots, T^{1k}(d_N)\} \right\}_{iter}$ the tax parameters in *iter* iteration of Nelder-Mead algorithm, T_{iter}^{Last} the $T^{1k}(d_0)$ that satisfies (1.42) under T_{iter}^{-1}, T_{iter} =

$\{T^{Last}_{iter}, T^{-1}_{iter}\}$, and $\{y^i_t \mid \forall i, t\}_{iter}$ the optimal income of all families at all times under T_{iter} tax policy. Here is the steps of the algorithm after $i_{iter\text{-}1}$ iteration in the Nelder-Mead algorithm.

Inputs: T^{-1}_{iter} and $\{y^i_t \mid \forall i, t\}_{iter\text{-}1}$

Steps to find T^{Last}_{iter}:

1. Check if the error term from (1.42) with T^{-1}_{iter} and the highest value for T^{Last}_{iter} ($d_0 = 0$) is positive. If not, Set the objective function to a negative punishment number proportional to error and exit.

2. Choose the initial value for T^{Last}_{iter}, by finding T^{Last}_{iter} that satisfies the government budget constraint (1.42) given $\{y^i_t \mid \forall i, t\}_{iter\text{-}1}$ (assuming no behavioral response from the previous iteration).

3. Solve the dynamic problem and find the error in (1.42) from T^{-1}_{iter} and initial value for T^{Last}_{iter}.

4. Update T^{Last}_{iter}, so that the error term in (1.42) would become T_{Up} of its current value, given no behavioral response from the new tax policy.

5. Repeat [3] with the updated tax policy. Repeat [4] to [5] until it converges (the difference in total tax payment is less than the tolerance tol_T).

I repeat this process until the Nelder-Mead algorithm converges with tol_{NM}.

1.D Details of Estimation

1.D.1 Estimation of Wage Processes

Rewriting (1.8) with measurement error, we have:

$$\log(\omega^j_{it}) = \beta(X^j_{it})' + \xi^{\omega,j}_t + \nu^j_{it} \tag{1.43}$$

Where X^j_{it} the vector of person's characteristics $(1, (Age^j_{it}), (Age^j_{it})^2, (Age^j_{it}) \times (Edu^j_{it}), (Age^K_{it}))$

and

$$\begin{bmatrix} \nu_{it}^{f} \\ \nu_{it}^{m} \end{bmatrix} \sim \mathbb{N}(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\nu^f}^2 & \rho_{\nu_{f,m}} \\ \rho_{\nu_{f,m}} & \sigma_{\nu^m}^2 \end{bmatrix}) \tag{1.44}$$

is the measurement error. Define $\tilde{\omega}_{it}^{j} = \log(\omega_{it}^{j}) - \beta(X_{it}^{j})'$ and $u_{it}^{j} = \xi_{it}^{\omega,j} + \nu_{it}^{j}$ the residual of regression $\log(\omega_{it}^{j})$ on X_{it}^{j}. The parameters of wage functions β can be estimated using this regressions

$$\log(\omega_{it}^{j}) = \beta' X_{it}^{j} + u_{it}^{j}. \tag{1.45}$$

For the estimation of variance and correlation of the permanent wage shocks I use GMM approach. Taking the s-order difference from u_{it}^{j}, we have

$$\Delta^s u_{it}^{j} = \sum_{k=0}^{s-1} \iota_{it-k}^{\omega,j} + \Delta^s \nu_{it}^{j} \tag{1.46}$$

Since my data of wages are biannual, I use $s = 2$ in my estimation.

I use the following moments of $\Delta^s u_{it}^{j}$ as my moments conditions.

$$\mathbb{E}\big(\Delta^s (u_{it}^{j})^2\big) = s \cdot \sigma_{\iota^{\omega,j}}^2 + 2\sigma_{\nu^j}^2 \tag{1.47}$$

$$\mathbb{E}\big(\Delta^s u_{it}^{j} \Delta^s u_{it-s}^{j}\big) = -\sigma_{\nu^j}^2 \tag{1.48}$$

$$\mathbb{E}\big(\Delta^s u_{it}^{f} \Delta^s u_{it}^{m}\big) = s \cdot \rho_{\iota_{\omega,f,m}} + 2\rho_{\nu_{f,m}} \tag{1.49}$$

$$\mathbb{E}\big(\Delta^s u_{it}^{f} \Delta^s u_{it-s}^{m}\big) = -\rho_{\nu_{f,m}} \tag{1.50}$$

1.D.2 Approximation of the Tax Function

To approximate the tax functions in the model, I utilize a non-parametric method. I approximate the tax functions using linear splines with fixed knots similar to Gayle and Shephard (2019). I fix the number of knots to $N_T = 10$ and the locations of the knots to $\{d_1, \cdots, d_N\}$ $= \{12500\$, 25000\$, 37500\$, 55000\$, 75000\$, 105000\$, 150000\$, 200000\$, 250000\$\}$. To lower computational cost of the optimal tax problem, I am assuming the tax credit for the first and second child are the same, even though they are slightly different in the data. I assume the tax credit for each additional child is the average of total child tax credits of the first and the second child and estimate the child tax credit using the linear splines with the same knots as in the tax function[37].

To estimate the parameters of the approximated tax function, I calculate the tax payment of families with income from 0 $ to 250,000 $ with 100 $ increments for 1980-2015 using NBER TAXIM[38]. I assume that the spouses who file taxes jointly and do not claim any child-care or other tax-exempt expenditures. I pool all the data together (after transforming all of them to 2015 $) and separately estimate the tax payment of childless married couples and the average total child tax credit using via regressions.

Figure 1.9 compares the approximated tax payment versus the actual tax payment in this period.

[37]Total child tax credit is defined as the change in the tax payment of a household given an additional child.

[38]The upper bound on income is based on my sample, in which only a small number of households in it have income above $ 250,000. Given the data limitations, this paper is not able to discuss marginal income tax of very high-income families.

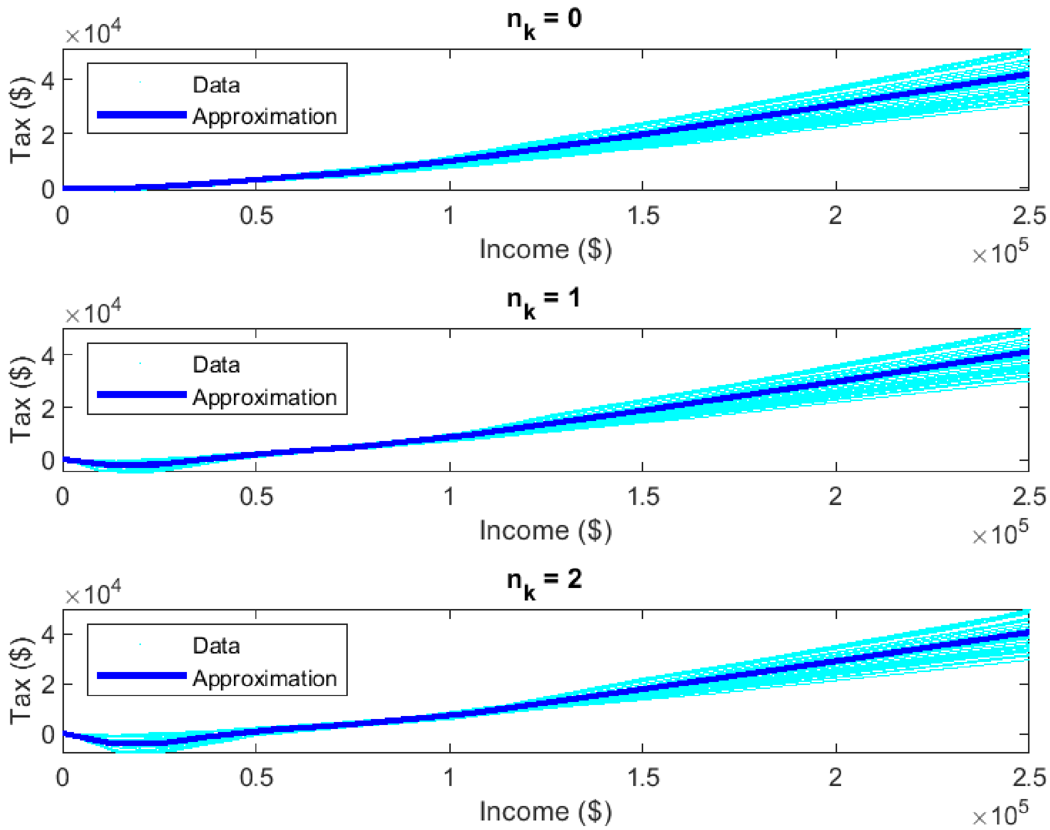


Figure 1.9: Comparing tax Data from 1980-2015 with the approximated tax functions

1.D.3 Fit of targeted Moments

Tables 1.15 to 1.18 report model-implied moments and compare them with the data. They show model matches in all the moments of asset holding, labor supply, and children's score used in the simulated method of moment estimation.

$Age_t^{i,m}$	n_t^i	$z_t^{i,1}$	$z_t^{i,2}$	Data	Model
$22 \leq . < 25$	0	-	-	15003.39	13994.82
$25 \leq . < 30$	0	-	-	16441.83	21356.42
$30 \leq . < 35$	0	-	-	33880.9	42214.71
$35 \leq . < 40$	0	-	-	64553.34	84740.93
$40 \leq . < 45$	0	-	-	79218.11	125588.7
$45 \leq . < 50$	0	-	-	105727.5	163315
$50 \leq . < 55$	0	-	-	144051.5	205530.7
$55 \leq . < 60$	0	-	-	170698.5	224492
$22 \leq . < 25$	1	All	-	12233.64	14020.03
$25 \leq . < 30$	1	All	-	21202.03	21106.43
$30 \leq . < 35$	1	All	-	40390.4	41905.23
$35 \leq . < 40$	1	All	-	64239.5	83286.71
$40 \leq . < 45$	1	All	-	91790.96	125677.5
$45 \leq . < 50$	1	All	-	121174.9	163990.6
$50 \leq . < 55$	1	All	-	146421.1	204810.9
$55 \leq . < 60$	1	All	-	212782.6	224677.5
$22 \leq . < 25$	2	All	All	16039.62	14872.17
$25 \leq . < 30$	2	All	All	27229.92	22233.94
$30 \leq . < 35$	2	All	All	57583.03	44301.3
$35 \leq . < 40$	2	All	All	100000	89598.72
$40 \leq . < 45$	2	All	All	145370.4	131643.7
$45 \leq . < 50$	2	All	All	180025.2	170561.8
$50 \leq . < 55$	2	All	All	257000	212277.3
$55 \leq . < 60$	2	All	All	323503.3	233130.4

Table 1.14: Model Fit of Asset Moments

Table 1.15: Model Fit of Work Hour Moments

	Data	Model
Mean of Work Hours of Male	40.28098	40.31501
Share of male in part time work	.1178348	.1178225
LFP rate of Female	.7548632	.7595825
with at least one child below 6 years old	.6725142	.6674231

Letter-Word score at first observation of child		
	Data	Model
Mean	6.189189	4.957592
Standard deviation	3.526391	2.991563
Letter-Word score at final age of child (17)		
	Data	Model
Mean	51.60976	52.60382
Standard deviation	3.375485	5.53345

Mean of Letter-Word score at different stages		
	Data	Model
$3 \leq AgesofChild < 7$	12.67485	13.45348
$7 \leq AgesofChild < 13$	40.26567	40.46244
$13 \leq AgesofChild < 18$	49.62465	49.62612

Table 1.16: Model Fit of Moments Related to Child Development Process (Mean and Standard Deviations)

Correlation of $log(1 + LW\ Score)$ and 5 years ago		
	Data	Model
$log(1 + LW\ Score)$	.5792834	.6531667
$log(1 + Family\ Income)$	.2674577	.2862937
Regression coefficients of $log(1 + LW\ Score)$ on		
	Data	Model
Constant	.2896065	.5641883
Age	.5572309	.521095
Age^2	-.0209444	-.0195821

Regression coefficients of $log(1 + LW\ Score)$ on 5 years ago		
	Data	Model
Constant	3.468318	3.497069
$log(1 + LW\ Score)$	.1901484	.1130985
Interaction of *Age* and (1+*LW Score*)	-.0171459	-.0062394
Interaction of Age^2 and (1+*LW Score*)	.0010029	.0006059
Regression coefficients of log(1+LW score) on last year		
	Data	Model
Constant	2.752746	3.070212
$log(1 + Family\ Income)$	-.1725475	-.1765085
Interaction of *Age* and (1+*Family Income*)	.0454663	.0426066
Interaction of Age^2 and (1+*Family Income*)	-.0018487	-.0017357
Number of children	.0200423	-.006454

Table 1.17: Model Fit of Moments Related to Child Development Process (Correlations and Partial Regressions)

Table 1.18: Model Fit of Moments Related to Child Development Process (Full Regression)

Regression coefficients of $log(1 + \textit{LW Score})$ on 5 years ago		
	Data	Model
Constant	3.740268	3.040915
Age	-.1222445	-.0286999
$log(1 + \textit{LW Score})$	-.0433168	.0245328
Interaction of *Age* and (1+*LW Score*)	.0340935	.0167934
$log(1 + \textit{Family Income})$	.0323854	.0605845
Interaction of *Age* and (1+*Family Income*)	-.0010639	-.0020681
Number of children	.0232005	-.0068075
Standard Error of Residuals	.0956739	.1034268
Autocorrolation of Residuals	.2279518	.3933409

Chapter 2

History Dependence of Pension Systems

2.1 Introduction

The modest numbers on the bottom right of social security checks are the main sources of income for many American retirees and the last barrier for keeping them out of poverty. In addition, the incentive system that they build shape the leading public policy in influencing the retirement age of workers, as well as their labor supply in old age.

A retiree's public pension benefit is based on the history of her earnings before retirement. In the US, like most countries, two functions working in the composite determine the amount of pension benefits. One function summarizes all of the history of earnings into one outcome, and the other finds the amount of benefit based on that outcome. In the US, ignoring some details, the history of earnings will be summarized by taking the average of the top 35 years of earnings of a worker. Then, this result will yield the retirement benefit after passing through a progressive benefit function. Studying the ramifications of the design of the history-dependent part of the pension system and exploring the paths to improve it is the main focus of this paper.

It is crucial that this design be guided by how it affects the redistribution, insurance, and incentives system that it creates for workers. As it will be discussed in Section 2.2, different countries institute different sets of rules to calculate the pension benefits from the history of earnings. This creates a different balance of these often conflicting criteria in each country.

- **Redistribution:** The social security system is an integral part of the overall redistribution system that taxes and transfers construct. Workers who have the less steep trajectory of earnings over the life cycle would benefit more if years of the highest earnings (which mostly happen late in the life cycle) have less influence in determining their pension benefits. In addition, the workers with less attachment to the workforce who experience many years of no earnings would be overlooked if an extended number of years are counted in determining the eligibility and the amount of the pension benefit. Moreover, as the only part of the redistribution system that is history-dependent, pension benefits can be used to correct the negative redistribution toward workers with high fluctuations in their earnings caused by progressive taxation.

- **Insurance:** If negative earning shocks during working life heavily influence the workers' resources during their retirement, the system is not providing enough insurance for workers. More insurance can be achieved by lowering the weight of low-earning years on determining retirement benefits and providing a non-history dependent part for workers with limited earning history (i. e. a consumption floor).

- **Incentives:** In old age, the decisions on how many hours to work and how much longer to stay in the labor force are heavily influenced by how these decisions affect the amount of pension benefits. If the earnings of these last years are highly weighted in determining the benefits, the system is providing more work incentives.

The history-dependent part of pension systems is vastly understudied. Although many papers have researched design of the benefit function, design of the summarizing function has been mostly ignored. It is noteworthy that this concern was raised by Peter Diamond in his Presidential Address on social security delivered during American Economic Association meeting in 2004 ((Diamond (2004)):

> "Depending on the nature of the underlying stochastic process of wage rates, both under weighting early years (relative to the use of interest rates) and not counting some low

> years may or may not help with insuring lifetime earnings. This is not an area that has received much research attention."

Understanding the consequences of the design of the history-dependent part of the pension system is the goal of this research. First, I utilize a long panel of data from the Panel Study of Income Dynamics (PSID) and introduce various stylized facts regarding retirement age, labor supply, and wages over the life cycle, and how different ways to summarise the history of earnings of individuals affect them. Some of the presented empirical regularities are updates of empirical facts in the literature with longer time series that span the whole life cycle of individuals, and some are new empirical regularities.

I show that when a smaller number the high-earning years are considered, the mean of average income increases and the standard deviation decreases. The more low-earning years are discarded, the system provides more insurance for the workers, as the negative years will not be considered in the function of the history of earnings. However, it also means more weight to the highest earning years which already represent more inequality than other years. The result is more inequality as fewer years are counted. On the other hand, workers' earnings fall for all workers late during their life cycle. Hence, when we use the latest years of earnings instead of the highest years of earnings, increasing the number of years decreases the mean of average earnings. As inequality is highest late in the life cycle, the concentration of average earnings also decreases by counting more years.

Next, I develop and parameterize a life cycle model of labor supply, saving, and retirement with labor market shocks. I utilize the data from the PSID to discipline the quantitative model. I employ the quantitative model and evaluate the effects of an important counterfactual policy. The new policy changes the current US pension policy which uses the top 35 years of earnings to account for all of the lifetime earnings, common in other OECD countries. Moving to the new policy regime causes 43 % rise in consumption. Both college-educated and non-college-educated groups benefit from this policy. Part of this higher consumption comes from higher pension benefits as the PIA rules remain the same. It also results in 64 % increase in hours of work, 5.7 % decrease in the age of retirement.

2.1.1 Related Literature

This paper mainly contributes to the literature on the design of pension systems (Diamond and Mirrlees (1978), Diamond and Mirrlees (1986) Huggett and Parra (2010), Golosov et al. (2013), Shourideh and Troshkin (2017), Grochulski and Kocherlakota (2010) and Ndiaye (2020)). None of these papers study the design of the history-dependence part of the pension systems. This paper also relates to the literature on labor supply and retirement literature decisions of workers (Fan et al. (2022), and Borella et al. (2023), O'Dea (2017), French (2005)). None of these papers consider the history-dependence of the pension systems or the optimal design of pension systems.

History-Dependent public transfer systems have been studied in the optimal taxation literature. Batzer (2021) studies optimal tax schedule when the tax function depends on all the history of income in a general equilibrium OLG model similar to Heathcote et al. (2017). Kapička (2022) studies the history dependence of taxation in a life cycle model and finds that a simple tax function that depends on the few past years can increase welfare. In the literature on income averaging and taxation on lifetime earnings, work of Vickrey (1947) is notwithstanding. Vickrey (1947) was concerned with the impact of progressive annual taxes on those with fluctuating incomes relative to those with constant incomes. However, using a longer period for determining taxes is likely to reduce the built-in stabilization from the income tax and lessen the easing of borrowing constraints. In Diamond (2009) absence of age and history-related tax rules counts as one of the reasons for a mandatory retirement pension System to redistribute based on lifetime between fluctuating and constant income individuals.

The remainder of this paper is organized as follows. In Section 2.2 I discuss history dependence of pension systems in the US and other countries. Section 2.3 describes the data. In Section 2.4, I present some empirical regularities. In Section 2.5, I develop the model and characterize the solution. In Section 2.6, I build the quantitative environment. In Section 2.7, I evaluate the effects of an important counterfactual policy. Section 2.8 concludes.

2.2 History Dependence of Pension Systems Around the World

In this section, I explain in more detail the history-dependence part of the US pension system and compare it to defined-benefit pension plans in other countries around the world.

2.2.1 History Dependence of Public Pension System in the US

In the US, the two functions that determine an individual's Social Security benefits are Average Indexed Monthly Earnings (AIME) and the Primary Insurance Amount (PIA). The top 35 years of earnings are used to calculate the AIME. At first, each year's earnings are indexed to reflect the growth in the economy and wage levels during workers' employment years. After the top 35 years of indexed earnings are determined, the indexed earnings will be summed and divided by the total number of months. Then, the average amount will be rounded down to the next lower dollar amount. The result is called AIME.

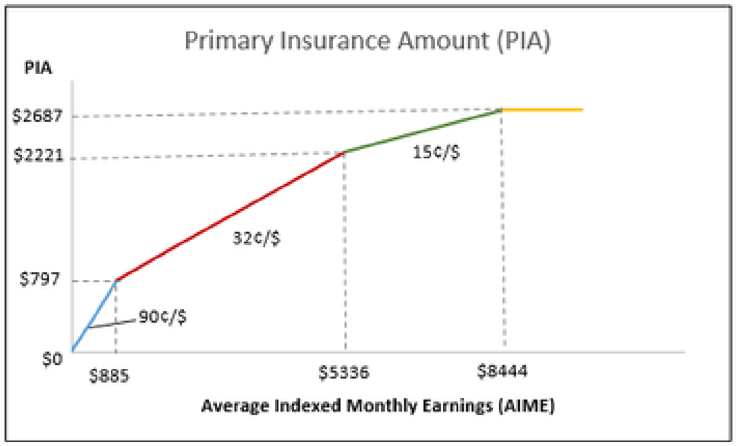
$$\underbrace{\text{AIME}}_{\substack{\text{Average Indexed}\\ \text{Monthly Earnings}}} = f(y_1, \cdots, y_T) = \frac{1}{12} \times \frac{1}{35} \times \sum_{\substack{\text{Top 35}\\ y_i \in \{\text{ years of }\}\\ \text{earnings}}} (y_i). \tag{2.1}$$

The second function takes AIME as given and determine the amount of monthly benefit (PIA).

$$\underbrace{\text{PIA}}_{\substack{\text{Primary Insurance}\\ \text{Amount}}} = g(AIME). \tag{2.2}$$

The current function g in the US is presented in 2.1.

Figure 2.1: Function g in the US: Determining PIA from AIME



2.2.2 History Dependence of Pension Around the world

There are three types of $f(.)$ that are most common across countries.

1. **Average Earnings of the Last L Years:** Although various OECD countries have moved from this method, this method is still employed among countries like France, Greece, Portugal, Spain, Norway, and Sweden. There are also numerous countries among developing countries that utilize this method. Thailand, Brazil, and Iran are some examples of developing counties.

2. **Average Life Time Earning, Excluding Some of the Lowest Earnings:** This method is used in the US in Canada. In the US, the highest 35 years determine the pension and in Canada, the lowest 15 % of earnings are excluded from the history.

3. **Average Life Time Earning:** In this method, the average of all of the earnings over the lifetime is used to calculate the pension benefit. Currently, this method is the most common method in OECD countries. Over the past decades, most OECD countries have moved from the second method (counting the last 10-25 years) to this method. Among developing countries, China, Indonesia, and Vietnam utilize all of their lifetime earnings.

Although the focus of this paper is public pension systems, it is also noteworthy that the first method is the most common method among private defined-benefit pension systems (like teacher unions)[1].

2.3 Data

2.3.1 Data Sources

To demonstrate empirical regularities and parameterize the empirical model, I use data from the Panel Study of Income Dynamics (PSID). The PSID is the longest running longitudinal household survey in the world. It began as a nationally representative sample of 5,000 families in the United States. I use the data from 1968-2019 and only consider males who are head of their household.

2.3.2 Summary statistics

Table 2.1 reports the descriptive summary of the final sample. All monetary variables are in 2015 dollars and all time variables are in annual units.

[1] For more information, see Ipp

Table 2.1: Descriptive Statistics of the Sample

	mean	25th percentile	Median	75th percentile	sd
Age	38.35	26.00	35.00	48.00	(15.62)
Years in School	13.17	12.00	13.00	16.00	(2.65)
College Degree	0.26	0.00	0.00	1.00	(0.44)
Work Hours (Before Retirement)	2068.78	1811.00	2058.00	2448.00	(741.24)
Labor Force Particpipation (Before Retirement)	0.89	1.00	1.00	1.00	(0.32)
Wage	24.93	13.42	20.61	31.18	(17.25)
Income (Before Retirement)	46188.68	20016.67	39400.19	63239.67	(38765.32)
Asset (Before Retirement)	234149.63	3129.88	43796.62	177682.78	(1124018.63)
Asset (After Retirement)	479893.73	28276.99	162900.42	489482.62	(1509194.48)
Social Security Income	13473.01	8682.77	12895.96	17435.62	(7085.43)
Age of Retirement (Based on Self Report)	50.01	39.00	59.00	63.00	(19.73)
Age of Retirement (Based on Receving SS Income)	57.94	41.00	65.00	72.00	(19.90)
Age of Retirement (Based on Stop Working)	60.07	55.00	61.00	65.00	(9.71)
Observations	394093				

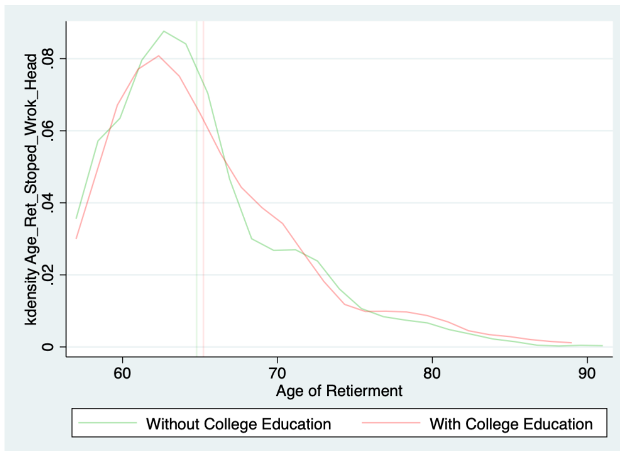
2.4 Empirical Regularities

In this section, I present some stylized facts regarding (1) retirement age, (2) labor supply and wages over the life cycle, and (3) how different ways to summarise the history of earnings of individuals would affect them. Sections 2.4.1 and 2.4.2 update some empirical regularities in the literature (i.e. Rupert and Zanella (2015)) with longer time series that span the whole life cycle of individuals. Section 2.4.3 shows some new empirical regularities.

2.4.1 Age of Retirement

Figure 2.2 shows the histogram of the age of retirement defined as when workers stopped working and never go back to work after that. If the whole history of earnings is not observed for an individual, I consider her retired at a given age if I observe at least a consensus 5 years of no income (at least one of them after age 55) after that year. The average age of retirement is 65.21 for college-educated workers and 64.78 for non-college-educated ones. There is noticeable heterogeneity in the age of retirement of workers. The standard deviation of the age of retirement of workers with a college degree is 6.05 years and for workers without a college degree is 5.69 years.

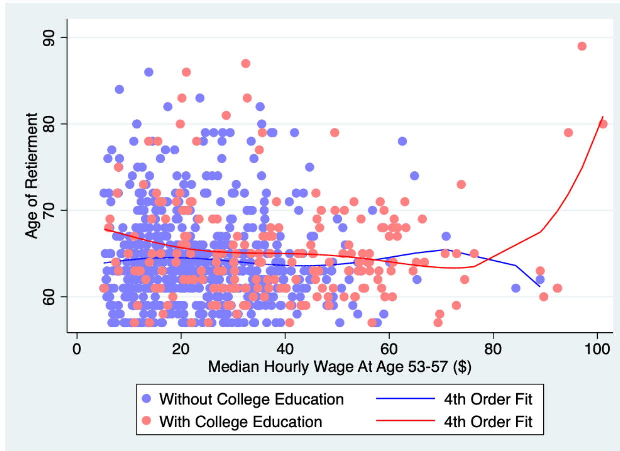
Figure 2.2: Histogram of Age of Retirement Conditional on Education level



Notes: The horizontal lines show the mean of variables.

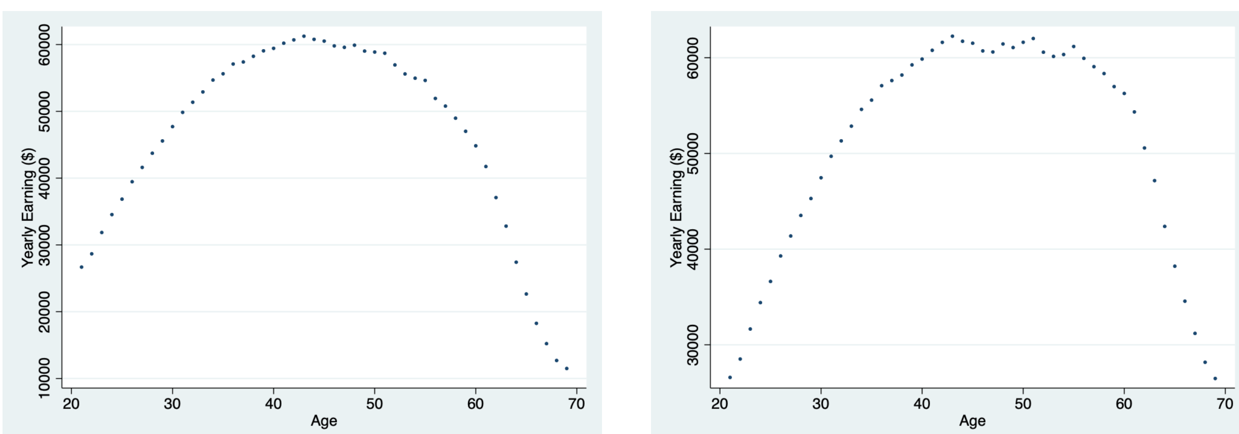
Figure 2.3 shows how the age of retirement varies by the hourly age around age 55 for college and non-college-educated individuals. For non-college-educated workers, the age of retirement is not affected by wages in old age. However, for college-educated workers, the age of retirement decreases by old-age wages and reverses for very high-wage workers.

Figure 2.3: Relationship Between Median Hourly Wage at Age 53-57 and Age of Retirement



2.4.2 Work Hours, Wage, and Earning over the Life Cycle

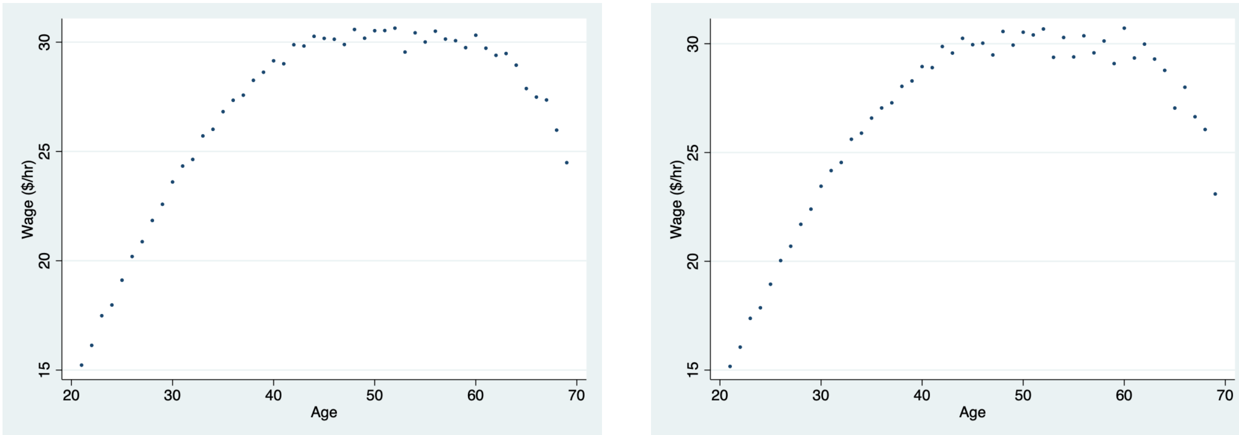
The figure in Figure 2.4 shows the earnings of workers over the life cycle from age 20 to 70. Earnings rise at the beginning of the life cycle, pick at age 42, and fall after that. To see how this pattern at the end of workers' careers is affected by workers leaving the workforce in old age, I only consider workers who stay in the labor market at least until age 70 next. For this sub-sample, the mean of earnings flattens from age 42 to 60, then falls. Hence, the early fall in earnings after age 42 can be explained by the lower earnings of workers leaving the labor market sooner.



(A) All Workers

(B) Workers Who Retire After Age 70

Figure 2.4: Mean of Workers' Earning over the Life Cycle



(A) All Workers

(B) Workers Who Retire After Age 70

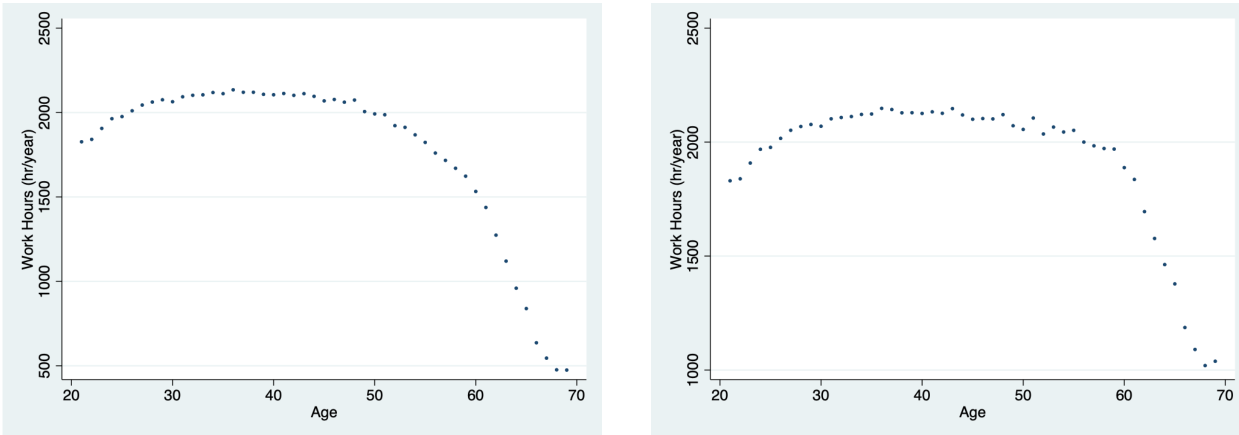
Figure 2.5: Mean of Workers' Wages over the Life Cycle

In Figure 2.7, I show how the life cycle pattern has changed for different generations. I group individuals born in the same 5 years period into the same generation. Given the data, the workers are divided into 9 generations, where the youngest generation was born in 1963 and the oldest one was born in 1923. For some of these generations, I have their whole full history of income[2]. As can be seen, the fall in earnings after age 42 is influenced the most by the generations born before 1942. Means of earnings in all groups represent hump shape patterns and drop after age 55.

To see how the pattern of earnings is shaped by changes in wages and work hours over the life cycle, I focus on these two components of earnings next. Figure 2.5 shows the wage profile of workers. It rises and picks at around age 42, but it reveals a much flatter trajectory after that, compared to earnings. It stays mostly flat for all workers and the sub-sample of workers who remain in the workforce at least until age 70 up until age 60. Then, it starts to fall for both samples. Figure 2.8 shows almost the same pattern for all cohorts.

Figure 2.6 shows the hours of work of individuals. The workers increase their work hours until age 35 and work almost the same hours until age 48 when the mean of work hours starts to decrease. Focusing on a sample of late retirees, they work almost the same hours until age 55, then decrease their work hours gradually after that until age 60 when their work hours start to rapidly decrease.

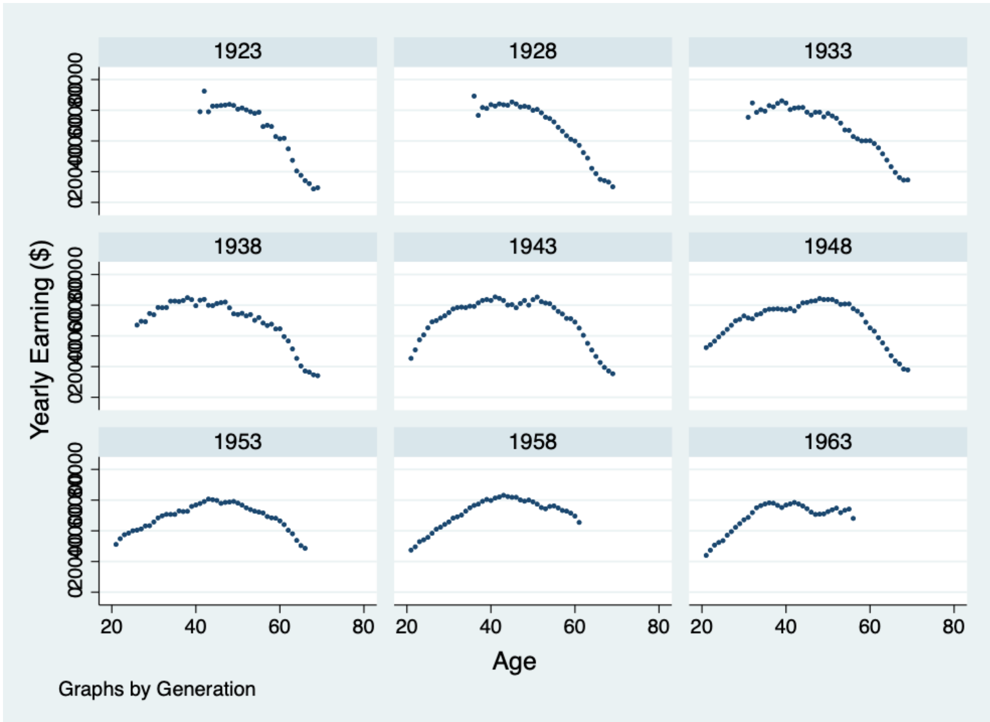
[2] After 1999, the PSID surveys biannually. For the outcomes between two survey years, I take the average of neighboring years.



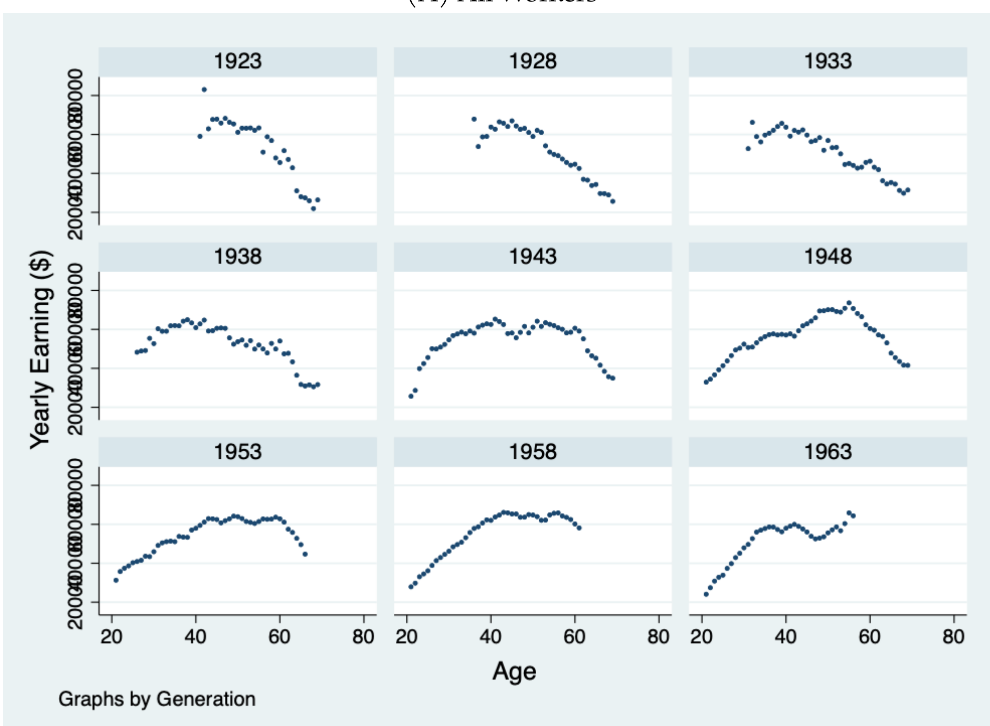
(A) All Workers (B) Workers Who Retire After Age 70

Figure 2.6: Mean of Workers' Work Hours over the Life Cycle

Figure 2.9 shows the change in work hours for different cohorts. We can see that earlier generations start decreasing their work hour sooner.



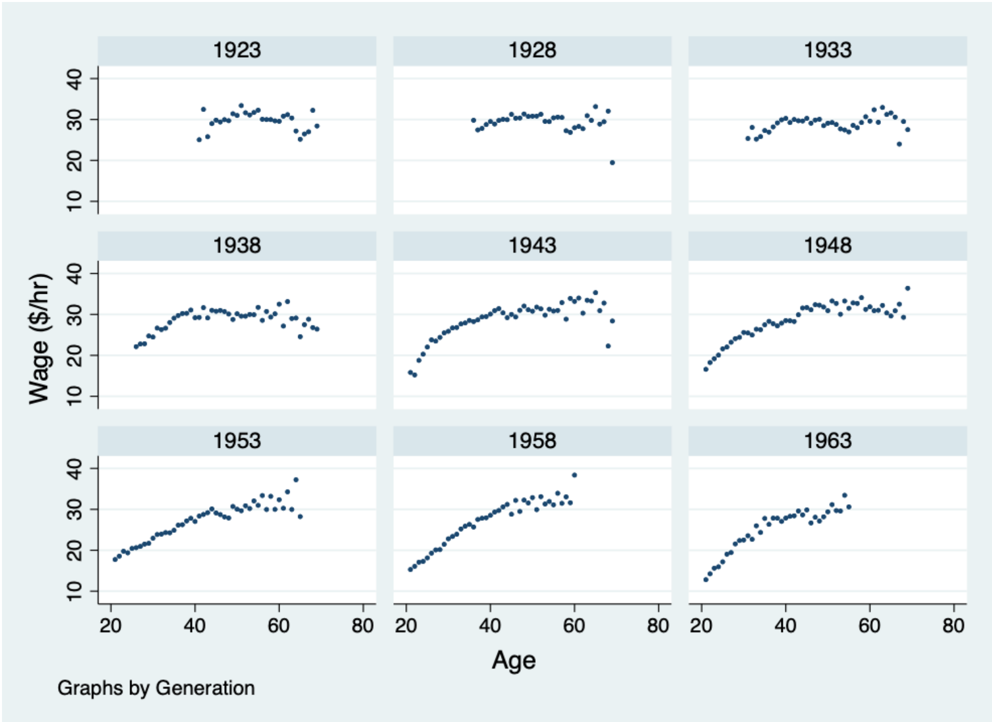
(A) All Workers



(B) Workers Who Retire After Age 70

Notes: Each generation is defined as workers who born on that year and the next 4 years.

Figure 2.7: Mean of Workers' Earning Over the Life Cycle Conditional on Generation

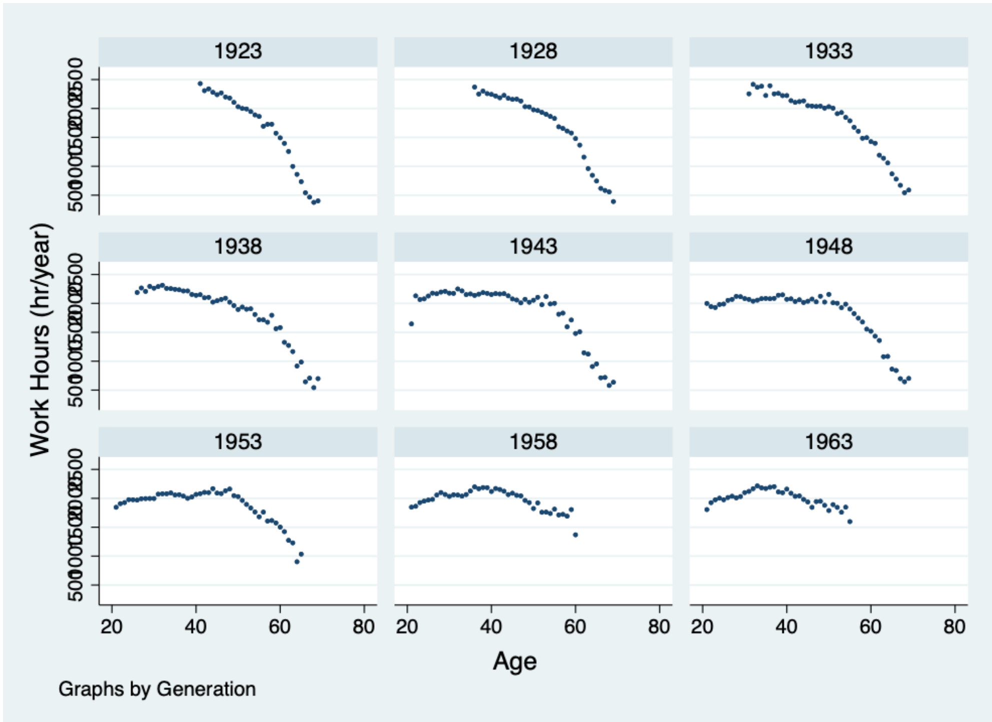


(A) All Workers

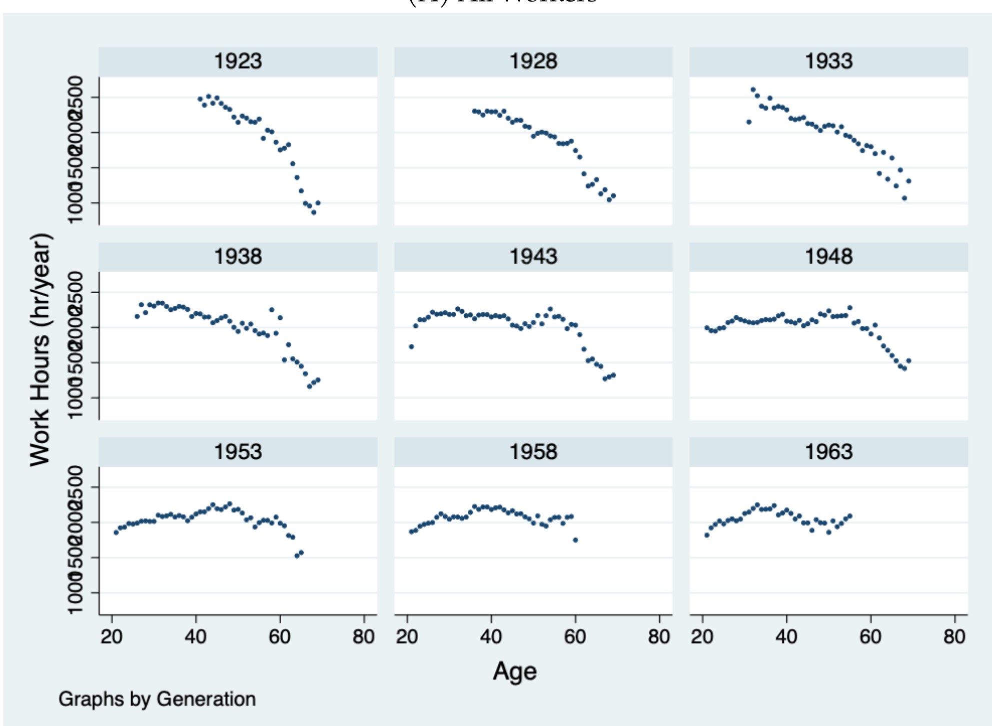
(B) Workers Who Retire After Age 70

Notes: Each generation is defined as workers who born on that year and the next 4 years.

Figure 2.8: Mean of Workers' Wages Over the Life Cycle Conditional on Generation



(A) All Workers



(B) Workers Who Retire After Age 70

Notes: Each generation is defined as workers who born on that year and the next 4 years.

Figure 2.9: Mean of Workers' Work Hours Over the Life Cycle Conditional on Generation

2.4.3 Comparing Different Functions of History of Earnings

In this section, I discuss how different ways to calculate the history of earnings would generate different results for workers. How many and which years counted for pension benefit calculation affect inequality of outcomes of workers in several dimensions (1) inequality between groups with a steep trajectory of wages over the life cycle (i.e. college-educated workers) and workers with a less upward trajectory of wages (i.e. high schooled dropouts); (2) inequality between groups with higher fluctuation in their earnings (i.e. farmers) and groups with more stable earning profiles (i.e. public employees); and (3) inequality between groups who have experienced long spells of unemployment (i. e. temporary workers) or have been out of the workforce for a long time (i. e. stay at home parents) and the workers with steady work history.

Utilizing Highest Earnings Years

Here, I discuss how different ways to summarise history of earnings affect inequality in general. In Table 2.2, I show the mean and standard deviation of outcomes of different functions. As expected, when more of the low-earning years are ignored, the mean becomes higher. The fourth column in Table 2.2 illustrates that the standard deviation of these variables decreases as more high-earning years are counted. The more low-earning years are discarded, it would provide more insurance for the workers as those bad years will not be counted in the function of the history of earnings. However, it also means more weight to the highest earning years which already represent more inequality than other years. The result is more inequality as more years are counted.

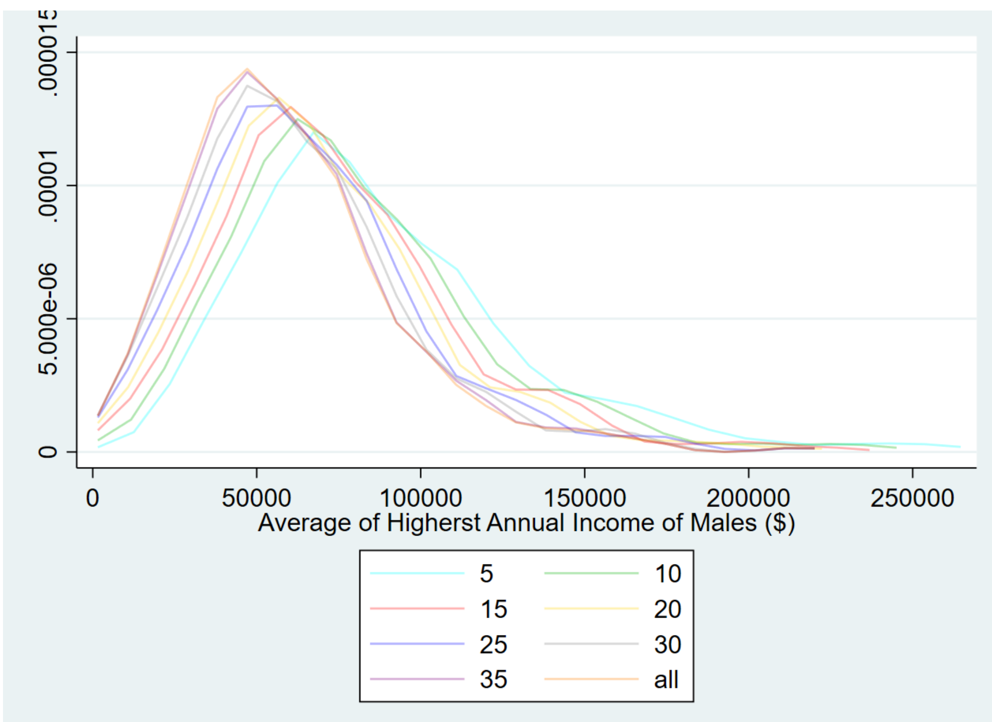
Table 2.2: Functions of History of Earnings Based on Highest Earnings

Mean of earning	Mean	Std. dev.
Highest 5 years	87830.09	42340.5
Highest 10 years	79764.62	38506.3
Highest 15 years	73878.37	36446.99
Highest 20 years	69272.62	34965.17
Highest 25 years	65449.93	33921.47
Highest 30 years	62401.4	33071.58
Highest 35 years	60402.91	32402.08
All	59815.76	32216.14

Notes: The number of observations are 814.

Looking at the histogram of outcomes of these functions in Figure 2.10, we can see more concentration of outcomes as the number of years increases.

Figure 2.10: Functions of History of Earnings Based on Highest Earnings



For the generations the full history of earnings are not observable, the highest earning years that we observe are not necessarily the highest earning years over all of the worker's life cycle.

Utilizing Latest Earnings Years

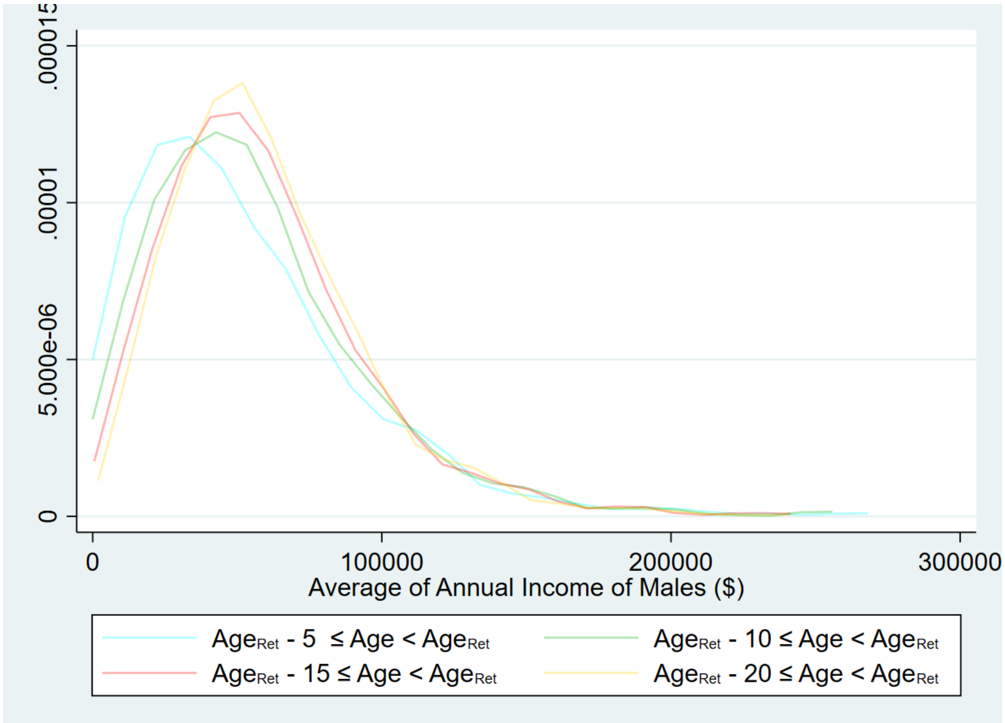
In Table 2.3 and Figure 2.11, I show the results for functions that consider the latest years of earnings using the first assumption. Similar to the highest earning method, the standard error of the outcomes decreases by increasing the number of years. Note as shown in Figure 2.4, the mean of earnings decreases late in the life cycle. Hence, the mean of average earnings will be higher as more years are taken into account. Note that I ignore the fact that the retirement age is endogenous and count the number of years backward from the observed age of retirement of the individual.

Table 2.3: Functions of History of Earnings Based on Latest Earnings from Actual Retirement Age

Mean of Earning	Mean	Std. dev.
Last 5 Years	52165.09	39807.09
Last 10 Years	56467.02	37854.27
Last 15 Years	59035.93	35301.48
Last 20 Years	60548.72	33644.44
All	61506.78	30177.98

Notes: The number of observations are 814.

Figure 2.11: Functions of History of Earnings Based on Latest Earnings from Actual Retirement Age



Inequality Among Different Groups with Different Methods and Rules

In this section, I discuss how the history-dependent part of the pension system shapes the inequality of pension benefits among different groups of workers. Table 2.4 illustrates the results of previous sections conditional on the education level of the workers. When the highest earning years are counted for the pension benefit, the more years counted, the less inequality between these two groups. This result can be explained by the fact that the difference between these two groups is not large when we consider the low-earning years of workers. However, the top-earning years of college-educated workers are much more than the top-earning years of non-college-educated workers. Similar results occur when we consider the latest years of earnings. As the number of years counted increases, the difference between the average earnings of no-college-educated workers and college-educated workers rises.

Table 2.4: Functions of History of Earnings Conditional on Education

	Without College Degree			With College Degree		
Mean of Earnings	Obs	Mean	Std. dev.	Obs	Mean	Std. dev.
Highest 5 Years		75895.08	33555.74		116374.6	47311.66
Highest 10 Years		68600.01	30392.11		106466.7	42590.16
Highest 15 Years		63215.37	28582.25		99380.73	40422.13
Highest 20 Years		58984.64	27461.92		93878.04	38581.33
Highest 25 Years		55630.34	26752.69		88935.12	37587.92
Highest 30 Year		53148.06	26284.26		84532.30	36969.00
Highest 35 Years		51604.20	25810.90		81446.47	36654.75
Latest 5 Years		44984	31628.65		69496.68	50776.2
Latest 10 Years		48468.04	29317.51		75564.58	47871.59
Latest 15 Years		50648.44	27480.97		79096.01	43075.57
Latest 20 Years		52031.33	26226.54		80919.48	40131.22
All	574	51169.77	25656.88	240	80494.08	36652.48

Table 2.5 presents the results based on the level of fluctuations in the earnings of workers. I divide the workers into two groups of high and low earning variability based on the median of the standard deviation of workers' earnings history. When only the top earnings of workers are cream skimmed the difference between the two groups is high, but as more years are considered the difference becomes smaller. This fact can be used to redistribute between these groups. The same pattern, to a lesser extent, applies to the latest years method.

Table 2.5: Functions of History of Earnings Conditional on Standard Error of Wages

	Below Median			Above Median		
Mean of Earnings	Obs	Mean	Std. dev.	Obs	Mean	Std. dev.
Highest 5 Years		57663.07	17601.06		117997.1	38150.68
Highest 10 Years		52130.44	17278.59		107398.8	33756.71
Highest 15 Years		47945.69	17492.57		99811.06	31713.2
Highest 20 Years		44763.66	17266.75		93781.58	30751.58
Highest 25 Years		41967.89	17073.76		88931.97	30118.52
Highest 30 Year		39800.83	16613.16		85001.96	29834.01
Highest 35 Years		38315.69	15804.46		82490.12	29571.96
Latest 5 Years		34505.15	19264.76		69956.5	46677.38
Latest 10 Years		37002.18	17860.3		75884.03	42318.53
Latest 15 Years		38425.02	16701.92		79646.85	36943.08
Latest 20 Years		39334.42	15605.36		81763.01	33479.08
All	407	37775.25	15414.81	407	81856.27	29440.41

Table 2.6 shows the results conditional on the level of workforce attachment of the workers. First, I calculate the share of the working history of each individual which they have zero earnings. Then, I divide them into two groups based on their position relative to the median of these shares. In the highest earning years method, the difference between these two groups becomes smaller, in general, as the number of years increases. As workforce attachment is highly associated with the level of earnings, the workers with lower workforce attachment do not benefit from counting less number of years. In the latest years method, the difference seems to not be affected greatly by the number of years. Note that the workforce attachment groups are categorized by working history during the whole life cycle of workers, which does not completely align with workforce attachment in the last 5 to 20 years of working life.

Table 2.6: Functions of History of Earnings Conditional on Share of Zero Work Hour Years

	Below Median			Above Median		
Mean of Earnings	Obs	Mean	Std. dev.	Obs	Mean	Std. dev.
Highest 5 Years		82300.88	41186.41		93386.53	42805.02
Highest 10 Years		74131.91	37435.28		85425.08	38782.09
Highest 15 Years		67983.74	35945.82		79802.04	36026.98
Highest 20 Years		63436.93	35083.38		75137.07	33889.18
Highest 25 Years		59935.63	34754.73		70991.4	32164.94
Highest 30 Year		58192.14	34599.45		66631.39	30931.23
Highest 35 Years		57594.57	34517.37		63225.08	29904.92
Latest 5 Years		52898.5	40443.33		51437.11	39201.66
Latest 10 Years		56067.85	39178.04		56867.17	36523.12
Latest 15 Years		57910.67	36541.25		60166.73	34017.72
Latest 20 Years		59251.38	34715.22		61852.44	32523.6
All	408	57594.57	34517.37	406	62047.9	29599.82

2.5 Model

2.5.1 Setup

The model starts at period 1, with a continuum of males (m) at age 26. Time is discrete and each time period (t) represents one year. Individuals are indexed by i which is excluded whenever it is not necessary for brevity. Each individual is either college-educated or not $(edu^i \in \{1, 0\})$. At each time period, individuals decide how much to consume versus save in a risk-free asset. During their working life, individuals also decide how many hours to work. Working individuals pay labor income and social security taxes. Each year after the early retirement age (age 62) until the full retirement age (age 66), they can decide to retire and receive their pension benefit (which depends on past incomes and retirement age and is determined according to pension rules) for the following years. Death is deterministic and happens at $T_L = 81$.

2.5.2 Preferences

The per-period utility of each individual is

$$u_t(c_t, h_t) = \frac{c_t^{1-\gamma}}{1-\gamma} - \psi \frac{{h_t}^{1+\frac{1}{\eta}}}{1+\frac{I}{\eta}} - \phi \mathbb{I}\{h_t > 0\} \tag{2.3}$$

The first term in $u_t(.)$ shows a person's utility from consumption (c_t). σ stands for the coefficient of relative risk aversion of consumption. The second term shows a person's disutility from work hours (h_t) in the intensive margin. ψ represents the disutility of work hours (h_t) and η is the Frisch elasticity of labor supply. The third term shows the fixed utility cost of working as determined by the coefficient ϕ.

2.5.3 Budget Constraint

The budget constraint before retirement age $(t < t_R)$ is

$$c_t = y_t - T^{inc}(y_t) - T^{SS}(y_t) + a_t - \frac{1}{1+r} a_{t+1},\ y_t = w_t h_t, a_t > 0 \tag{2.4}$$

Where w_t, y_t, and a_t represent wage, income, and asset, respectively. $T^{inc}(y_t)$ and $T^{SS}(y_t)$ are income and social security taxes. For each individual i at period t, her wage follows

$$\log(w_t) = \mu(edu, t) + \theta_t,\ \theta_t = \theta_{t-1} + \epsilon_t \tag{2.5}$$

$$\epsilon_t \sim N(0, {\sigma_\epsilon}^2),\ \theta_0 = 0. \tag{2.6}$$

Where $\mu(edu, t)$ is the deterministic part of wage and has a quadratic form, and θ_t is the stochastic part of wage which follows an AR(1) process with normal shocks ϵ_t.

After an individual decided to retire, her budget constraint follows

$$c_t = b(t_R, \{y_s\}_{s=1}^{t_R}) + a_t - \frac{1}{1+r} a_{t+1}. \tag{2.7}$$

where $b(.)$ is the benefit function that determines the amount of annual pension payment according to social security pension rules based on the retirement age and past incomes.

2.5.4 Individual Problem

The value function of non-retired individuals is

$$V_t^W(\{y_s\}_{s=1}^{t-1}, a_t, \theta_t, edu) = \max_{c_t,h_t} u_t(c_t, h_t) + \beta\mathbb{E}\Big[V_{t+1}^W(\{y_s\}_{s=1}^{t}, a_{t+1}, \theta_{t+1}, edu)\Big] \tag{2.8}$$

$$a_{t+1} = (1+r)\Big(a_t + w_t h_t - T^{inc}(y_t) - T^{SS}(y_t) - c_t\Big),\ \log(w_t) = \mu(edu, t) + \theta_t \tag{2.9}$$

$$\theta_{t+1} = \theta_t + \epsilon_{t+1} \tag{2.10}$$

$$\{y_s\}_{s=1}^{t+1} = \Big(\{y_s\}_{s=1}^{t}, y_{t+1}\Big),\ y_{t+1} = w_{t+1}h_{t+1} \tag{2.11}$$

Where $\{y_s\}_{s=1}^{t}$ is vector history of earnings. However, given that in the current social security regime and the counterfactuals studied in this paper the pension benefit is not a function of full earning history, it is possible to calculate the pension benefit with a lower number of the state variable. For instance, in the current rules and model environment, knowing AIME up to each year and the 5 lowest earrings so far are enough to keep track of the pension benefit after retirement. In this case, the problem simplifies to:

$$V_t^W(AIME_t, \{y_t^{min,k}\}_{k=1}^5, a_t, \theta_t, edu) = \max_{c_t,h_t} u_t(c_t, h_t) \tag{2.12}$$

$$+ \beta\mathbb{E}\Big[V_{t+1}^W(AIME_{t+1}, \{y_{t+1}^{min,k}\}_{k=1}^5, a_{t+1}, \theta_{t+1}, edu)\Big] \tag{2.13}$$

$$a_{t+1} = (1+r)\Big(a_t + w_t h_t - T^{inc}(y_t) - T^{SS}(y_t) - c_t\Big),\ \log(w_t) = \mu(edu, t) + \theta_t \tag{2.14}$$

$$\theta_{t+1} = \theta_t + \epsilon_{t+1} \tag{2.15}$$

$$AIME_{t+1} = \begin{cases} AIME_t + \frac{1}{T_{AIME}} y_t & \textit{if } y_t > y_t^{min,5}, \\ AIME_t + \frac{1}{T_{AIME}} y_t^{min,5} & \textit{if } y_t \leq y_t^{min,5}. \end{cases} \tag{2.16}$$

(2.17)

$$\{y_{t+1}^{min,k}\}_{k=1}^{5} = \begin{cases} y_{t+1}^{min,1} = y_t \quad , y_{t+1}^{min,2} = y_t^{min,1}, y_{t+1}^{min,3} = y_t^{min,2}, y_{t+1}^{min,4} = y_{t+1}^{min,3}, y_{t+1}^{min,5} = y_t^{min,4} \\ \textit{if} \quad\quad y_t < y_t^{min,1} & , \\ y_{t+1}^{min,1} = y_t^{min,1}, y_{t+1}^{min,2} = y_t \quad , y_{t+1}^{min,3} = y_t^{min,2}, y_{t+1}^{min,4} = y_{t+1}^{min,3}, y_{t+1}^{min,5} = y_t^{min,4} \\ \textit{if } y_t^{min,1} \leq y_t < y_t^{min,2} & , \\ y_{t+1}^{min,1} = y_t^{min,1}, y_{t+1}^{min,2} = y_t^{min,2}, y_{t+1}^{min,3} = y_t \quad , y_{t+1}^{min,4} = y_{t+1}^{min,3}, y_{t+1}^{min,5} = y_t^{min,4} \\ \textit{if } y_t^{min,2} \leq y_t < y_t^{min,3} & , \\ y_{t+1}^{min,1} = y_t^{min,1}, y_{t+1}^{min,2} = y_t^{min,2}, y_{t+1}^{min,3} = y_t^{min,3}, y_{t+1}^{min,4} = y_t \quad , y_{t+1}^{min,5} = y_t^{min,4} \\ \textit{if } y_t^{min,3} \leq y_t < y_t^{min,4} & , \\ y_{t+1}^{min,1} = y_t^{min,1}, y_{t+1}^{min,2} = y_t^{min,2}, y_{t+1}^{min,3} = y_t^{min,3}, y_{t+1}^{min,4} = y_t^{min,4}, y_{t+1}^{min,5} = y_t \\ \textit{if } y_t^{min,4} \leq y_t < y_t^{min,5} & , \\ y_{t+1}^{min,1} = y_t^{min,1}, y_{t+1}^{min,2} = y_t^{min,2}, y_{t+1}^{min,3} = y_t^{min,3}, y_{t+1}^{min,4} = y_t^{min,4}, y_{t+1}^{min,5} = y_t^{min,5} \\ \textit{if } y_t^{min,5} \leq y_t & , \end{cases} \tag{2.18}$$

where $y_t^{min,k}$ is the k^{th} lowest past earning and T_{AIME} is the number of years utilized to calculate AIME.

The value function of an individual after being retired is

$$V_t^R(AIME_{t_R}, a_t, t_R) = \max_{c_t} u_t(c_t, 0) + \beta V_t^R(AIME_{t_R}, a_{t+1}, t_R) \tag{2.19}$$

$$a_{t+1} = (1+r)\Big(a_t + PIA(AIME_{t_R}, t_R) - c_t\Big). \tag{2.20}$$

Where $AIME_{t_R}$ is the AIME at the time of retirement and $PIA(.)$ is the function that calculate the pension benefit based on AIME. Between the early retirement age and full retirement age, the value function follows

$$V_t^E(AIME_{t_R}, a_t, t_R) = \max_{Ret\in\{0,1\}} \Big\{V_t^W(.), V_t^R(.)\Big\}, \tag{2.21}$$

as the individuals decides whether to retire *Ret* $= 1$ or not *Ret* $= 0$. The age when *Ret* $= 1$ is the retirement age (t_R).

2.6 Quantitative Environment

2.6.1 Preference Parameters

The preference parameters taken from the literature are $\beta = 0.98803$ and $\gamma = 1.66$ (Huggett and Parra (2010)), $\eta = 0.5$ as it is common in the literature (i.e. **?**). I estimate $\psi = 0.04$ and $\phi = 0.0006$ to match mean working hours of men and labor supply participation of men above age 50 in the sample.

2.6.2 Budget Constraint Parameters

Total endowment hours in each year is assumed to be 8760. Furthermore, I am assuming $r = 0.042$ similar to Huggett and Parra (2010).

2.6.3 Wage Process

I estimate the wage function via GMM method. The wage function is assumed to take a quadratic form

$$mu(Age, edu) = \beta_0^w + \beta_1^w Age + \beta_2^w Age^2 + \beta_3^w edu + \beta_4^w edu \cdot Age + \beta_5^w edu \cdot Age^2. \tag{2.22}$$

The coefficients of the wage function $(\beta_0^w, \beta_1^w, \beta_2^w, \beta_3^w, \beta_4^w, \beta_5^w)$ are $(1.6698, 0.0605, -.0006, -.3780, 0.03214, -0.0002)$ The difference between the observed and model-generated wages is assumed to measurement error. The variance of persistent shocks is estimated to $\sigma_\epsilon^2 = 0.02601$.

2.6.4 Tax and Social Security Rules

Similar to Heathcote et al. (2017), I assume the Income Tax takes the form of $T^{SS}(y_t) = y_t - \kappa {y_t}^{(1-\tau)}$ and estimate its parameters using TAXIM data: $\kappa = 2.716084$ and $\tau_T = 0.1029$. The social Security rules is assumed to be same as 2000 rules for all years. the socail security tax is $T^{SS}(y_t) = \tau_{SS} y_t = 0.106$ until the wage base = 76200. The PIA (primary insurance amount) has three bend points $a = \{6372, 38422, 76200\}$ and slopes are $b = \{0.9, 0.32, 0.15\}$ with the wage base of $= 76200$. The early retirement reeducation's are $re = \{0.75, 0.80, 0.867, 0.933\}$. AIME (Average Index monthly earning) is $\frac{1}{12}$ of mean of top 35 years of earnings.

2.7 Numerical Exercise

In this section, I utilize the presented quantitative model and evaluate the effects of a counterfactual policy that changes the current US pension policy which uses the top 35 years of earnings to account for the lifetime earnings, common in other OECD countries. Table 2.7 presents the result both separately for non-college-educated and college-educated workers and when they are pooled together. Accounting for the full history of earnings causes 42 % increase in consumption and 82 % increase in hours of work. Age of retirement decreases by 5.7 % as the age of retirement in the new regime will reach the lower constraints allowed in the model.

Table 2.7: Results of the Numerical Exercise

	Mean of Status Quo			Mean of Counterfactual			% change in Mean		
edu	0	1	all	0	1	all	0	1	all
c	30871	22884	29352	44167	47405	44783	43.06	107.15	52.56
h	1611	1236	1539	2872	2549	2811	78.30	106.16	82.56
a	223861	186657	216786	223126	308285	239321	-0.32	65.16	10.39
PIA	12203	10389	11858	11745	11627	11722	-3.75	11.92	-1.14
Age_R	65.38	65.95	65.49	62	62	62	-5.17	-5.99	-5.33

Notes: Consumption (c), assert (a), and PIA are in the unit of $, work hours (h) is in the unit of hours, and age of retirement (Age_R) is in the unit of years.

Moving to the pension system that considers the full earning history increases the marginal benefit of working in the years that are not among the top 35 years of earnings as now those years would affect the pension benefits. On the other hand, since the weight of each of the top 35 earning years is now lower, the marginal benefit of working during those years decreases. If the years between the early and normal retirement ages are among the top 35 years of earning (which is the case for most of the workers as it was shown in Figure 2.4) there is more incentive to retire earlier. If workers delay retirement, their pension benefits will become lower by the addition of the low-wage years of their 60s to their working history which lowers their average full-life earnings.

Note that due to the high number of continuous state variables, approximating the value function of working individuals is a major challenge to numerically solve this problem. Here, I utilize a quadratic polynomial approximation method to overcome this problem. However, this approximation method is only used for the current policy regime and not for the counterfactual one. Given the lower number of state variables needed to calculate pension benefits in the full history method, there is no longer a need for polynomial approximation of value functions. The use of different approximation methods can potentially distort the comparison of the results in these two cases. Moreover, the model parameters are calibrated for the status quo environment. Hence, these results should be seen as a numerical exercise and are not directly applicable to policymaking. Improving the approximation method is left for future research.

2.8 Conclusion

In most countries that use a defined-benefit system, two functions working in the composite determine the amount of pension benefits. One function summarizes all of the history of earnings into one outcome, and the other finds the amount of benefit based on that outcome. In the US, ignoring some details, the history of earnings will be summarized by taking the average of the top 35 years of earnings of a worker. Then, this result will yield the retirement benefit after passing through a progressive benefit function. History dependence of pension systems influences labor supply at old age and the retirement age, the level of redistribution among workers based on their full history of earnings, and consumption insurance of retirees. Studying the effects of the design of the history-dependent part of the pension system and ways to improve it (an understudied area of research) is the main goal of this paper. In this research, I introduce some new stylized facts and show how utilizing different methods to calculate pension benefits from the history of earnings would affect different workers. Next, I examine how different ways to summarize the history of earnings affect workers. I develop a dynamic model of labor supply, saving, and retirement with labor market shocks and evaluate a counterfactual policy that changes the current US pension policy which uses the top 35 years of earnings to account for lifetime earnings.

Chapter 2 Appendices

2.A Numerical Algorithm

2.A.1 Grids

The grid for work hour consist of 5 points at equal distance between 0 and 3500 hours a year. Minim or maximum past incomes are stored in 5 point grid between 0 $ and 200k $ with 2/3 of points at the first 1/3 of the domain. The grid for AIME has 6 points at equal distance to each other from 0 $ to the wage base of social security rules. The shocks to permanent wages is approximated by 5 points according to Gauss-Hermite weights and locations. The permanent wage is stored in 8 points uniform grid. The girds for asset consist of 12 points from 0 to 750k $ with 2/3 of the grids in the first 1/5 of the domain.

2.A.2 Solution Algorithm

I use backward induction to numerically solve for the value and policy functions. Here I explain each step of the Algorithm at each t used to solve the dynamic programming problem. **Inputs:** S_t grid points, h_t grid, a^{t+1} grid, and V^*

1. Regress V^* on vector on quadratic polynomial of state variables and save b_t^V
2. Based on law of motions of S_t, find S_{t+1} for each S grid point and each choices of h_t,a_{t+1},Ret_t
3. Calculate $V_t(S_t, h_t, a_{t+1}, Ret_t) = u(.) + \mathbb{E}[V^{t+1}(S')]$ based on b_V and analytical integral
4. find $h_t^*(S)$,$a_{t+1}^*(S)$,$Ret_t^*(S)$ that maximizes $V(.)$.
5. Regress $h_t^*(S)$,$a_{t+1}^*(S)$,$Ret_t^*(S)$ on quadratic polynomial of state variables and save coefficient for simulation.

Bibliography

Aaberge, R., U. Colombino, and S. Ström (1999). Labour supply in italy : An empirical analysis of joint household decisions , with taxes and quantity constraints. *Journal of applied econometrics 14*, 403–422.

Abbott, B. (2017). The substitutability of parental investments of time and expenditure. *Working Paper*.

Adda, J., C. Dustmann, and K. Stevens (2017). The career costs of children. *Journal of Political Economy 125*, 293–337.

Agostinelli, F. and M. Wiswall (2016). Estimating the technology of children's skill formation. *Working Paper*.

Aizawa, N. (2017). Labor market sorting and health insurance. *Working Paper*.

Aizawa, N., S. Kim, and S. Rhee (2019). Labor market screening and social insurance program design for the disabled. *Working Paper*.

Akerlof, G. A. (1978). The economics of " tagging " as applied to the optimal income tax, welfare programs, and manpower planning. *American Economic Review 68*, 8–19.

Almond, D. and J. Currie (2010). Human capital development before age 5. *NBER WORKING PAPER SERIES*.

Almond, D., J. Currie, and V. Duque (2018). Childhood circumstances and adult outcomes : Act ii †. *56*, 1360–1446.

Altonji, J., A. A. Smith, and I. Vidangos (2013). Modeling earnings dynamics. *Econometrica 81*, 1395–1454.

Attanasio, O., P. Levell, H. Low, and V. Sánchez-Marcos (2018). Aggregating elasticities: Intensive and extensive margins of women's labor supply. *Econometrica 86*, 2049–2082.

Attanasio, O. P., J. Banks, C. Meghir, and G. Weber (1999). Humps and bumps in lifetime consumption. *Journal of Business and Economic Statistics 17*, 22–35.

Banks, B. J., R. Blundell, and S. Tanner (1998). Is there a retirement-savings puzzle ? *American Economic Review 88*, 769–788.

Banks, J., R. Blundell, and I. A. N. Preston (1992). Adult equivalence scales: A life-cycle perspective. *Fiscal Policy*.

Batzer, R. (2021). Letting your past define your taxes: Optimal history-dependent income taxation in general equilibrium. *Working Paper*, 1–62.

Bernal, R. (2008). The effect of maternal employment and child care on children's cognitive development. *International Economics Review, forthcoming 49*, 1173–1209.

Bick, A. and S. Choi (2013). Revisiting the effect of household size on consumption over the life-cycle. *Journal of Economic Dynamics and Control 37*, 2998–3011.

Blumkin, T., Y. Margalioth, and E. Sadka (2010). Taxing children: The re-distributive role of child benefits - revisited. *CESifo Working Paper Series*.

Blundell, R., M. Browning, and C. Meghir (1994). Consumer demand and the life-cycle allocation of household expenditures. *Review of Economic Studies 61*, 57–80.

Blundell, R., P.-A. Chiappori, and C. Meghir (2005). Collective labor supply with children. *Journal of Political Economy 113*.

Blundell, R., M. C. Dias, C. Meghir, and J. Shaw (2016). Female labor supply, human capital, and welfare reform. *Econometrica 84*, 1705–1753.

Blundell, R. and A. Lewbel (1991). The information content of equivalence scales. *Journal of Econometrics 50*, 49–68.

Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2016). Consumption inequality and family labor supply. *American Economic Review 106*, 387–435.

Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2018). Children, time allocation, and consumption insurance. *Journal of Political Economy 126*, S73–S115.

Blundell, R. and A. Shephard (2011). Employment , hours of work and the optimal taxation of low income families. *Review of Economic Studies*, 1–31.

Boca, D. D., C. Flinn, and M. Wiswall (2014). Household choices and child development. *Review of Economic Studies 81*, 137–185.

Boca, D. D., C. Flinn, and M. Wiswall (2016). Transfers to households with children and child development. *Economic Journal 126*, F136–F183.

Borella, M., M. D. Nardi, and F. Yang (2023). Are marriage-related taxes and social security benefits holding back female labor supply. *The Review of Economic Studies 90*, 102–131.

Cagetti, M. (2003). Wealth accumulation over the life cycle and precautionary savings. *Journal of Business and Economic Statistics 0*.

Carneiro, P. and R. Ginja (2016). Partial insurance and investments in children. *Economic Journal 126*, F66–F95.

Caucutt, E. M. and L. Lochner (2012). Early and late human capital investments, borrowing constraints, and the family. *NBER Working Paper Series*.

Cigno, A., A. Luporini, and A. Pettini (2003). Transfers to families with children as a principal-agent problem. *Journal of Public Economics 87*, 1165–1177.

Cigno, A. and A. Pettini (2002). Taxing family size and subsidizing child-specific commodities? *Journal of Public Economics 84*, 75–90.

Conesa, C., S. Kitao, and D. Krueger (2009). Taxing capital ? not a bad idea after all! *American Economic Review 99*, 25–48.

Corinth, K., B. D. Meyer, M. Stadnicki, and D. Wu (2021). The anti-poverty, targeting, and labor supply effects of the proposed child tax credit expansion. *NBER Workign Paper Series*.

Cremer, H., A. Dellis, and P. Pestieau (2003). Family size and optimal income taxation. *The Economic Journal 16*, 37–54.

Cremer, H., F. Gahvari, and J. M. Lozachmeur (2010). Tagging and income taxation: Theory and an application. *American Economic Journal: Economic Policy 2*, 31–50.

Cuhna, F., J. Heckman, and S. Schennch (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*, 883–931.

Cunha, F. (2013). Investments in children when markets are incomplete. *Working Paper*.

Cunha, F. and J. J. Heckman (2009). The economics and psychology of inequality and human development. *Journal of the European Economic Association 7*, 320–364.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006). Chapter 12: Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education 1*, 697–812.

Cunha, F., J. J. Heckman, and S. Schennch (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*, 883–931.

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

Diamond, P. (2004). Social security. *American Economic Review: Papers Proceedings*, 1–24.

Diamond, P. (2009). Taxes and pensions. *Southern Economic Journal 76*, 2–15.

Diamond, P. A. . and J. A. . Mirrlees (1986). Payroll-tax financed social insurance with variable retirement. *The Scandinavian Journal of Economics 88*, 25–50.

Diamond, P. A. and J. A. Mirrlees (1978). A model of social insurance with variable retirement. *Journal of Public Economics 10*, 295–336.

Erosa, A. and M. Gervais (2002). Optimal taxation in life-cycle economies. *Journal of Economic Theory 105*, 338–369.

Fan, X., A. Seshadri, and C. Taber (2022). Estimation of a life-cycle model with human capital, labor supply and retirement. *NBER Working Papers 29905*.

Farhi, E. and I. Werning (2007). Inequality and social discounting. *Journal of Political Economy 115*, 365–402.

Farhi, E. and I. Werning (2013). Insurance and taxation over the life cycle. *Review of Economic Studies 80*, 596–635.

French, E. (2005). The effects of health, wealth, and wages on labour supply and retirement behaviour. *The Review of Economic Studies 72*, 395–427.

Gayle, G.-L. and A. Shephard (2019). Optimal taxation, marriage, home production, and family labor supply. *Econometrica 87*, 291–326.

Gelber, A. and M. Weinzierl (2016). Optimal taxation when children's abilities depend on parents' resources. *69*, 11–40.

Golosov, M., A. Shourideh, M. Troshkin, and A. Tsyvinski (2013). Optimal pension systems with simple instruments. *American Economic Review: Papers Proceedings 103*, 502–507.

Golosov, M., M. Troshkin, and A. Tsyvinski (2016). Redistribution and social insurance. *American Economic Review 106*, 359–386.

Grochulski, B. and N. Kocherlakota (2010). Nonseparable preferences and optimal social security. *Journal of Economic Theory 145*, 2055–2077.

Guner, N., R. Kaygusuz, and G. Ventura (2012). Taxation and household labour supply. *Review of Economic Studies 79*, 1113–1149.

Guner, N., R. Kaygusuz, and G. Ventura (2020). Child-related transfers, household labour supply, and welfare. *Review of Economic Studies 87*, 2290–2321.

Heathcote, J., K. Storesletten, and G. L. Violante (2017). Optimal tax progressivity: An analytical framework. *Quarterly Journal of Economics 132*, 1693–1754.

Heathcote, J., K. Storesletten, and G. L. Violante (2020). Optimal progressivity with age-dependent taxation. *Journal of Public Economics 189*, 104074.

Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica 42*, 679–694.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*, 411–482.

Ho, C. and N. Pavoni (2020). Efficient child care subsidies. *American Economic Review 110*, 162–199.

Hosseini, R., L. E. Jones, and A. Shourideh (2013). Optimal contracting with dynastic altruism: Family size and per capita consumption. *Journal of Economic Theory 148*, 1806–1840.

Hoynes, H. W. and D. W. Schanzenbach (2009). Consumption responses to in-kind transfers: Evidence from the introduction of the food stamp program. *American Economic Journal: Applied Economics 1*, 109–39.

Huggett, M. and J. Parra (2010). How well does the u . s . social insurance system provide social insurance ? *Journal of Political Economy 118*, 76–112.

Imai, S. and M. P. Keane (2004). Intertemporal labor supply and human capital accumulation. *International Economic Review 45*, 601–641.

Kapička, M. (2022). Quantifying the welfare gains from history dependent income taxation. *Working Paper*.

Karabarbounis, M. (2016). A road map for efficiently taxing heterogeneous agents. *American Economic Journal: Macroeconomics 8*, 182–214.

Keane, M. P. (2011). Labor supply and taxes: A survey. *Journal of Economic Literature 49*, 961–1075.

Keane, M. P. and K. I. Wolpin (2001). The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review 42*, 1051–1103.

Kleven, H. J., C. T. Kreiner, and E. Saez (2009). The optimal income taxation of couples. *Econometrica 77*, 537–560.

Kremer, M. (2001). Should taxes be independent of age? *Working Paper*.

Kurnaz, M. (2021). Optimal taxation of families: Mirrlees meets becker. *Economica 13*, 957–971.

Kurnaz, M. and M. Soytas (2019). Intergenerational income mobility and income taxation. *SSRN Electronic Journal*, 1–46.

Lee, S. Y. T. and A. Seshadri (2019). On the intergenerational transmission of economic status. *Journal of Political Economy 127*, 855–921.

Low, H., C. Meghir, and L. Pistaferri (2010). Wage risk and employment over the life cycle. *American Economic Review 100*, 1432–1467.

Low, H., C. Meghir, L. Pistaferri, and A. Voena (2018). Marriage, labor supply and the social safety net. *Working Paper*.

Low, H. W. and L. Pistaferri (2015). Disability insurance and the dynamics of the incentive-insurance trade-off. *American Economic Review 105*, 2986–3029.

Malkov, E. (2021). Optimal income taxation of singles and couples. *Working Paper*, 1–59.

Mankiw, N. G. and M. Weinzierl (2009). The optimal taxation of height: A case study of utilitarian income redistribution. *NBER Workign Paper Series*, 1–28.

Meghir, C. and L. Pistaferri (2004). Income variance dynamics and heterogeneity. *Econometrica 72*, 1–32.

Moffitt, R. A., B. J. Phelan, and A. E. Winkler (2017). Welfare rules, incentives, and family structure. *NBER Working Papers 21257*, 1–56.

Mullins, J. (2019). Designing cash transfers in the presence of children's human capital formation. *Working Paper*.

Ndiaye, A. (2020). Flexible retirement and optimal taxation. *Working Paper*.

O'Dea, C. (2017). Insurance , efficiency and design of public pensions. *Working Paper*.

olivier Gourinchas, B. Y. P. and J. A. Parkeri (2002). Consumption over the life cycle. *Econometrica 70*, 47–89.

Parodi, F. (2019). Taxation of durables, non-durables and earnings with heterogeneous preferences. *SSRN Electronic Journal*, 1–69.

Pavoni, N. (2007). Optimal welfare-to-work programs. *Review of Economic Studies*, 283–318.

Petit, B. (2019). Intergenerational effects of child-related tax benefits in the us. *Working Paper*.

Piketty, T. and E. Saez (2013). Optimal labor income taxation. *Handbook of Public Economics 5*, 391–474.

Rupert, P. and G. Zanella (2015, 5). Revisiting wage, earnings, and hours profiles. *Journal of Monetary Economics 72*, 114–130.

Saez, E. (2001). Using elasticitles to derive optimal income tax rates. *The review of economic studies 68*, 205–229.

Saez, E. (2002). Optimal income transfer programs: Intensive versus extensive labor supply responses. *Quarterly Journal of Economics 117*, 1039–1073.

Saez, E. and S. Stantcheva (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review 106*, 24–45.

Scholz, J. K. and A. Seshadri (2009). Children and household wealth. *Working Paper*.

Shourideh, A. and M. Troshkin (2017). Incentives and efficiency of pension systems. *Working Paper*.

Stantcheva, S. (2013). Optimal taxation and human capital policies over the life cycle. *Journal of Political Economy 125*, 1–10.

Stantcheva, S. (2015). Optimal income, education, and bequest taxes in an intergenerational model. *NBER Working Paper*.

Stantcheva, S. (2020). Dynamic taxation. *Forthcoming Annual Review of Economics*.

Verriest, E. (2019). Household labor supply, child development and childcare policies in the united states. *SSRN Electronic Journal*.

Vickrey, W. (1947). *Agenda for Progressive Taxation*.

Weinzierl, M. (2011). The surprising power of age-dependent taxes. *Review of Economic Studies 78*, 1490–1518.

Weinzierl, M. C. (2012). Why do we redistribute so much but tag so little? the principle of equal sacrifice and optimal taxation. *NBER Workign Paper Series*.

Wilde, P. (2001). Understanding the food stamp benefit formula. *Economic Research Service/USDA*.

Wu, C. and D. Krueger (2021). Consumption insurance against wage risk: Family labor supply and optimal progressive income taxation. *American Economic Journal: Macroeconomics 13*, 79–113.

Zhou, A. (2021). Building future generations: The macroeconomic consequences of family policies. *Working Paper*, 1–59.