Estimation of Dynastic Life-Cycle Discrete Choice Models

George-Levi Gayle
Department of Economics, Washington University in St. Louis

Limor Golan
Department of Economics. Washington University in St. Louis

Mehmet A. Soytas
Graduate School of Business, Ozyegin University

October 21, 2014

Abstract

Dynastic models have long provided a framework for the study of equilibria with intergenerational transfers, social mobility, and inequality. However with a few exceptions, there has been very little work on the estimations of these models. With the advent of data sets like the PSID that tracks households over more than one generation estimation of these models are now feasible. This paper explores the estimation of a class of life-cycle discrete choice intergenerational models. It proposes a new semiparametric estimation technique that circumvents the need for full solution of the dynamic programming problem. As is standard in this class of estimators, we show that it is $\sqrt{N}$ consistent and asymptotically normally distributed. We compare our estimator to a modified version of the full solution maximum likelihood estimator in a Monte Carlo study. Our estimator performs comparable to ML in finite sample but greatly reduces the computational cost. To demonstrate the applicability of the estimator, a dynastic model of intergenerational transmission of human capital with unitary households is estimated. (Preliminary and Incomplete)

Keywords: Intergenerational Models, Estimation, Discrete Choice, Human Capital, PSID
JEL classification: C13, J13, J22, J62

1 Introduction

The importance of parents' altruism toward their children and children's altruism towards their parents has long been recognized as an important factor underlying economic behavior of individuals. Economic models that incorporate these intergenerational links are normally referred to as dynastic models. Many important economic behavior and hence the welfare effect of many public policies critically depend on whether these intergenerational links are explicitly modeled. For example, several papers have documented that the distribution of wealth is more concentrated than that of the labor earnings and that it is characterized by a smaller of fraction of households owning a larger fraction of total wealth over time. There are different models of intergenerational transfers explaining persistence in wealth and income across generations (for example, Loury 1981 model of transmission of human capital and Laitner 1992 model of bequests) however, in these models fertility is exogenous. Becker and Barro (1988) Barro and Becker (1989) develop dynastic models with endogenous fertility, however, in their model endogenizing fertility leads to lack of persistence in earnings and wealth because wealthier households have more children and therefore intergenerational transfers do not depend on wealth and income. In the data, there is clearly persistence in income across generations. Since then dynastic models with endogenous fertility that capture the intergenrational persistence of income and wealth have been analyzed extensively, but such models have not been estimated mainly because of computational feasibility considerations. This paper develops an estimator for dynastic models of intergenerational transfers and estimates a model quantifying the different factors generating persistence of income.

While the study of dynastic models have been wide spread in the economic literature these studies have been largely theoretical or quantitative theory. However, the estimation of these models and the use of these estimated models to conduct counterfactual policy analysis are nonexistent. There are two main reasons for this gap, the first is the data limitation and the second is computational feasibility. Ideally, one would need data on the choices and characteristics of multiple generations linked across time in order to estimate these dynastic models. The number of generations needed for estimation can be reduced to two by analyzing the stationary equilibrium properties of these models and recently data on the choices and characteristics of at least two generations have become available in the Panel Survey of Income Dynamics (PSID) and other data sets. The standard estimation algorithm for these types of model uses a nested fixed point algorithm. The major limitation of this estimation procedure is that it suffers from the curse of dimensionality, i.e. as the number of states in the state space increases the number of computations increases at a rate faster than linear. Dynastic models add an additional loop to this estimation procedure, i.e. a nested fixed point squared. Therefore, this estimation procedure suffers from the curse of dimensionality squared.

This paper develops an estimation that partially overcomes this curse of dimensional by exploiting properties of the stationary equilibrium. It provides a framework to estimate a rich class of dynastic models which includes investment in children's human capital, monetary transfers, unitary households, endogenous fertility and a life-cycle within each generation. This is an extension of methods used in the literature for the estimation of standard single agent model to the dynastic setting (see for example Hotz and Miller (1993) and Hotz et. al. (1994)). This estimation technique makes this estimation and empirical assessment of proposed counterfactual policy reform feasible. The paper compares the performance of the proposed estimator to a nested fixed point estimator using simulations and provides estimation results from an application of intergenerational transmission of human capital.

There are several types of dynastic models that our framework incorporates. In some models fertility is endogenous, as in Barro and Becker (1988, 1989), but cannot generates persistence of wealth across generations. Other models capture intergenerational transfers and persistence in wealth across generations but fertility is exogenous as in Laitner (1992) and Loury (1981). Alvarez (1999) combines the main features of the above mentioned models together by incorporating fertility decision into the Laitner (1981) and Loury (1992) intergenerational transfer models. On the other hand, some models, as Laitner (1981), incorporate an elaborate finite life cycle model for adults in each generation, while in other models there is one period of childhood and one period of adulthood. The framework we study incorporates all these elements and develops a model in which altruistic parents make discrete choices of birth, labor supply, and discrete and continuous investment in children. In particular, to accommodate many models in the literature, parents choose time with children and a continuous monetary investment in the child every year over their life-cycle. The model can also be extended to include bequests. The model is a partial equilibrium model, and as in most dynastic models and in the basic setup there is one decision maker in a household, however, we show that it can be easily extended to a unitary household. The empirical application presents estimates of a dynastic model of intergenerational transmission of human capital with unitary households.

There are two main estimators used in the literature to estimate dynamic discrete choice models: A full solution method using "nested fixed point" (NFXP) (see Wolpin (1984), Miller (1984), Pakes (1986) and Rust (1987) early examples) and "conditional choice probability" (CCP) (see Hotz and Miller (1993), Altug and Miller (1998), Aguirregabiria (1999)) estimators that do not require the solution to the fixed points. More recently Aguirregabiria and Mira (2002) showed that an appropriately formed CCP based estimator, "nested pseudo likelihood" (NPL), is asymptotically equivalent to a NFXP estimator. However even with a CCP estimator or a NPL estimator, estimation of the intergenerational model requires dealing with further complications which are not present in single agent dynamic discrete choice models. The first difficulty is the representation of the value functions of the problem. This difficulty is associated with the non-standard nature of the problem. An intergenerational model has finite (T) periods in the life-cycle in each generation and infinitely many generations are linked by the altruistic preferences. The problem in this framework does not fit into a finite horizon dynamic discrete choice model since in the last period T, there is a continuation value associated with the next generation's problem which is linked to current generation by the transfers and the discount factor. Therefore we need to find a representation for the next generation's continuation value if we want to treat the problem as a standard T period problem and solve by backwards induction[1]. In this paper we propose a new estimation procedure which enables us to derive representations of the period value functions in terms of period primitives. In particular, we show that an appropriately defined alternative representation of the continuation value enables us to derive the representations and one can apply a CCP estimator to the intergenerational model.

2 Theoretical Framework

The theoretical framework is developed to allow for estimation of a rich group of dynastic models and allows for addressing many relevant policy questions. This section develops a model of altruistic parents that make transfers to their children. The transfers are discrete, and they can allow for time investment in children (time is discritized) and modeling discrete labor supply choice allows for monetary investment with discrete levels. Section 4 then extend the basic framework to allow for continuous choice of transfers. This will allow to use the framework to analyze bequests or any continuous monetary transfers parents make to children. We incorporate two important aspects of the problem: first fertility is endogenous. Endogenous fertility has important implications to intergenerational transfers, and the quantity-quality trade-offs parents makes when they choose transfers as well as number of offsprings. Second, we include a life-cycle for each generations. Life-cycle is important to understanding fertility behavior, and spacing of children, as well as timing of different types of investments. This section analyzes a model with one genderless decision maker. We later extend this framework to a unitary household.[2]

We, build on previously developed dynastic models that analyze transfers and intergenerational transmission of human capital. In some models, such as Loury (1981) and Becker and Tomes (1986), fertility is exogenous while in others, such as Becker and Barro (1988) and Barro and Becker (1989), fertility is endogenous. The Barro-Becker framework is extended in our model by incorporating a life-cycle behavior model, based on previous work such as Heckman, Hotz and Walker (1985) and Hotz and Miller (1988) into an infinite horizon model of dynasties. Our life cycle model includes individuals choices about time allocation decisions, investments in children, and fertility. We formulate a partial equilibrium discrete choice model that incorporates life-cycle considerations of individuals from each generation into the larger framework. Adults in each generation derive utility from their own consumption, leisure, and from the utility of their adult offspring. The utility of adult offspring is determined probabilistic by the educational outcome of childhood, which in turn is determined by parental time and monetary inputs during early childhood, parental characteristics (such as education), and luck. Parents make decisions in each period about fertility, labor supply, time spent with children, and monetary transfers. For simplicity, the only intergenerational transfers are transfers of human capital, as in Loury (1981). However, the framework can include any other choice of transfer which is discrete. We assume no borrowing or savings for simplicity.

In the model adults live for T periods. Each adult from generation $g \in \{0, ...\infty\}$ makes discrete choices about labor supply, h_t, time spent with children, d_t, and birth, b_t, in every period $t = 1...T$. For labor time individuals choose no work, part time, or full time ($h_t \in (0, 1, 2)$ and for time spent with children individuals choose none, low and high ($d_t \in (0, 1, 2)$.. The birth decision is binary ($b_t \in (0, 1)$. The individual does not make any choices during childhood, when $t = 0$. All the discrete choices can be combined into one set of mutually exclusive discrete choice, represented as k, such that $k \in (0, 1...17)$. Let I_{kt} be an indicator for a particular choice k at age t; I_{kt}

[1] Obviously we can always solve the problem by NFXP; if we assume the next generation's period 0 value function is same as the current generation's value function in period 0. This is another way of saying the problem is stationary in the generations. In this case the solution to the dynamic programming problem requires solving the fixed point problem for the period value functions. However as one can easily anticipate we encounter the same computational burden of full solution. Therefore our specific interest is on CCP type estimators.

[2] Treatment of households with two decisions makers with separate utility functions, marriage and divorce is involved and is beyond the scope of this paper. See Gayle, Golan and Soytas (2014) for more details on one such model.

takes the value 1 if the k^{th} choice is chosen at age t and 0 otherwise. These indicators are defined as follows:

$$\begin{aligned} I_{0t} &= I\{h_t = 0\}I\{d_t = 0\}I\{b_t = 0\},\ I_{1t} = I\{h_t = 0\}I\{d_t = 0\}I\{b_t = 1\},\ ..., \\ I_{16t} &= I\{h_t = 1\}I\{d_t = 2\}I\{b_t = 1\},\ I_{17t} = I\{h_t = 2\}I\{d_t = 2\}I\{b_t = 1\} \end{aligned} \tag{1}$$

Since these indicators are mutually exclusive then $\sum_{k=0}^{17} I_{kt} = 1$. We define a vector, x, to include the time invariant characteristics of education, skill, and race of the individual. Incorporating this vector, we further define the a vector z to include all past discrete choices and as well as time invariant characteristics, such that $z_t = (\{I_{k1}\}_{k=0}^{17}, ..., \{I_{kt-1}\}_{k=0}^{17}, x)$.

We assume the utility function is the same for adults in all generations. An individual receives utility from discrete choice and from consumption of a composite good, c_t. The utility from consumption and leisure is assumed to be additively separable because the discrete choice, I_{kt}, is a proxy for the leisure, and is additively separable from consumption. The utility from I_{kt} is further decomposed in two additive components: a systematic component, denoted by $u_{1kt}(z_t)$, and an idiosyncratic component, denoted by ε_{kt}. The systematic component associated with each discrete choice k represents an individual's net instantaneous utility associated with the disutility from market work, the disutility/utility from parental time investment, and the disutility/utility from birth. The idiosyncratic component represents preference shock associated with each discrete choice k which is transitory in nature. To capture this feature of ε_{kt} we assume that the vector $(\varepsilon_{0t}, .., \varepsilon_{17t})$ is independent and identically distributed across the population and time, and is drawn from a population with a common distribution function, $F_\varepsilon(\varepsilon_{0t}, .., \varepsilon_{17t})$. The distribution function is assumed to be absolutely continuous with respect to the lebesgue measure and has a continuously differentiable density.

Per-period utility from the composite consumption good is denoted $u_{2t}(c_t, z_t)$. We assume that $u_{2t}(c_t, z_t)$ is concave in c, i.e. $\partial u_{2t}(c_t, z_t)/\partial c_t > 0$ and $\partial^2 u_{2t}(c_t, z_t)/\partial c_t^2 < 0$. Implicit in this specification is intertemporally separable utility from the consumption good, but not necessarily for the discrete choices, since u_{2t} is a function of z_t, which is itself a function of past discrete choices, but is not a function of the lagged values of c_t. Altruistic preferences are introduced under the same assumption as the Barro-Becker model: Parents obtain utility from their adult offsprings' expected lifetime utility. Two separable discount factors capture the altruistic component of the model. The first, β, is the standard rate of time preference parameter, and the second, $\lambda N^{1-\nu}$, is the intergenerational discount factor, where N is the number of offspring an individual has over her lifetime. Here λ $(0 < \lambda < 1)$ should be understood as the individual's weighting of his offsprings' utility relative to her own utility[3]. The individual discounts the utility of each additional child by a factor of $1-\nu$, where $0 < \nu < 1$ because we assume diminishing marginal returns from offspring.[4]

We let wages, w_t, be given by the earnings function $w_t(z_t, h_t)$, which depends on the individual's time invariant characteristics, choices that affect human capital accumulated with work experience, and the current level of labor supply, h_t. The choices and characteristics of parents are mapped onto offspring's characteristics, x', via a stochastic production function of several variables. The offspring's characteristics are affected by parents' time invariant characteristics, parents' monetary and time investments, and presence and timing of siblings. These variables are mapped into the child's skill and educational outcome by the function $M(x'|z_{T+1})$. Since z_{T+1} includes all parents choices and characteristics and contains information on the choices of time inputs and monetary inputs. Because z_{T+1} also contains information on all birth decisions, it captures the number of siblings and their ages. We assume there are four mutually exclusive outcomes of offspring's characteristics: Less than high school (LH), High school (HS), Some college (SC) and College (Coll). Therefore $M(x'|z_{T+1})$ is a mapping of parental inputs and characteristics into a probability distribution over these four outcomes.

We normalize the price of consumption to 1. Raising children requires parental time, d_t, and also market expenditure. The per-period cost of expenditures from raising a child is denoted pc_{nt}. Therefore the per period

[3]technically this definition is assuming he has one period left in his lifetime and only have one child.

[4]Note that this formulation can be written as an infinite discounted sum (over generations) of per-period utilities as in the Barro-Becker formulation.

budget constraint is given by:

$$w_t \geq c_t + pc_{nt} \tag{2}$$

.The sequence of optimal choice for both discrete choice and consumption is denoted as I^o_{kt} and c^o_t respectively. We can thus denote the expected lifetime utility at time $t = 0$ of a person with characteristics x in generation g, excluding the dynastic component, as:

$$U_{gT}(x) = E_0\left[\sum_{t=0}^{T} \beta^t [\sum_{k=0}^{17} I^o_{kt}\{u_{1kt}(z_t) + \varepsilon_{kt}\} + u_{2t}(c^o_t, z_t)]|x\right] \tag{3}$$

The total discounted expected lifetime utility of an adult in generation g including the dynastic component is:

$$U_g(x) = U_{gT}(x) + \beta^T \lambda N^{-\nu} E_0\left[\sum_{n=1}^{N} U_{g+1,n}(x'_n)|x\right], \tag{4}$$

where $U_{g+1,n}(x'_n)$ is the expected utility of child n ($n = 1, .., N$) with characteristics x'. In this model individuals are altruistic and derive utility from offsprings' utility, subject to discount factors β,and $\lambda N^{1-\nu}$.

To simplify presentation of the model we assume that pc_{nt} is proportional to individual's current wages and the number of children, but we allow this proportion to depend on state variables. This assumption allows us to capture the differential expenditures on children made by individuals with different incomes and characteristics. Practically this allows us to observe differences in social norms of child rearing among different socioeconomic classes .[5]Explicitly we assume that

$$pc_{nt} = \alpha_{Nc}(z_t)(N_t + b_t)w_t(x, h_t) \tag{5}$$

and, incorporating the assumption that individuals can not borrow or save and equation (5) the budget constraint becomes:

$$w_t(x, h_t) = c_t + \alpha_{Nc}(z_t)(N_t + b_t)w_t(x, h_t) \tag{6}$$

Solving for consumption from equation (6) and substituting for consumption in the utility equation, we can rewrite the third component of the per-period utility function, specified as $u_{2kt}(z_t)$, as a function of just z_t:

$$u_{2kt}(z_t) = u_t[w_t(x, h_t) - \alpha_{Nc}(z_t)(N_t + b_t)w_t(x, h_t), z_t] \tag{7}$$

Note that the discrete choices now map into different levels of utility from consumption. Therefore we can get rid of the consumption as choice and write the systematic contemporary utility associated with each discrete choice k as:

$$u_{kt}(z_t) = u_{1kt}(z_t) + u_{2kt}(z_t). \tag{8}$$

Incorporating the budget constrain manipulation, we can rewrite the Equation (3) as:

$$U_{gT}(x) = E_0\left[\sum_{t=0}^{T} \beta^t \sum_{k=0}^{17} I^o_{kt}[u_{kt}(z_t) + \varepsilon_{kt}]|x\right]. \tag{9}$$

Discussion Alvarez (99) analyzes and generalizes the conditions under which dynastic models with endogenous fertility lead to intergenerational persistence in income and wealth. Following his analysis, we show which assumptions are relaxed in our model and lead to persistence in income. The first is constant cost per-child. In our model the per-period cost of raising and transferring human capital is the costs described in Equations 5 and 6, as well as the opportunity costs of time input in children d_t: $w(x, 1 - d_t - leisure_t)$. Time input in children as well as labor market time are models as discrete choice with three levels. This introduces non-linearities. Even if we were able to capture the proportional increase in time with children as the number of children increases, the non-linearity in labor supply decisions implies that the opportunity cost of time with children is not linear. Thus the

[5] In general, individuals can choose expenditures on children but we do not observe spending in our data used for estimation in this proposal.

cost of transfer of human capital per child are not constant. Furthermore, in contrast to standard dynastic models and those analyzed in Alvarez (99) we incorporate dynamic elements of the life-cycle, that involve age effect and experience. The opportunity cost of time with children therefore incorporate returns to experience, which are non-linear. The non-linearities involved in labor supply are realistic, parents labor market time is often not proportional to the number of children they have, and hours in the labor market, for a given wage rate are not always flexible and depend on occupation and jobs. Furthermore, fertility decisions are made sequentially, and due to age effects, the cost of a child vary over the life-cycle. The second condition is non-separability in preferences, aggregation of the utilities from children and the feasible set. In our model, the latter is relaxed; that is, the separability of the feasible set across generations. This is because the opportunity costs of the children depend on the their education and labor market skill. However, education and labor market skills of children are linked with their parents' skills and education through the production function of education.

2.1 Optimal Discrete Choice

The individual then chooses the sequence of alternatives yielding the highest utility by following the decision rule $I(z_t, \varepsilon_t)$ where ε_t is the vector $(\varepsilon_{0t}, ..., \varepsilon_{17t})$. The optimal decision rules are given by:

$$I^o(z_t, \varepsilon_t) = \arg\max_I E_I \left[\sum_{t=0}^{T} \beta^t \{ \sum_{k=0}^{17} I_{kt}[u_{kt}(z_t) + \varepsilon_{kt}]\} + \beta^T \lambda N^{-\nu} \sum_{n=1}^{N} U_{g+1,n}(x'_n)|x \right] \tag{10}$$

where the expectations are taken over the future realizations of z and ε induced by I^o. In any period $t < T$, the individual' maximization problem can be decomposed into two parts: the utility received at t plus the discounted future utility from behaving optimally in the future. Therefore we can write the value function of the problem, which represents the expected present discounted value of life time utility from following I^o, given z_t and ε_t, as:

$$V(z_{t+1}, \varepsilon_{t+1}) = \max_I E_I \left(\sum_{t'=t+1}^{T} \beta^{t'-t} \sum_{k=0}^{17} I_{kt'}[u_{kt'}(z_{t'}) + \varepsilon_{kt'}] + \beta^{T-t'} \lambda N^{-\nu} \sum_{n=1}^{N} U_{g+1,n}(x'_n)|z_t, \varepsilon_t \right) \tag{11}$$

By Bellman's principle of optimality, the value function can be defined recursively as:

$$\begin{aligned} V(z_t, \varepsilon_t) &= \max_I \left[\sum_{k=0}^{17} I_{kt} \{u_{kt}(z_t) + \varepsilon_{kt} + \beta E(V(z_{t+1}, \varepsilon_{t+1})|z_t, I_{kt} = 1)\} \right] \\ &= \sum_{k=0}^{17} I^o_{kt}(z_t, \varepsilon_t)[u_{kt}(z_t) + \varepsilon_{kt}] + \beta \sum_{z_{t+1}} \int V(z_{t+1}, \varepsilon_{t+1}) f(\varepsilon_{t+1}) d\varepsilon_{t+1} F(z_{t+1}|z_t, I^o_{kt} = 1)] \end{aligned} \tag{12}$$

where $f(\varepsilon_{t+1})$ is the continuously differentiable density of $F_\varepsilon(\varepsilon_{0t}, .., \varepsilon_{17t})$, and $F(z_{t+1}|z_t, I_{kt} = 1)$ is a transition function for state variables which is conditional on choice k. In this simple version, the transitions of the state variables are deterministic given the choices of labor market experience, time spent with children and number of children.

Since ε_t is unobserved, we further define the *ex ante* (or integrated) value function, $V(z_t)$, as the continuation value of being in state z_t before ε_t is observed by the individual. Therefore $V(z_t)$ is given by integrating $V(z_t, \varepsilon_t)$ over ε_t. Define the probability of choice k at age t by $p_k(z_t) = E[I^o_{kt} = 1|z_t]$, the *ex ante* value function can be written as

$$V(z_t) = \sum_{k=0}^{17} p_k(z_t) \left[u_{kt}(z_t) + E_\varepsilon[\varepsilon_{kt}|I_{kt} = 1, z_t] + \beta \sum_{z_{t+1}} V(z_{t+1}) F(z_{t+1}|z_t, I_{kt} = 1) \right] \tag{13}$$

In this form $V(z_t)$ is now a function of the conditional choice probabilities, the expected value of the preference shock, the per-period utility, the transition function, and the *ex ante* continuation value. All components expect the conditional probability and the *ex ante* value function are primitives of the initial decision problem. By writing the conditional choice probabilities as a function of just the primitives and the ex ante value function, we can characterize the optimal solution of problem (i.e. the ex ante value function) as implicitly dependent on just the primitives of the original problem.

To create such a model we define the conditional value function, $\upsilon_k(z_t)$, as the present discounted value (net of ε_t) of choosing k and behaving optimally from period $t = 1$ on:

$$\upsilon_k(z_t) = u_{kt}(z_t) + \beta \textstyle\sum_{z_{t+1}} V(z_{t+1})F(z_{t+1}|z_t, I_{kt} = 1) \tag{14}$$

The conditional value function is the key component to the conditional choice probabilities. Equation(10) can now be rewritten using the individual's optimal decision rule at t to solve:

$$I^o(z_t, \varepsilon_t) = \arg\max_I \textstyle\sum_{k=0}^{17} I_{kt}[\upsilon_k(z_t) + \varepsilon_{kt}] \tag{15}$$

Therefore the probability of observing choice k, conditional on z_t is $p_k(z_t)$ and is found by integrating out ε_t from the decision rule in Equation (15):

$$p_k(z_t) = \int I^o(z_t, \varepsilon_t) f_\varepsilon(\varepsilon_t)d\varepsilon_t = \int \left[\textstyle\prod_{k\neq k'} 1\{\upsilon_k(z_t) - \upsilon_{k'}(z_t) \geq \varepsilon_{kt} - \varepsilon_{tk'}\}\right] f_\varepsilon(\varepsilon_t)d\varepsilon_t \tag{16}$$

Therefore $p_k(z_t)$ is now entirely a function the primitives of the model (i.e. $u_{kt}(z_t), \beta, F(z_{t+1}|z_t, I_{kt} = 1)$, and $f_\varepsilon(\varepsilon_t)$) and the *ex ante* value function. Hence substituting equation (16) into Equation (46) gives a implicit equation defining the *ex ante* value function as a function only the primitives of the model:

3 A Generic Estimator of the Life-Cycle Dynastic Discrete choice Model

We use a partial solution, multi-stage estimation procedure to accommodate the non-standard features of the model. By assuming stationarity across generations and discrete state space in the dynamic programming problem we obtain an analytic representation of the valuation function. The alternative valuation function depends on the conditional choice probabilities (CCP), the transition function of the state variable, and the structural parameters of the model. In the first stage we estimate the conditional choice probabilities and the transition function. The second stage forms either moment conditions or likelihood functions to estimate the remaining structural parameters using a Pseudo Maximum Likelihood (PML) or a Generalized Methods of Moment (GMM) respectively. For each iteration in the estimation procedure the CCP is used generate valuation representation to form the terminal value in the life-cycle problem, which can then be solved by backward induction to obtain the life-cycle valuation functions.

3.1 An Alternative Representation of the Problem

The alternative representation of the continuation value of the intergenerational problem is developed below. The Hotz and Miller estimation technique for standard single agent problems is adapted to the dynastic problem using the following representation.

Proposition 1 *There exists an alternative representation for the ex-ante conditional value function at time t which is a function of only the primitives of the problem and the conditional choice probability as:*

$$\begin{aligned}\upsilon_k(z_t) \;\; = \;\; & u_{kt}(z_t) + \textstyle\sum_{t'=t+1}^{T} \beta^{t'-t} \sum_{s=0}^{17} \sum_{z_{t'}} p_s(z_{t'})[u_{st'}(z_{t'}) + E_\varepsilon(\varepsilon_{st'}|I_{st'} = 1, z_{t'})]F_k^o(z_{t'}|z_t) \\ & +\lambda\beta^{T-t} N_T{}^{-\nu} \textstyle\sum_{n=1}^{N_T} \sum_x V(x) \sum_{s=0}^{K_T} \sum_{z_T} M_k^n(x'|z_T)p_s(z_T)F_k^o(z_T|z_t)\end{aligned} \tag{17}$$

where $F_k^o(z_{t'}|z_t)$ is the $t' - t$ period ahead optimal transition function, recursively defined as:

$$F_k^o(z_{t'}|z_t) = \begin{cases} F(z_{t'}|z_t, I_{kt} = 1) & \text{for } t' - t = 1 \\ \sum_{r=0}^{17} \sum_{z_{t'-1}} p_r(z_{t'-1})F(z_{t'}|z_{t'-1}, I_{rt'-1} = 1)F_k^o(z_{t'-1}|z_t) & \text{for } t' - t > 1 \end{cases}$$

where N_T is the number children induced from Z_T, K_T is the number of possible choice combinations available to the individual in the terminal period (in which birth is no longer feasible) and $M_k^n(x'|z_T) = M(x'|z_T)$ conditional on $I_{kT} = 1$ for the n^{th} child born in a parent's life-cycle.

Let $e_k(p,z)$ represent the expected preference shocks conditional on choice k being optimal in state z. The expected preference shocks are written in this notation to convey the shock as a function of the conditional choice probability (see Hotz and Miller (1993)). For example, in the Type 1 extreme value case, $e_k(p,z)$ is given by $\gamma - \ln[p_k(z)]$ where γ is Euler's constant. From the representation in Proposition 1 we can define the ex-ante conditional lifetime utility at period t, excluding the dynastic component as:

$$U_k(z_t) = u_{kt}(z_t) + \sum_{t'=t+1}^{T} \beta^{t'-t} \sum_{s=0}^{17} \sum_{z_{t'}} p_s(z_{t'})[u_{st'}(z_{t'}) + e_s(p, z_{t'})]F_k^o(z_{t'}|z_t)$$

Because $U_k(z_t)$ is a function of only the primitives of the problem and the conditional choice probabilities,we can write an alternative representation for the ex-ante value function at time t :

$$V(z_t) = \sum_{k=0}^{17} p_k(z_t)[U_k(z_t) + e_k(p,z_t) + \lambda\beta^{T-t} {N_T}^{-\nu} \sum_{n=1}^{N_T} \sum_x V(x) \sum_{s=0}^{K_T} \sum_{z_T} M_k^n(x'|z_T) p_s(z_T) F_k^o(z_T|z_t)] \tag{18}$$

Equation (18) is satisfied at every state vector z_t. The problem is stationary over generation, so $z_t = x$ at period $t = 0$ because there is no history of decisions in the state space, and hence the initial state space has finite support on the integers $\{1, ..., X\}$. We define the optimal lifetime intergenerational transition function as $M_k^o(x'|x) = \sum_{n=1}^{N_T} \sum_{s=0}^{K_T} \sum_{z_T} p_s(z_T) M_k^n(x'|z_T) F_k^o(z_T|x)$. M_k^o can be interpreted as the probability that average descendent of the individual with characteristic x', given that his parents have characteristics x, chooses decision k in the first period and behaves optimally from period 1 to T of parent's life-cycle. Now, we can express the components of Equation (18) in vector or matrix form:

$$V_0 = \begin{bmatrix} V(1) \\ . \\ . \\ . \\ V(X) \end{bmatrix}, \quad U(k) = \begin{bmatrix} U_k(1) \\ . \\ . \\ . \\ U_k(X) \end{bmatrix}, \quad E(k) = \begin{bmatrix} e_k(p,1) \\ . \\ . \\ . \\ e_k(p,X) \end{bmatrix}, P(k) = \begin{bmatrix} p_k(1) \\ . \\ . \\ . \\ p_k(X) \end{bmatrix},$$

$$\imath_X = \begin{bmatrix} 1 \\ . \\ . \\ . \\ 1 \end{bmatrix}_{Xx1}, \text{and } M^o(k) = \begin{bmatrix} M_k^o(1|1) & ... & M_k^o(X|1) \\ . & & \\ . & & \\ . & & \\ M_k^o(1|X) & ... & M_k^o(X|X) \end{bmatrix}$$

Using these components the vector of *ex ante* value function can be expressed as:

$$V_0 = \sum_{k=0}^{17} P(k) * [U(k) + E(k) + \lambda\beta^T N_T^{-\nu} M^o(k)]V_0 \tag{19}$$

where $*$ refers to element by element multiplication. Rearranging the terms and solving for V_0 we obtain:

$$V_0 = [I_X - \lambda\beta^T N_T^{-\nu} \sum_{k=0}^{17} \{P(k)\imath_X'\} * M^o(k)]^{-1} \sum_{k=0}^{17} P(k)[U(k) + E(k)] \tag{20}$$

where I_X denote as the $X \times X$ identity matrix. Equation (20) is based on the dominant diagonal property, which implies that the matrix $I_X - \lambda\beta^T N_T^{-\nu} \sum_{k=0}^{17} \{P(k)\imath_X'\} * M^o(k)$ is invertible.

3.2 Estimation

We parameterized the period utility by a vector θ_2, $u_{kt}(z_t, \theta_2)$, the period transition on the observed states is parameterized by a vector θ_3, $F(z_t|z_{t-1}, I_{kT} = 1, \theta_3)$, the intergenerational transitions on permanent characteristics is parameterized by a vector θ_5, $M^n(x'|z_{T+1}, \theta_4)$, and the the earnings function is characterized by a vector θ_5, $w_t(x, h_t, \theta_5)$. Therefore the conditional value functions, decision rules, and choice probabilities now also depend on $\theta \equiv (\theta_2, \theta_3, \theta_4, \theta_5, \beta, \lambda, \nu)$. Standard estimates of dynamic discrete choice models involve forming the likelihood functions from the conditional choice probability derived in Equation (16). This involves solving the value function for each iteration of the likelihood function. The method used to solve the valuation function depends on the nature of the optimization problems, and normally falls into one of two cases

1. Finite horizon problems: The problem has an end date (like in a standard life-cycle problem) and hence future value function is obtained via backwards recursion.

2. Stationary infinite horizon problem: The valuation is obtained via a contraction mapping.

A dynastic discrete choice model in unusually because it involves both a finite horizon problem and an infinite horizon problem. Solving both problems for each iteration of the likelihood function is a computationally infeasible for all but the simplest of models. We avoid solving the stationary infinite horizon problem in estimation by replacing the terminal value in the life-cycle problem with Equation (20). This alters the problem to a finite horizon problem which can be solved via backwards recursion since the flow utility function is:

$$\upsilon_k(z_T) = u_{kT}(z_T) + \lambda {N_T}^{-\nu} \sum_x V(x) \sum_{n=1}^{N_T} M_k^n(x'|z_T) \tag{21}$$

Since $u_{kT}(z_T)$ is parameterized by θ_2, the transition $M_k^n(x'|z_T)$ is known since can be estimated from the data. Observing $F_\varepsilon(\varepsilon_{0t}, .., \varepsilon_{17t})$ and calculating $V(x)$ via equation (20)[6], we can calculate the *ex ante* value function at T using $V(z_T) = \sum_{k=0}^{17} \int I_{kI}^0(z_T, \varepsilon_T)[\upsilon_k(z_T) + \varepsilon_{kT}] f_\varepsilon(\varepsilon_T) d\varepsilon_T$. The conditional value function for T-1 is given by $\upsilon_k(z_{T-1}) = u_{kT-1}(z_{T-1}) + \beta \sum_{z_T} V(z_T) F(z_T|z_{T-2}, I_{kT} = 1)$. This is continued backwards given $\upsilon_k(z_{T-1})$ to form value function at T-2, and so on.

The backward induction procedure outlined above shows that only $M_k^n(x'|z_T)$ in equation (21) and (20) depends on the next generation's outcome. Thus we can estimate the intergenerational problem with only two generations of data, as is the case in the standard stationary discrete choice models (see Rust (1987) for example). To estimate the intergenerational problem we let I_{dtg}, z_{dtg},and ε_{dtg} respectively indicate the choice, observed state, and unobserved state at age t in the generation g of dynasty d. Forming the conditional choice probabilities for each individual in the first observed generation of dynasty d at all age t yields the components necessary for estimation. Estimation proceeds in two steps.

Step 1: In the first step we estimate the CCP, transition, earnings functions necessary to compute the inversion in Equation (20). The expectation of observed choices conditional on the observed state variable gives an empirical analogue to the conditional choice probabilities at the true parameter values of the problem, θ_1^o, allowing us to estimate the CCPs, we denote this estimate by $\widehat{p_k(z_{dt1})}$. We also estimates θ_3, θ_4, and θ_5 which parameterize the transition and earnings functions $F(z_t|z_{t-1}, I_{kT} = 1, \theta_3)$, $M^n(x'|z_{T+1}, \theta_4)$ and $\omega_t(x, h_t, \theta_5)$ respectively in this step.

Step 2: The second step can be estimated two ways, the first is a pseudo maximum likelihood (PML) and the second is a generalized method of moment (GMM). We can use a pseudo maximum likelihood method and not a pure maximum likelihood estimator because part of the likelihood function is concentrated out using the data. With D dynasties, the PML estimates of $\theta_0 = (\theta_2, \beta, \lambda, \nu)$ are obtained via:

$$\widehat{\theta}_{0PML} = \underset{\theta_0}{\arg\max} \left(\sum_{dt1=1}^{D} \sum_{t=0}^{T} \sum_{k}^{17} I_{dt1} \ln[p_k(z_{dt1}; \theta_0, \widehat{\theta}_3, \widehat{\theta}_4, \widehat{\theta}_5)] \right) \tag{22}$$

where $p_k(z_{dt1}; \theta_0, \widehat{\theta}_3, \widehat{\theta}_4, \widehat{\theta}_5)$ is the CCP defined in equation (16) with the conditional value function replaced with $\upsilon_k(z_{dt1}, \theta_0, \widehat{\theta}_3, \widehat{\theta}_4, \widehat{\theta}_5)$; which is calculated using the backward recursion using the estimated choice probabilities and the transition functions outlined in step 1.

[6]This manipulation is possible because the alternative value function in equation (20) is a function of only the parameters of the model and the CCP. Since the CCP can be estimated directly from the data, backward recursion becomes possible because the decision in the last period, T, is similar to a static problem when the value of children is replaced with equation (20).

An alternative second step GMM estimator is formed using the inversion found in Hotz and Miller (1992). Under the assumption that ε is distributed independently and identically as type I extreme values, then Hotz and Miller inversion implies that

$$\log\left(p_k(z_{dt1};\theta_0,\widehat{\theta}_3,\widehat{\theta}_4,\widehat{\theta}_5)/p_K(z_{dt1};\theta_0,\widehat{\theta}_3,\widehat{\theta}_4,\widehat{\theta}_{53})\right)=\upsilon_k(z_{dt1},\theta_0,\widehat{\theta}_3,\widehat{\theta}_4,\widehat{\theta}_5)-\upsilon_K(z_{dt1},\theta_0,\widehat{\theta}_3,\widehat{\theta}_4,\widehat{\theta}_5) \quad (23)$$

for any normalized choice K. We can use $\widehat{p_k(z_{dt1})}$, estimated from stage 1, to form an empirical counterpart to equation (23) and estimate the parameters of our model. The moment conditions can be obtained from the difference in the conditional valuation functions calculated for choice k and the base choice 0. The following moment conditions are produced for an individual at age $t \in \{17, ..., 55\}$:

$$\xi_{jdt}(\theta_0)\equiv\upsilon_k(z_{dt1},\theta_0,\widehat{\theta}_3,\widehat{\theta}_4,\widehat{\theta}_5)-\upsilon_0(z_{dt1},\theta_0,\widehat{\theta}_3,\widehat{\theta}_4,\widehat{\theta}_5)-\ln\left[\widehat{p_k(z_{dt1})}/\widehat{p_0(z_{dt1})}\right] \quad (24)$$

therefore there are 17 orthogonality conditions therefore $j = 1, ..., 17$. Let $\xi_{dt}(\theta_0)$ be the vector of moment conditions at t, these vectors are defined as $\xi_{dt}(\theta_0) = (\xi_{1dt}(\theta_0), \xi_{2dt}(\theta_0), ...\xi_{17dt}(\theta_0))'$. Therefore $E[\xi_{dt}(\theta_0^o)|z_{dt}]$ converges to 0 for every consistent estimator of true conditional choice probability, $p_k(z_{dt1};;\theta_0,\widehat{\theta}_3,\widehat{\theta}_4,\widehat{\theta}_5)$, for $t \in \{17, ..., 55\}$, and where θ_0^o is the true parameter of the model. Define $\xi_d(\theta_0) \equiv \left(\xi_{d1}(\theta_0)', ..., \xi_{dT}(\theta_0)'\right)'$ as the the vector of moment restrictions for a given individual over time and define a weight matrix as $\Phi(\theta_0) \equiv E_t[\xi_d(\theta_0)\xi_d(\theta_0)']$. Then the GMM estimate of θ_0 is obtained via:

$$\widehat{\theta}_{02SGMM}=\underset{\theta_0}{\arg\min}[1/D\textstyle\sum_{d=1}^{D}\xi_d(\theta_0)]'\widehat{\Phi}[1/D\sum_{d=1}^{D}\xi_d(\theta_0)]. \quad (25)$$

where $\widehat{\Phi}$ is a consistent estimator of $\Phi(\theta^o)$.

3.3 Monte Carlo Study

To compare the dynamics of the model in a numerical example and to examine the performance of the estimation, we use a simple human capital investment model with intergenerational transfers which has the two period model structure of Section 1. We generate simulated data from the model for given parameter values, compare the dynamics, and estimate the model parameters for the generated data set. We estimated the parameters using the Nested Fixed Point (NFP) and Pseudo Maximum Likelihood (PML) estimators described above. The estimations are repeated for both algorithms for different specifications of the model in terms of sample size (i.e., for 1000, 10, 000, 20, 000, 40, 000). The number of structural parameters estimated including the discount factors are 3.

For illustrative purposes we start with the model in which the period utility function, $u_k(z_t)$, has the following linear form: the individual chooses whether to invest or not $I_k \in \{0, 1\}$ in each period $t \in \{0, 1\}$. We assume that individuals may have only one child, $N \leq 1$, and receive the following utilities associated with each choice:

$$u_k(z_t)=\left\{\begin{array}{cc} z_t & \text{if } k=0 \\ (1-\theta)z_t & \text{if } k=1 \end{array}\right\}$$

where $F_\varepsilon(\varepsilon_t)$ is the choice specific, unobservable part of the utility and assumed to be independently distributed type 1 extreme values.

The value of the vector z_t is subject to change each period because of different choices made in each period and because individual characteristics like skill and education, given in the vector x, may transition over time. In the example, environment the individual starts the life-cycle with a particular set of character traits, which can be denoted as $z_t \in (0.5, 0.6, 0.7, 0.8, 0.9)$. Note that at $t = 0$ the individual has not made any choices yet, so the vector z_0 depends fully on initial characteristics x. The value of z_1 is given by the transformation function

$F_k(z_t|z_{t-1})$ In **??** we modeled the transformation function $F_k(z_t|z_{t-1})$ as deterministic, but here we assume a stochastic process given by the transition matrix:

$$F_0(z_t|z_{t-1}) = \begin{pmatrix} .85 & .13 & .02 & 0 & 0 \\ .04 & .85 & .09 & .02 & 0 \\ .01 & .04 & .85 & .09 & .01 \\ 0 & .01 & .05 & .85 & .09 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \text{ and } F_k(z_t|z_{t-1}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ .1 & .9 & 0 & 0 & 0 \\ .13 & .27 & .6 & 0 & 0 \\ .01 & .11 & .28 & .6 & 0 \\ 0 & .04 & .13 & .23 & .6 \end{pmatrix}.$$

The individual's character traits in the next period are determined by the probabilities in the corresponding row,

where each row corresponds to one of the initial values $z_0 \in (0.5, 0.6, 0.7, 0.8, 0.9)$, and each column represents character traits in the next period, $z_1 \in (0.5, 0.6, 0.7, 0.8, 0.9)$. The transition is such that an individual with character traits $z_0 = 0.5$ with who chooses not to have a child such that the choice vector $I_0 = 0$ will have characteristics $z_1 = .5$ with a probability of 0.85. In this simplified model, the next generation's (offspring's) initial characteristics z_0' depend only on the sum of the financial investment decisions in the life-cycle where z_{T+1}. This educational outcome of the offspring is determined by the intergenerational transition function:

$$M(z_0' \mid z_{T+1}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & .1 & .4 & .4 & .1 \\ 0 & 0 & .04 & .06 & .9 \end{pmatrix}$$

where z_{T+1} can take values in {0,1,2}.The next generation's starting character traits are determined by the prob-

abilities given in the row, where each row corresponds to one of the values of $z_{T+1} \in (0, 1, 2)$ and the first row represents investment level $z_{T+1} = 0$. If the individual invests nothing, then the next generation will have the lowest consumption value with complete certainty. The transition is such that an individual who opts to invest 2 times in the life-cycle has a faces a probability of 0.9 that the next generation will start his life-cycle with the characteristics $z_0' = 0.9$.

We simulated the model for a given values of the parameters of the model, $(\theta_2, \beta, \lambda) = (0.25, 0.8, 0.95)$, where θ is the structural parameter of interest which gives the marginal cost of investment, and λ and β are the generational and time discount factors respectively. We solve the dynamic problem for data sets of $1,000$, $10,000$, $20,000$, $40,000$ individual dynasties and repeat the simulation 100 times. For the conditional choice probability (CCP) estimation, the initial consistent estimates are estimated nonparametrically using the generated sample. Next, we estimate the model by NFP and PML.[7] Table 2 presents the result of the estimation for each specification. We find, not surprisingly, that the finite sample properties of the estimators improve monotonically with sample size. In the NFP estimation, the Mean Square Error (MSE) of the parameter θ drops quickly as the sample size increases. The results for the discount factors are similar: MSE fall as sample size increases. In the PML estimation we observe a similar pattern for all estimators. We obtain similar results from the NFP and PML estimations. For the sample size of $1,000$ the PML estimate of the MSE of θ_0 is 0.00249, compared with 0.00288 from the NFP. The PML estimate of the MSE of λ is 0.01253 compared to 0.00901, and PML estimate of the MSE of β is 0.00396 compared to 0.00305. For the sample sizes $10,000$, $20,000$ and $40,000$, the MSE obtained from PML estimation is lower than the MSE obtained from the NFP, but the magnitudes are still very close. In terms of biases, the two estimation algorithms are also quite similar. The major difference between the two estimation algorithms computational time, which vary greatly between the NSP and PML even though we simulate a very simple model. The average computational time for the NFP for a sample of 1000 is 347.6 seconds, but only 0.65 seconds for the PML estimation, meaning the PML was 530 times faster. For the sample size of $40,000$ computation times are 509.8 and 12.6 seconds for the NFP and PML respectively, a ratio of 40.4.

[7] As illustrated in the estimation section, intergenerational models at the final step can be estimated either by PML or generalized method of moments (GMM). For this simulation study we used the PML because it is more comparable to the Full Solution Maximium Likelihood.

3.4 Large Sample Properties

It is well known in the econometric literature that under certain regularity conditions, pre-estimation does not have any impact on the consistency of the parameters in the subsequent steps of a multistage estimation (Newey, 1984; Newey and McFadden, 1994; Newey, 1994). The asymptotic variance, however, is affected by the pre-estimation. In order to conduct inference in this type of estimation, one has to correct the asymptotic variance for the pre-estimation. The method used for correcting the variance in the final step of estimation depends on whether the pre-estimation parameters are of finite or infinite dimension. Unfortunately, our estimation strategy combines both finite- and infinite-dimensional parameters. Combining results from two sources (Newey, 1984; Newey and McFadden, 1994), however, allows us to derive the corrected asymptotic variance for our estimator.

We proposed two estimators: A PML-estimator and a GMM-estimator. The PML-estimator can also be written as a GMM-estimator by using the first order condition of the optimization problem as the moment condition. As such we will only derive the large sample property for a moment based estimator where the moment can either be the first condition of the PML-estimator defined in Equation (22) or the orthogonal condition defined in Equation (24). For ease of notation we will use the same notation to represents these two types of orthogonality conditions. Following Newey (1984), we can write the sequential-moments conditions for the first- and third-step estimation as a set of joint moment conditions:

$$\overline{\xi}_d(Z_d, \theta_0, \theta_3, \theta_4, \theta_5, \psi) = \left[\xi_{dF}(Z_d, \theta_3), \xi_{dM}(Z, \theta_4), \xi_{dW}(Z, \theta_5), \xi_d(Z_d, \theta_0, \theta_3, \theta_4, \theta_5, \psi)\right]',$$

where$\xi_{dF}(Z_d, \theta_3)$ is the orthogonality condition from the estimation of the life-cycle transition function, $\xi_{dM}(Z, \theta_4)$ is the orthogonality condition from the estimation of the generation transition function, $\xi_{dW}(Z, \theta_5)$ is the orthogonality condition from the estimation of the earnings equation, and $\xi_d(Z_d, \theta_0, \theta_3, \theta_4, \theta_5, \psi)$ is the moment conditions from the second-step estimation defined in Equation (24). Regardless of the estimation method used to estimate θ_3, θ_4, and θ_5 they can always be expressed as moment conditions. Let $\theta = (\theta_0, \theta_3, \theta_4, \theta_5)'$, with the true value denoted by θ^o. Each element of infinite dimensional parameter, ψ, can be written as a conditional expectation. Redefine each element as $\psi^k(z^k) = f_{z^k}(z^k) E\left[\widetilde{I}_{dk} \mid z^k\right]$, where $\widetilde{I}_{dkt} = [1, I_{dkt}]'$ for the estimation of $p_k(z_{dt})$. Therefore, $\psi^{k(D)}(z^k) = \frac{1}{D}\sum_{d=1}^{D} \widetilde{I}_{dk} J_{\delta_N}(z^k - z_d^k)$. The conditions below ensure that $\psi^{(D)}$ is close enough to ψ^o for D large enough, in particular that $\sqrt{D}\left\|\psi^{(N)} - \psi^o\right\|^2$ converges to zero.

A3: *There is a version of* $\psi^o(z)$ *that is continuously differentiable of order* κ, *greater than the dimension of* z *and* $\psi_1^o(z) = f_z(z)$ *is bounded away from 0.*

A4: $\int J(u)\,\mathrm{d}u = 1$ *and for all* $j < \kappa$, $\int J(u)\left(\bigotimes_{s=1}^{j} u\right)\mathrm{d}u = 0.$

A5: *The bandwidth,* δ_D, *satisfies* $D\delta_D^{2\dim(z)}/(\ln(D))^2 \to \infty$ *and* $D\delta_D^{2\kappa} \to 0$.

A6: *There exists a* $\Psi(Z)$, $\epsilon > 0$, *such that*

$$\left\|\nabla_\theta \overline{\xi}_d(Z, \theta, \psi) - \nabla_\theta \overline{\xi}_d(Z, \theta^o, \psi^o)\right\| \le \Psi(Z)\left[\left\|\theta - \theta^o\right\|^\epsilon + \left\|\psi - \psi^o\right\|^\epsilon\right]$$

and $E[\Psi(Z)] < \infty$.

A7: $\theta^{(D)} \to \theta^o$ *with* Θ^o *in the interior of its parameter space.*

A8: (Boundedness)

(i) Each element of $\overline{\xi}_d(Z, \theta, \psi)$ is bounded almost surely: $E[\left\|\overline{\xi}_d(Z, \theta, \psi)\right\|^2] < \infty$;

(ii) $p_{dkt} \in (0, 1)$, for all k.

(iii) $\xi_{dF}(Z_d, \theta_3)$, $\xi_{dM}(Z, \theta_4)$ and $\xi_{dW}(Z, \theta_5)$ are continuously differentiable in θ_3,θ_4, and θ_5 respectively.

Proposition 2 *Under A1–A8 and the influence,* $\Phi(Z)$*, defined in the appendix,*

$$\sqrt{N}\left(\theta^{(D)}-\theta^{o}\right)\Rightarrow N(0,\,\Sigma(\theta^{o})),$$

where

$$\begin{aligned}\Sigma(\theta^{o}) &= E\left[\nabla_{\theta}\overline{\xi}_{d}(Z)\Omega_{d}^{-1}\nabla_{\theta}\overline{\xi}_{d}(Z)'\right]^{-1}E\left[\nabla_{\theta}\overline{\xi}_{d}(Z)\Omega_{d}^{-1}\left\{\overline{\xi}_{d}(Z)+\Phi(Z)\right\}\left\{\overline{\xi}_{d}(Z)+\Phi(Z)\right\}'\Omega_{d}^{-1}\nabla_{\theta}\overline{\xi}_{d}(Z)'\right]\\ &\quad\times E\left[\nabla_{\theta}\overline{\xi}_{d}(Z)\Omega_{d}^{-1}\nabla_{\theta}\overline{\xi}_{d}(Z)'\right]^{-1}.\end{aligned}$$

Assumptions A3–A8 are standard in the semiparametric literature, see Newey and McFadden (1994) for details. One can now use Theorem 1 to calculate the standard for all the parameters in our estimation. The proof of Theorem 1 will follow from checking the conditions for Theorem 8.12 in Newey and McFadden (1994).

4 Extensions

The dynastic framework developed so far in this paper has three major drawbacks. First, part of parental investment and transfer from parent to children are monetary in nature. Monetary investment and/or parental transfer, such as paying for college, purchasing a house, are most naturally characterized as a continuous choice. Second, the framework assume that gender does not matter. However, there are significant differences in the cost, choices and opportunities over an individual lifetime which are gender specific. Third, which is related to gender but a specific to it, is that individuals normally form households and it take a man and a woman to reproduces, and fertility is central in the model. In this section we consider extensions to the basic framework that account for these three shortcomings.

4.1 Continuous Choice and Transfer

In order for the estimation technique developed above to be applicable to a dynastic framework two features were present. First, all choices were discrete, and second, all systematic state variables, at the initial stage and in every period during the life-cycle have a discrete state space. We replace these assumptions with two weaker assumptions. The first is that there must be at least one discrete choice variable. This requirement is easily satisfied, as birth decision is naturally discrete. The second, is that the initial systematic state variable, i.e. endowment that an individual starts the adult life with, must belong to a finite set with discrete support. This is weaker than the original assumption but is a more restrictive requirement but is satisfied in a non-trivial number of economic dynastic models. For example, in models where human capital is the major intergenerational transfers and even in model of bequests once the amount transferred is discretized. In practice, most dynamic programing models the state space in normally discretized. This requirement, however, relaxes the assumption that state space is discrete for the entire lifetime and that all choice variable are discrete. While bequest and initial wealth still has to be discrete, the framework allows for any transfer and investment of parents make during over their life and map into discrete initial conditions of the child such as education, and houses or other assets discrete in nature are allowed.

We extend our framework by assuming that we observed data on the per-period expenditures of raising a child, pc_{nt}, which is continuous. Lets further assume that this expenditure is potentially productive, i.e. higher expenditure increases the probability of a higher level of education of the child. Lets redefined the vector of state variable z, to capture these new assumptions, $z_t = (\{I_{k1}\}_{k=0}^{17}, ..., \{I_{kt-1}\}_{k=0}^{17}, pc_{n1}, ..., pc_{nt-1}, x)$ with $x \in \{x_1, ..., x_{|X|}\}$, a discrete set with finite support. As before $M(x' \mid z_{T+1})$ is the integenerational transition probability of x conditional on parent's endowment, x, and parent's choices over his/her lifetime.

Let I_{kt}^{o} and pc_{nt}^{o} be the sequence of optimal choice over the parents lifetime. Also lets redefine the systematic part of current utility in Equation(8) as

$$u_{kt}(z_t,\, pc_{nt}){=}u_{1kt}(z_t)+u_t[\omega_t(x,h_t)-pc_{nt},z_t]. \tag{26}$$

There the lifetime expected utility excluding the dynastic component at the start of an adult's life becomes:

$$U_{gT}(x) = E_0\left[\sum_{t=0}^{T}\beta^t[\sum_{k=0}^{17} I_{kt}^o\{u_{1kt}(z_t, pc_{nt}^o) + \varepsilon_{kt}\}]|x\right]. \tag{27}$$

As before we can write the value function of the problem, which represents the expected present discounted value of life time utility from following I^o and pc_{nt}^o, given z_t and ε_t, as:

$$\begin{aligned} V(z_{t+1}, \varepsilon_{t+1}) \quad = \quad & \max_{I, pc_{nt}} E_{I,pc_n}\left(\left\{\sum_{t'=t+1}^{T}\beta^{t'-t}\sum_{k=0}^{17} I_{kt'}[u_{kt'}(z_{t'}, pc_{nt'}) + \varepsilon_{kt'}]\right.\right. \\ & \left.\left. +\beta^{T-t'}\lambda N^{-\nu}\sum_{n=1}^{N} E_T[U_{g+1,n}(x_n')|z_{T+1}]\right\} |z_{t+1}, \varepsilon_{t+1}\right) \end{aligned} \tag{28}$$

By Bellman's principle of optimality, the value function can be defined recursively as:

$$V(z_t, \varepsilon_t) = \sum_{k=0}^{17}\left(I_{kt}^o(z_t, \varepsilon_t)[u_{kt}(z_t, pc_{nt}^o(z_t)) + \varepsilon_{kt}] + \beta\int\left[\int V(z_{t+1}, \varepsilon_{t+1}) f_\varepsilon(\varepsilon_{t+1})d\varepsilon_{t+1}\right] dF_k(z_{t+1}|z_t, pc)]\right)$$

where $f_\varepsilon(\varepsilon_{t+1})$ is the continuously differentiable density of $F_\varepsilon(\varepsilon_{0t}, .., \varepsilon_{17t})$, and $F_k(z_{t+1}|z_t, pc)$ is a transition function for state variables which is conditional on choices $I_{kt}^o = 1$ and $pc_{nt}^o = pc$. Note that $I_{kt}^o(z_t, \varepsilon_t)$ is a function of z_t and ε_t while $pc_{nt}^o(z_t)$ is only a function z_t. This is a consequence of additive separability of the preferences shock which will not affect the continuous choice which will be demonstrated below. The *ex ante* value function is then

$$V(z_t) = \sum_{k=0}^{17} p_k(z_t)\left[u_{kt}(z_t, pc_{nt}^o(z_t)) + E_\varepsilon[\varepsilon_{kt}|I_{kt} = 1, z_t] + \beta\int V(z_{t+1})dF_k(z_{t+1}|z_t, pc)\right] \tag{29}$$

In this form $V(z_t)$ is now a function of the conditional choice probabilities, the continuous choice decision rule, the expected value of the preference shock, the per-period utility, the transition function, and the *ex ante* continuation value. All components expect the conditional probability, the continuous choice decision rule and the *ex ante* value function are primitives of the initial decision problem. By writing the conditional choice probabilities and the continuous choice decision rule as a function of just the primitives, and the ex ante value function, we can characterize the optimal solution of problem (i.e. the ex ante value function) as implicitly dependent on just the primitives of the original problem. Let define the conditional value function, $\upsilon_k(z_t, pc_{nt})$:

$$\upsilon_k(z_t) = \max_{pc_{nt}}\left[u_{kt}(z_t, pc_{nt}) + \beta\int V(z_{t+1})dF_k(z_{t+1}|z_t, pc)\right] \tag{30}$$

Therefore the probability of observing choice k, conditional on z_t, $p_k(z_t)$, is still given by:

$$p_k(z_t) = \int\left[\prod_{k\neq k'} 1\{\upsilon_k(z_t) - \upsilon_{k'}(z_t) \geq \varepsilon_{kt} - \varepsilon_{tk'}\}\right] f_\varepsilon(\varepsilon_t)d\varepsilon_t \tag{31}$$

However, the optimal continuous choice is found in two steps. First find the optimal choice conditional on $I_{kt} = 1$, called it $pc_{knt}(z_t)$. This characterized by the following Euler equation:

$$\frac{\partial u_{kt}(z_t, pc_{nt})}{\partial pc_{nt}} = -\beta\frac{\partial\int V(z_{t+1})dF_k(z_{t+1}|z_t, pc)}{\partial pc_{nt}}. \tag{32}$$

Then plug it into the conditional valuation function:

$$\upsilon_k(z_t) = \left[u_{kt}(z_t, pc_{knt}(z_t)) + \beta\int V(z_{t+1})dF_k(z_{t+1}|z_t, pc)\right] \tag{33}$$

and find the optimal discrete choice:

$$I^o(z_t, \varepsilon_t) = \arg\max_I \sum_{k=0}^{17} I_{kt}[\upsilon_k(z_t) + \varepsilon_{kt}].$$

Finally we obtain the optimal continuous choice by sets $pc_{nt}^o(z_t) = pc_{knt}(z_t)$ if $I_{kt}^o(z_t, \varepsilon_t) = 1$.

We now can find an alternative valuation function which is only a function of $p_k(z_t)$, $pc_{knt}(z_t)$, and primitives of the model. We can now state a more general version of Proposition3

Proposition 3 *There exists an alternative representation for the ex-ante conditional value function at time t which is a function only of the primitives of the problem and the conditional choice probability as:*

$$\begin{aligned} \upsilon_k(z_t) &= u_{kt}(z_t, pc_{knt}(z_t)) + \sum_{t'=t+1}^{T} \beta^{t'-t} \sum_{s=0}^{17} \int [p_s(z_{t'})[u_{st'}(z_{t'}, pc_{knt'}(z_t)) + E_\varepsilon(\varepsilon_{st'}|I_{st'}=1, z_{t'})] dF_k^o(z_{t'}|z_t) \\ &\quad + \lambda \beta^{T-t} N_T{}^{-\nu} \sum_{n=1}^{N_T} \sum_x V(x) \sum_{s=0}^{K_T} \int [M_k^n(x'|z_T) p_s(z_T)] dF_k^o(z_T|z_t) \end{aligned} \tag{34}$$

where $F_k^o(z_{t'}|z_t)$ is the $t'-t$ period ahead optimal transition function, recursively defined as:

$$F_k^o(z_{t'}|z_t) = \begin{cases} F(z_{t'}|z_t, I_{kt}=1, pc_{knt}(z_t)) & \textit{for } t'-t=1 \\ \sum_{r=0}^{17} \sum_{z_{t'-1}} p_r(z_{t'-1}) F(z_{t'}|z_{t'-1}, I_{rt'-1}=1, pc_{knt'-1}(z_{t'-1})) F_k^o(z_{t'-1}|z_t) & \textit{for } t'-t>1 \end{cases}$$

where N_T is the number children induced from Z_T, K_T is the number of possible choice combinations available to the individual in the terminal period (in which birth is no longer feasible) and $M_k^n(x'|z_T) = M(x'|z_T)$ conditional on $I_{kT}=1$ for the n^{th} child born in a parent's life-cycle.

This representation is similarly to the one in Proposition3 except for the inclusion of $pc_{knt}(z_t)$ and the replacement of integral for a summation deal with the continuous state variables over the life-cycle. The inversion and hence the estimation follows the through as before except we now need a first stage consistent estimate of $pc_{knt}(z_t)$ as well. This is obtained as $pc_{knt}(z_t) = E[pc_{nt}|z_t, I_{kt}=1]$. See Altug and Miller (1998) and Gayle and Golan (2012) for applications with continuous and discrete choices.

4.2 Household and Gender

We extend the basic framework to include household decisions and gender. To the best of our knowledge no other paper estimates dynastic models with household decisions. There are many model of household decisions; here we show how extend the model incorporate a unitary decisions maker. The framework can be extended to deal with collective household decisions, see Gayle Golan and Soytas (2014) for an application of this estimation techniques to a non-corporative collective model of household behavior. Let individual's gender, subscripted as σ, takes the value of m for a male and f for a female: $\sigma = \{f, m\}$. Gender is included in the vector of invariant characteristics x_σ. Let K describe the number of possible combinations of actions available to each household. Individuals get married at time 0, and for simplicity we assume that there is no divorce (see Gayle, Golan, and Soytas 2014 for application with marriage and divorce). Households are assumed to live for T periods and die together. Time zero is normalized to take account of the normal age gap within married couples, which would imply that men has a longer childhood than female. All individual variables and earnings are indexed by a the gender subscript σ. We omit the gender subscript when a variable refers to the household (both spouses). The state variables are extended to include the gender of the offspring. Let the vector ζ_t indicate the gender of a child born at age t, where $\zeta_t = 1$ if the child is a female and $\zeta_t = 0$ otherwise. The vector of state variables is expanded to include the gender of the offspring:

$$z_t = (\{I_{k1}\}_{k=0}^K, ..., \{I_{kt-1}\}_{k=0}^K, \zeta_0, .., \zeta_{t-1}, x_f, x_m).$$

We assume households invest time and money in the children in the household. The function $\omega_{\sigma t}(z_t, h_{\sigma t})$ denotes the earnings function; the only difference from the single agent problem is that gender is included in z_t and can thus affect wages. The total earnings is the sum of individual earnings as $\omega_t(z_t, h_t) = \omega_{1t}(z_t, h_{ft}) + \omega_{2t}(z_t, h_{mt})$ where $h_t = (h_{ft}, h_{mt})$. The educational outcome of the parents offsprings is mapped from the same parental inputs as the single agent model: income and time investment, number of older and younger siblings, and parents' characteristics such as education, race, and labor market skill. In the extension gender is also included as a parental characteristic. Thus the production function is still denoted by $M(x'|z_{T+1})$ where z_{T+1} represents the state variables at the end of the parents' life-cycle, T.

In the household, the total per period expenditures cannot exceed the combined income of the spouses. The budget constraint for the household is given by

$$w_t \geq c_t + \alpha_{Nc}(z_t)(N_t+b_t)w_t(z_t, h_t) \tag{35}$$

The right hand side represents expenditures on personal consumption of the parents, c_t, and on children. Parents pay for the children living in their household, regardless of the biological relationship, and do not transfer money to any biological children living outside of the household.

As in the single agent model, we can eliminate the continuous choice in the lifetime utility problem so that household face a purely discrete choice problem. Recall that the budget constraint for the household, assuming no borrowing or saving, is:

$$w_t(z_t, h_t) - \alpha_N(z_t)(N_t+b_t)w_t(z_t, h_t) = c_t \tag{36}$$

and, as in the single agent problem, we may substitute for consumption in u_2 and obtain the following household utility function:

$$u_{kt}(z_t) = \theta_k(z_t) + u_t[w_t(z_t, h_t)(1 - \alpha_N(z_t)(N_t+b_t)), z_t]. \tag{37}$$

For notation simplicity let $x_f, \in \{f\}_{f=1}^{F}$ and $x_m \in \{m\}_{m=1}^{M}$ and P_{fm} be the probability that type f female married type m make at age 0. We can then defined the expected lifetime utility for a type (f, m) household at age 0, excluding the dynastic component, as:

$$U_T(f, m) = E_0\left[\sum_{t=0}^{T} \beta^t \sum_{k=0}^{K} I_{kt}^0\{u_{kt}(z_t) + \varepsilon_{kt}\}\right], \tag{38}$$

and the expected lifetime utility for a type (f, m) household at age 0 as

$$U(f, m) = U_T(f, m) + \beta^T \lambda E_0\left[N^{-\nu} \sum_{n=1}^{N} \sum_{f'=1}^{F} \sum_{m'=1}^{M} P_{f'm'}U_n(f', m')| f, m\right]. \tag{39}$$

As in the single individual version of the model we can define the expected present discounted value of the lifetime utility of the household as any period t as

$$V(z_t, \varepsilon_t) = \max_I E_I\left(\sum_{s=t+1}^{T} \beta^{s-t} \sum_{k=0}^{K} I_{ks}[u_{ks}(z_s) + \varepsilon_{ks}] + \beta^{T-s}\lambda N^{-\nu} \sum_{n=1}^{N} \sum_{f'=1}^{F} \sum_{m'=1}^{M} P_{f'm'}U_n(f', m')|z_t, \varepsilon_t\right) \tag{40}$$

This can be written recursively as

$$V(z_t, \varepsilon_t) = \sum_{k=0}^{K} I_{kt}^o(z_t, \varepsilon_t)[u_{kt}(z_t) + \varepsilon_{kt}] + \beta \sum_z \int V(z, \varepsilon) f_\varepsilon(\varepsilon) d\varepsilon F(z|z_t, I_{kt}^o = 1)]$$

where $f_\varepsilon(\varepsilon)$ is the continuously differentiable density of $F_\varepsilon(\varepsilon_{0t}, .., \varepsilon_{17t})$, $F(z|z_t, I_{kt} = 1)$ is a transition function for state variables which is conditional on choice k, and $I_{kt}^o(z_t, \varepsilon_t)$ is optimal household decision rule.. Similar to Equation (48) we can define the conditional choice household probability as by $p_k(z_t) = E[I_{kt}^o = 1|z_t]$ and the ex ante value function as:

$$V(z_t) = \sum_{k=0}^{K} p_k(z_t)\left[u_{kt}(z_t) + E_\varepsilon[\varepsilon_{kt}|I_{kt} = 1, z_t] + \beta \sum_z V(z)F(z|z_t, I_{kt} = 1)\right]. \tag{41}$$

The rest of the estimation carries through as in the single individual case.

Discussion The addition of the two household members to the model captures important issues of the degree of specialization in housework and labor market work in household with different composition of education. The importance of which spouse spends time with children (and the levels of time) depends on the production function of education of children and whether time of spouses is complement or substitute. Furthermore, we capture patterns of assortative mating which may amplify the persistence of income across generations relative to a more random matching patterns. Since in our model there is potentially correlation of the cost of transfers to children (time input) with both parents' characteristics, assortative mating patterns imply that if children of more educated parents are more likely to be more educated, they are also more likely to have a more educated spouse which increases the family resources and their children educational outcomes.

5 Empirical Application

To illustrate the estimation method, we estimate the above model using a data set compiled from two sources, the Panel Study of Income Dynamics (PSID) and the American Time Use Survey (ATUS). The PSID is a longitudinal survey compiled from in person interviews between 1968 and 1972, and by telephone interview thereafter. The ATUS is a cross sectional survey compiled from telephone interviews conducted between 2003 and 2011. From the PSID collection we use data from the Family-Individual File of the Michigan PSID and from ATUS data we use the entire collection. The PSID provides a long panel of matched data on individuals labor market hours, earnings, housework hours, marriage, and childbirth histories, for overlapping cohorts and generations. Our initial sample from the PSID contains 423,631 individual-year observations. Where an observation was missing for a parent or spouse, the entire panel for that household was excluded. To select relevant data we began by creating a variable called "Relationship to Head [of household]" and setting the variable equal to either "head," "wife," "son,"or "daughter" based on survey responses. We further narrowed the sample to white and black individuals between the ages of 17 and 55, taking 17 as a lower bound for high school graduates and 55 as the upper bound for fertility decisions. We excluded anyone under 17 years of age as an eligible parent and we excluded individuals with less than 5 years of sequential observations because the earnings equation we plan to estimate requires a least 4 observed labor-market participation decisions. Finally we excluded all observations of parents whose children were older than 16 years in the first panel wave to ensure the data represents parental investment in a child's *early* life. These exclusions reduced the number of individual-year observations to 139,827, and produced a sample of panel data containing 12,051 individual males and 17,744 individual females, all of whom were observed for at least one year during our sample period.

A shortcoming of the PSID is that housework hours are recorded in aggregate, so the PSID not contain information on time spent on different household activities. The PSID does not provide data on time spent on child care or other kinds of housework. Time spent with children is estimated using a variation of the approach used in previous literature such as Hill and Stafford (1974, 1980), Leibowitz (1974), and Datcher-Loury (1988). Hours spent with children are computed as the deviation of average housework hours for parents with children from the average housework hours of individuals without children. We compute the average hours spent with children for each gender, education level, and age and set hours spent with children to zero where there is a negative value and when individuals have no children. We have benchmarked this variable with actual time spent with children from the American Time Use Survey (ATUS) and the basic patterns are similar. See the companion paper, Gayle, Golan and Soytas (2014), for more details on this comparison.

Table 1 presents the summary statistics for our sample; Column (1) summarizes the overall sample, Column (2) shows only data from parents, and Column (3) summarizes data of the their children. The first generation is on average 7 years older than the second generation. As a consequence, a higher proportion are married in the first generation. The male-female ratio is similar across generations (about 55 percent female), and this ratio is higher in our sample than in the general population because, females are more likely to maintain responsibility for children in cases of divorce. Our sample contains a higher proportion of blacks than the general population, which is consistent with PSID survey procedures, and the second generation has an even higher proportion of blacks than the first generation (about 29 percent in the second and 20 percent in the first generation) because of higher fertility rates among blacks in our sample. There are no significant differences across generations in completed years of education. The second generation in our sample has a lower average age than the first generation, so the second generation also has a lower marriage rate and a lower average for number of children, annual labor income, labor market hours, housework hours, and mean time spent with children. Our second generation sample spans the same age range, 17 to 55, as the first sample.

5.1 Empirical Implementation

This section describes the choice set specifications and functional forms of the model that we estimate. We assume that all individuals enter the first period of the life-cycle married. That is, they transition into a married household immediate after becoming adult. When individuals transition into a married household, their spouses' characteristics are drawn from the known matching function $G(x_{-\sigma} | x_\sigma)$. Since the matching function depends on the individual's state variables– it separately captures the effect of number of children and past actions that affect labor market experience for example, on the spouse's characteristics.

We set the number of an adult's periods in each generation to $T = 30$ and measure the individual's age where $t = 0$ is age 25. This is because at this age most individuals would have completed their education and started their family in the data. As discussed above, we assume that parents receive utility from adult children,whose educational outcome is revealed at the last period of their life regardless of the birth date of the child. This assumption is similar to the Barro-Becker assumptions. We avoid situations where the outcome of an older child is revealed while parents make fertility and time investment decisions to ensure that (i) these decisions are not affected by adult child outcomes, and (ii) that adult children's behavior and choices do not affect investment in children and fertility of the parents, in which case solutions to the problems are significantly more complicated and it is not clear whether a solution exists.

The three levels of labor supply correspond to working 40 hours a week; an individual working fewer than three hours per week is classified as not working, individuals working between 3 and 20 hours per week are classified as working part-time, while individuals working more than 20 hours per week are classified as working full-time. There are three levels of parental time spent with children corresponding to no time, low time, and high time. To control for the fact that females spend significantly more time with children than males, we use a gender-specific categorization. We use the 50th percentile of the distribution of parental time spent with children as the threshold for low versus high parental time with children, and the third category is 0 time with children. This classification is done separately for males and females. Finally, birth is a binary variable; it equals 1 if the mother gives birth in that year and 0 otherwise. Therefore the household choice are a combination of labor supply and time with children for males and females in household plus the birth decision.

Labor Market Earnings An individual's earnings depend on the subset of his or her characteristics, $z_{\sigma t}$. These include age, age squared, and dummy variables indicating whether the individual has high school, some college, or college (or more) education interacted with age respectively; the omitted category is less than high school. Let η_σ be the individual-specific ability, which is assumed to be correlated with the individual-specific time-invariant observed characteristics. Earnings are assumed to be the marginal productivity of workers and are assumed to be exogenous, linear additive, and separable across individuals in the economy. The earnings equations are given by

$$w_{\sigma t} = \exp(\delta_{0\sigma} z_{\sigma t} + \sum_{s=0}^{\rho} \delta_{\sigma,s}^{pt} \sum_{k_{t-s} \in \mathcal{H}_{P\sigma}} I_{k_{t-s}\sigma} + \sum_{s=1}^{\rho} \delta_{\sigma,s}^{ft} \sum_{k_{t-s} \in \mathcal{H}_{F\sigma}} I_{k_{t-s}\sigma} + \eta_\sigma) \tag{42}$$

where $\mathcal{H}_{P\sigma}$ and $\mathcal{H}_{F\sigma}$ are the set of choices for part-time and full-time work, respectively. Therefore, the earnings equation depends on experience accumulated while working part-time and full-time and the current level of labor supply. Thus, $\delta_{\sigma,s}^{pt}$ and $\delta_{\sigma,s}^{ft}$ capture the depreciation of the value of human capital accumulated while working part-time and full time, respectively. In the estimation we assume $\rho = 4$ given that the effect of experience with higher lags is insignificant (Gayle and Golan, 2012; Gayle and Miller, 2013).

Production function of children We assume that race is transmitted automatically to children and rule out interracial marriages and fertility. This is done because there is insufficient interracial births in our sample to study this problem. Therefore, parental home hours when the child is young affect the future educational outcome of the child, which is denoted by Ed'_σ [8], and innate ability, η'_σ, both of which affect the child's earnings (see equation 42).The state vector for the child in the first period of the life-cycle is determined by the intergenerational state transition function $M(x'|z_{T+1})$; specifically, we assume that

$$M(x'|z_{T+1}) = \left[\Pr(\eta'_\sigma \mid Ed'_\sigma), 1\right] \Pr(Ed'_\sigma \mid z_{T+1}) \tag{43}$$

Thus, we assume that the parental inputs and characteristics (parental education and fixed effects) determine educational outcomes according to the probability distribution $\Pr(Ed'_\sigma \mid z_{T+1})$. In our empirical specification the state vector of inputs, z_{T+1}, contains the parental characteristics, the cumulative investment variables (low time and high time) of each parent up to period T, the permanent income of each parent, and the number of siblings..In the data, we observe only total time devoted to children each period; thus, we assign each child age 5 or younger in the household the average time investment, assuming all young children in the household receive the same time input. Parental characteristics include the education of the father and mother, their individual-specific effects, and race. Once the education level is determined, it is assumed that the ability η'_σ is determined according to the probability distribution $\Pr(\eta'_\sigma \mid Ed'_\sigma)$. The above form of the transition allows us to estimate the equations separately for the production function of children given as the first two probabilities and the marriage market matching given as the last term.

Contemporaneous Utility We assume that the per-period utility from consumption is linear; therefore, Equation (37) the utility for a single parent utility from consumption and children (after substituting the budget constraint), becomes

$$u_{kt}(z_t) = \theta_k(z_t) + \alpha w_t(z_t, h_t) - \alpha\alpha_N(z_t)(N_t + b_t) \tag{44}$$

where $\theta_k(z_t)$ are the coefficients associated with each combination of time allocation choice, thus capturing the differences in the value of non-pecuniary benefits/costs associated with the different activities. The vector of decisions includes birth; thus, we allow the utility associated with different time allocations to depend on whether there is a birth or not. As discussed earlier, this utility captures not only the level of leisure but also the non-pecuniary costs/benefits associated with the different activities; for example, we do not rule out that time spent with children may be valued and that the non-pecuniary costs/benefits depend on birth events and levels of labor supply.

We assume no borrowing and saving, one consumption good with price normalized to 1, and risk neutrality. The first term represents the utility from own consumption. The second term, however, represents the net utility/cost from having young children in the household. In general, given our assumptions, we can use a budget constraint to derive the coefficients on income and number of children and a separate, non-pecuniary utility from children and monetary costs. However, since we do not have data on consumption or expenditures on children, the coefficients on the number of children also capture non-pecuniary utility from children and cannot be identified separately from the monetary costs of raising children. The interaction of income with the number of children and education captures differences in the cost of raising children by the socioeconomic status of parents. By assuming a linear utility function, we abstract from risk aversion and insurance considerations that may affect investment in children, fertility, as well as the labor supply. For families, we ignore the insurance aspects of marriage and divorce. While these issues are potentially important, we abstract from them and focus on transmission of human capital. The no borrowing and savings assumption is extreme and allows us to test whether (i) income is important in the production function of education of children, and (ii) whether the timing of income is important.

[8] Level of education, Ed_σ, is a discrete random variable in the model where it can take 4 different values: less than high school (LHS), high school (HS), some college (SC), and college (COL).

5.2 Empirical Results

This section presents results of estimation and analysis of the structural model. First, we present estimates from Step 1 of our estimation procedure. Second, we present estimates from Step 2 of the estimation. Third, we present results that assess how well our model fits the data. Finally, we present counterfactual value of each type of household, which can be interpreted also as the return to parental investment in children; the valuation function of the children includes the value of their education, earnings, as well as the spouse they married and his or her income.

5.3 First stage estimation

The first stage estimates include estimates of the earnings equation, the unobserved skills function, the intergenerational education production function, and the marriage assignment functions. All these functions are fundamental parameters of our model which are estimated outside the main estimation of the preference, discounts factors, and the net costs of raising children parameters. The first stage estimates also include equilibrium objects such as the conditional choice probabilities. Below we present estimates on the main earnings equation, the unobserved skills function, the intergenerational education production function. The estimates of the marriage assignment functions and the conditional choice probabilities are included in a supplementary appendix.

Earnings equation and unobserved skills Table 3 presents the estimates of the earnings equation and the function of unobserved (to the econometrician) individual skill (see also Gayle, Golan and Soytas, 2014). The top panel of the first column shows that the age-earnings profile is significantly steeper for higher levels of completed education; the slope of the age-log-earnings profile for a college graduate is about 3 times that of an individual with less than a high school education. However, the largest gap is for college graduates; the age-log-earnings profile for a college graduate is about twice that of an individual with only some college. These results confirm that there are significant returns to parental time investment in children in terms of the labor market because parental investment significantly increases the likelihood of higher education outcomes, which significantly increases lifetime labor market earnings.

The bottom panel of the first column and the second column of Table 4 show that full-time workers earn 2.6 times more than part- time workers for males, and 2.3 times more than part time workers for females (see also Gayle, Golan and Soytas (2014)). It also shows that there are significant returns to past full-time employment for both genders; however, females have higher returns to full-time labor market experience than males. The same is not true for part-time labor market experience; males' earnings are lower if they worked part time in the past while there are positive returns to the most recent female part-time experience. However, part-time experiences 2 and 3 years in the past are associated with lower earnings for females; these rates of reduction in earnings are, however, lower than those of males. These results are similar to those in Gayle and Golan (2012) and perhaps reflect statistical discrimination in the labor market in which past labor market history affects beliefs of employers on workers' labor market attachment in the presence of hiring costs.[9] These results imply there are significant costs in the labor market in terms of the loss of human capital from spending time with children, if spending more time with children comes at the expense of working more in the labor market. This cost may be smaller for female than males because part-time work reduces compensation less for females than males. If a female works part-time for 3 years, for example, she loses significantly less human capital than a male working part-time for 3 years instead of full-time. This difference may give rise to females specializing in child care; this specialization comes from the labor market and production function of a child's outcome as is the current wisdom.

The unobserved skill (to the econometrician) is assumed to be a parametric function of the strictly exogenous time-invariant components of the individual variables. This assumption is used in other papers (such as those by

[9] These results are also consistent with part-time jobs difffering more than full-time jobs for males more than for females.

MaCurdy, 1981; Chamberlain, 1986; Nijman and Verbeek, 1992; Zabel, 1992; Newey, 1994; Altug and Miller, 1988); and Gayle and Viauroux, 2007). It allows us to introduce unobserved heterogeneity to the model while still maintaining the assumption on the discreteness of the state space of the dynamic programming problem needed to estimate the structural parameters from the dynastic model. The Hausman statistic shows that we cannot reject this correlated fixed effect specification. Column (3) of Table 4 presents the estimate of the skill as a function of unobserved characteristics; it shows that blacks and females have lower unobserved skill than whites and males. This could capture labor market discrimination. Education increases the level of the skill but it increases at a decreasing rate in the level of completed education. The rates of increase for blacks and females with some college and a college degree are higher than those of their white and male counterparts. This pattern is reversed for blacks and females with a high school diploma. Notice that the skill is another transmission mechanism through which parental time investment affects labor market earnings in addition to education.

Intergenerational education production function A well-known problem with the estimation of production functions is the simultaneity of the inputs (time spent with children and income). As is clear from the structural model, the intergenerational education production function suffers from a similar problem. However, because the output of the intergenerational education production (i.e., completed education level) is determined across generations while the inputs, such as parental time investment, are determined over the life-cycle of each generation, we can treat these inputs as predetermined and use instruments from within the system to estimate the production function.

Table 4 presents results of a Three Stage Least Squares estimation of the system of individual educational outcomes; the estimates of the two other stages are in the supplementary appendix. The system includes the linear probabilities of the education outcomes equation as well as the labor supply, income, and time spent with children equations. The estimation uses the mother's and father's labor market hours over the first 5 years of the child's life as well as linear and quadratic terms of the mother's and father's age on the child's fifth birthday as instruments. The estimation results show that controlling for all inputs, a child whose mother has a college education has a higher probability of obtaining at least some college education and a significantly lower probability of not graduating from high school relative to a child with a less-educated mother; while the probability of graduating from college is also larger, it is not statistically significant. If a child's father, however, has some college or college education the child has a higher probability of graduating from college. This is consistent with the findings of Rios-Rull and Sanchez-Marcus (2002).

We measure parental time investment as the sum of the parental time investment over the first 5 years of the child's life. The total time investment is a variable that ranges between 0 and 10 since low parental investment is coded as 1 and high parental investment is code as 2. The results in Table 5 show that while a mothers' time investment significantly increases the probability of a child graduating from college or having some college education, a father's time investment significantly increases the probability of the child graduating from high school or having some college education. These estimates suggest that while a mother's time investment increases the probability of a high educational outcome, a father's time investment truncates low educational outcome. However, time investment of both parents is productive in terms of their children's education outcomes. It is important to note that mothers' and fathers' hours spent with children are at different margins, with mothers providing significantly more hours than fathers. Thus, the magnitudes of the discrete levels of time investment of mothers and fathers are not directly comparable since what constitutes low and high investment differs across genders.

5.3.1 Second stage estimation

This section presents estimates of the intergenerational and intertemporal discount factors, the preference parameters, and child care cost parameters. Table 5 presents the discount factors. It shows that the intergenerational discount factor, λ, is 0.795. This implies that in the second to last period of the parent's life, a parent valuation of their child's utility is 79.5% of their own utility. The estimated value is in the same range of values obtained in the literature calibrating dynastic model (Rios-Rull and Sanchez-Marcos, 2002; Greenwood, Guner, and Knowles, 2003). However, these models do not include life-cycle. The estimated discount factor, β, is 0.81. The discount factor is smaller than typical calibrated values, however, few papers that estimate it find lower values (for example, Arcidiacono, Sieg, and Sloan, 2006, find it to be 0.8).[10] Lastly, the discount factor associated with the number children, υ, is 0.25. It implies that the marginal increase in value from the second child is 0.68 and of the third child is 0.60.

Table 5 also presents the marginal utility of income which is positive and increasing with number of children except for a household with a college graduate wife and husband with at least a high school education. This rationalized the negative education in education. Also husband's education decreases the marginal of income for families with children. The marginal utility of income for families with children is also lower for black families.

The right hand panel of Table 5 presents our estimates of the dis/utility from the household. As is usual in discrete choice models they are estimated relative to an outside choice with in case are both spouses not work, giving birth, or spending any time with young children. We also use an additive specification in which the cost of birth, work and time with children are additively separable. First, every labor supply choice of the household carries with it a disutility relative to the reference choice; the exception is for households in which both spouses works full time — which statistically is not different from no work, no birth and no time with children – and when the wife does not work and husband works full time. In the data if both spouses spend low time with children and there is no birth then both spouses are equally likely to be observed working full time as not working and hence the equal utility both set of choices. Second, there are no distinct patterns to utility from time with children; these estimates are highly nonlinear, perhaps reflecting that it is a mixture of leisure and disutility. However, giving birth provides has a positive utility. This implies that, although parents get utility from the quality of their children, they also get some instantaneous utility from a birth.

5.3.2 Model fit and Value of Different Household

In this section, we first assess the ability of our model to reproduce the basic stylized facts by race, gender, and marital status as summarized in Section 2. We assess how well our model predicts the choices of labor supply, home hours with young children, and birth. Our model is over-identified and passes the standard over-identifying restrictions J-test. In the estimation, the conditional choice probabilities are targeted; in the simulation we simulate a sample of individuals and determine whether the individuals in our simulated sample behave like the individuals in our data. In some regards, this exercise is equivalent to a graphical summary of our model's over-identification test. Next, we calculate the counterfactual value of different household types to see whether it can rationalize the observed marriage pattern in the data.

Table 6 presents the model's fit. The model matches the labor supply patterns between gender and across race well. While it also matches the variation across race and gender for parental time with children, the levels are not similar in all cases. Examining the birth decisions, the model produces the differences in birth rates across household of different race, but the underpredicts the fecundity of whites by about a half. This lower birth rate is partly rationalizes the lower time with children predicted by the model. Nevertheless, our empirical model specification is very parsimonious: We do not include race, education, or marital status in the preference parameters for the disutility/utility of the different choices. In addition, the only unobserved heterogeneity is estimated from the earnings equations. Still, the model performs well in replicating the data based primarily on the economic interactions embodied in it.

Next, we turn to the value of different household types; the results are presented in Table 7. It shows that there is not much difference between married households by race. The only exception is that a black male with at least some college education receives a negative lifetime utility from marrying a female with less than high

[10] We are not aware of dynastic models in which the time discount factor is estimated.

school education. While the qualitative pattern is similar for white couples, it is not as strong. Overall, households with more education have higher values. For whites, the maximum value is achieved in two college graduate households, while for blacks it is a college educated husband and a wife with some college.

6 Conclusion

This paper develops an estimation that partially overcomes this curse of dimensional by exploiting properties of the stationary equilibrium. It provides a framework to estimate a rich class of dynastic models which includes investment in children's human capital, monetary transfers, unitary households, endogenous fertility and a life-cycle within each generation. This is an extension of methods used in the literature for the estimation of non-dynastic model to the dynastic setting. This estimation technique makes this estimation and empirical assessment of proposed counterfactual policy reform feasible. The paper compares the performance of the proposed estimator to a nested fixed point estimator using simulations and provides estimation results from an application of intergenerational transmission of human capital. The application provides plausible estimates of the intergenerational discount factors and matching the data very well.

A Appendix

Proof of Proposition 1. Recall the conditional value function in Equation (Link to Model Section):

$$\upsilon_k(z_t) = u_{kt}(z_t) + \beta \sum_{z_{t+1}} V(z_{t+1}) F(z_{t+1}|z_t, I_{kt} = 1) \tag{45}$$

We begin by noting that

$$V(z_{t+1}) = \sum_{s=0}^{17} p_s(z_{t+1}) \left[u_{st+1}(z_{t+1}) + E_\varepsilon(\varepsilon_{st+1}|I_{st+1} = 1, z_{t+1}) + \beta \sum_{z_{t+2}} V(z_{t+2}) F(z_{t+2}|z_{t+1}, I_{st+1} = 1) \right] \tag{46}$$

Substituting 46 into $\sum_{z_{t+1}} V(z_{t+1}) F(z_{t+1}|z_t, I_{kt} = 1)$ gives:

$$\begin{aligned}\sum_{z_{t+1}} V(z_{t+1}) F(z_{t+1}|z_t, I_{kt} = 1) = &\sum_{z_{t+1}} \sum_{s=0}^{17} p_s(z_{t+1}) \left[u_{st+1}(z_{t+1}) + E_\varepsilon(\varepsilon_{st+1}|I_{st+1} = 1, z_{t+1}) \right] F(z_{t+1}|z_t, I_{kt} = 1) \\ &+\beta \sum_{z_{t+1}} \sum_{s=0}^{17} p_s(z_{t+1}) \left[\sum_{z_{t+2}} V(z_{t+2}) F(z_{t+2}|z_{t+1}, I_{st+1} = 1) \right] F(z_{t+1}|z_t, I_{kt} = 1)\end{aligned}$$

Substitute the above into Equation (45) gives:

$$\begin{aligned}\upsilon_k(z_t) = u_{kt}(z_t) &+ \beta \sum_{z_{t+1}} \sum_{s=0}^{17} p_s(z_{t+1}) \left[u_{st+1}(z_{t+1}) + E_\varepsilon(\varepsilon_{st+1}|I_{st+1} = 1, z_{t+1}) \right] F(z_{t+1}|z_t, I_{kt} = 1) \\ &+\beta^2 \sum_{z_{t+1}} \sum_{s=0}^{17} p_s(z_{t+1}) \left[\sum_{z_{t+2}} V(z_{t+2}) F(z_{t+2}|z_{t+1}, I_{st+1} = 1) \right] F(z_{t+1}|z_t, I_{kt} = 1)\end{aligned} \tag{47}$$

Similarly

$$V(z_{t+2}) = \sum_{r=0}^{17} p_r(z_{t+2}) \left[u_{rt+2}(z_{t+2}) + E_\varepsilon(\varepsilon_{rt+2}|I_{rt+2} = 1, z_{t+2}) + \beta \sum_{z+3} V(z_{t+3}) F(z_{t+3}|z_{t+2}, I_{rt+2} = 1) \right] \tag{48}$$

Then

$$\begin{aligned}&\sum_{z_{t+2}} V(z_{t+2}) F(z_{t+2}|z_{t+1}, I_{st+1} = 1) = \\ &\sum_{z_{t+2}} \sum_{r=0}^{17} p_r(z_{t+2}) \left[u_{rt+2}(z_{t+2}) + E_\varepsilon(\varepsilon_{rt+2}|I_{rt+2} = 1, z_{t+2}) \right] F(z_{t+2}|z_{t+1}, I_{rt+1} = 1) \\ &+\beta \sum_{z_{t+2}} \sum_{r=0}^{17} p_r(z_{t+2}) \left[\sum_{z+3} V(z_{t+3}) F(z_{t+3}|z_{t+2}, I_{rt+2} = 1) \right] F(z_{t+2}|z_{t+1}, I_{st+1} = 1)\end{aligned} \tag{49}$$

Substitute into (47) gives:

$$
\begin{aligned}
v_k(z_t) = u_{kt}(z_t) + \beta \textstyle\sum_{z_{t+1}} \sum_{s=0}^{17} p_s(z_{t+1}) \left[u_{st+1}(z_{t+1}) + E_\varepsilon(\varepsilon_{st+1}|I_{st+1}=1, z_{t+1})\right] F(z_{t+1}|z_t, I_{kt}=1) \\
+\beta^2 \textstyle\sum_{z_{t+1}} \sum_{s=0}^{17} p_s(z_{t+1}) \sum_{z_{t+2}} \sum_{r=0}^{17} p_r(z_{t+2}) \left[u_{rt+2}(z_{t+2}) + E_\varepsilon(\varepsilon_{rt+2}|I_{rt+2}=1, z_{t+2})\right] \\
\times F(z_{t+2}|z_{t+1}, I_{st+1}=1) F(z_{t+1}|z_t, I_{kt}=1) \\
+\beta^3 \textstyle\sum_{z_{t+1}} \sum_{s=0}^{17} p_s(z_{t+1}) \sum_{z_{t+2}} \sum_{r=0}^{17} p_r(z_{t+2}) \sum_{z+3} V(z_{t+3}) F(z_{t+3}|z_{t+2}, I_{rt+2}=1) \\
\times F(z_{t+2}|z_{t+1}, I_{st+1}=1) F(z_{t+1}|z_t, I_{kt}=1)
\end{aligned} \tag{50}
$$

WLOG we assume $t+3=T$ then :

$$
V(z_T, \varepsilon_T) = \max_I E\left(\textstyle\sum_{k=0}^{17} I_{kT}[u_{kT}(z_T) + \varepsilon_{kT} + \lambda N_k^{-\nu} \sum_{n=1}^{N_k} \sum_{x_n} U_{g+1,n}(x_n)]|z_T, \varepsilon_T\right)
$$

Now

$$
\begin{aligned}
V(z_T) &= \textstyle\int V(z_T, \varepsilon_T) f_\varepsilon(\varepsilon_T) d\varepsilon_T \\
&= \textstyle\int \max_I E\left(\sum_{j=0}^{17} I_{jT}[u_{jT}(z_T) + \varepsilon_{jT} + \lambda N_j^{-\nu} \sum_{n=1}^{N_j} \sum_{x_n} U_{g+1,n}(x_n)]|z_T, \varepsilon_T\right) f_\varepsilon(\varepsilon_T) d\varepsilon_T \\
&= \textstyle\sum_{j=0}^{17} p_j(z_T)[u_{kT}(z_T) + E_\varepsilon(\varepsilon_{jT}|z_T, I_{jT}=1) + \lambda N_j^{-\nu} \sum_{n=1}^{N_j} \sum_{x_n} U_{g+1,n}(x_n) M(x_n'|z_T, I_{jT}=1)]
\end{aligned} \tag{51}
$$

From the value function representation we know that $U_{g+1,n}(x_n) = V(x_n)$ therefore

$$
V(z_T) = \textstyle\sum_{j=0}^{17} p_j(z_T)[u_{jT}(z_T) + E_\varepsilon(\varepsilon_{jT}|z_T, I_{jT}=1] + \lambda N_j^{-\nu} \sum_{n=1}^{N_j} \sum_{x_n} V(x_n) M(x_n|z_T, I_{jT}=1)] \tag{52}
$$

Substitute the above into (50) and rearranging we get:

$$
\begin{aligned}
v_k(z_t) = u_{kt}(z_t) + \beta \textstyle\sum_{z_{t+1}} \sum_{s=0}^{17} p_s(z_{t+1}) \left[u_{st+1}(z_{t+1}) + E_\varepsilon(\varepsilon_{st+1}|I_{st+1}=1, z_{t+1})\right] F(z_{t+1}|z_t, I_{kt}=1) \\
+\beta^2 \textstyle\sum_{z_{t+2}} \sum_{r=0}^{17} p_r(z_{t+2}) \left[u_{rt+2}(z_{t+2}) + E_\varepsilon(\varepsilon_{rt+2}|I_{rt+2}=1, z_{t+2})\right] \\
\times \textstyle\sum_{z_{t+1}} \sum_{s=0}^{17} p_s(z_{t+1}) F(z_{t+2}|z_{t+1}, I_{st+1}=1) F(z_{t+1}|z_t, I_{kt}=1) \\
+\beta^3 \textstyle\sum_{z_T} \sum_{j=0}^{17} p_j(z_T)[u_{jT}(z_T) + E_\varepsilon[\varepsilon_{jT}|z_T, I_{jT}=1] \sum_{r=0}^{17} \sum_{z_{t+2}} p_r(z_{t+2}) F(z_{t+3}|z_{t+2}, I_{rt+2}=1) \\
\times \textstyle\sum_{s=0}^{17} \sum_{z_{t+1}} p_s(z_{t+1}) F(z_{t+2}|z_{t+1}, I_{st+1}=1) F(z_{t+1}|z_t, I_{kt}=1) \\
+\lambda\beta^3 \textstyle\sum_{z_T} \sum_{j=0}^{17} p_j(z_T) N_j^{-\nu} \sum_{n=1}^{N_j} \sum_{x_n} V(x_n') M(x_n'|z_T, I_{jT}=1) \sum_{z_{t+2}} \sum_{r=0}^{17} p_r(z_{t+2}) F(z_T|z_{t+2}, I_{rt+2}=1) \\
\times \textstyle\sum_{z_{t+1}} \sum_{s=0}^{17} p_s(z_{t+1}) F(z_{t+2}|z_{t+1}, I_{st+1}=1) F(z_{t+1}|z_t, I_{kt}=1)
\end{aligned} \tag{53}
$$

Using the definition of the optimal transition function the above simplifies to:

$$
\begin{aligned}
v_k(z_t) = u_{kt}(z_t) + \beta \textstyle\sum_{s=0}^{17} \sum_{z_{t+1}} p_s(z_{t+1}) \left[u_{st+1}(z_{t+1}) + E_\varepsilon[\varepsilon_{st+1}|I_{st+1}=1, z_{t+1}]\right] F^o(z_{t+1}|z_t, I_{kt}=1) \\
+\beta^2 \textstyle\sum_{s=0}^{17} \sum_{z_{t+2}} p_s(z_{t+2}) \left[u_{rs+2}(z_{t+2}) + E_\varepsilon[\varepsilon_{st+2}|I_{st+2}=1, z_{t+2}]\right] F^o(z_{t+2}|z_t, I_{kt}=1) \\
+\beta^3 \textstyle\sum_{s=0}^{17} \sum_{z_T} p_s(z_T)[u_{sT}(z_T) + E_\varepsilon[\varepsilon_{sT}|z_T, I_{sT}=1]] F^o(z_T|z_t, I_{kt}=1) \\
+\lambda\beta^3 \textstyle\sum_{s=0}^{17} \sum_{z_T} p_s(z_T) N_s^{-\nu} \sum_{n=1}^{N_s} \sum_{x_n} V(x_n) M(x_n|z_T, I_{sT}=1) F^o(z_T|z_t, I_{kt}=1)
\end{aligned} \tag{54}
$$

Under the assumption that parents are infertility in the final period of their life-cycle simplifies to:

$$
\begin{aligned}
v_k(z_t) = u_{kt}(z_t) + \beta \textstyle\sum_{s=0}^{17} \sum_{z_{t+1}} p_s(z_{t+1}) \left[u_{st+1}(z_{t+1}) + E_\varepsilon[\varepsilon_{st+1}|I_{st+1}=1, z_{t+1}]\right] F^o(z_{t+1}|z_t, I_{kt}=1) \\
+\beta^2 \textstyle\sum_{s=0}^{17} \sum_{z_{t+2}} p_s(z_{t+2}) \left[u_{rs+2}(z_{t+2}) + E_\varepsilon[\varepsilon_{st+2}|I_{st+2}=1, z_{t+2}]\right] F^o(z_{t+2}|z_t, I_{kt}=1) \\
+\beta^3 \textstyle\sum_{s=0}^{17} \sum_{z_T} p_s(z_T)[u_{sT}(z_T) + E_\varepsilon[\varepsilon_{sT}|z_T, I_{sT}=1]] F^o(z_T|z_t, I_{kt}=1) \\
+\lambda\beta^3 N^{-\nu} \textstyle\sum_{n=1}^{N} \sum_{x_n} V(x_n) \sum_{s=0}^{K_T} \sum_{z_T} M(x_n|z_T, I_{sT}=1) p_s(z_T) F^o(z_T|z_t, I_{kt}=1).
\end{aligned} \tag{55}
$$

■

Proof of Proposition 2. We first check the various boundedness requirements of Theorem 8.12 in Newey and McFadden (1994). By assumption A8(i), we have that $E[\left\|\overline{\xi}_d(Z,\theta,\psi)\right\|^2] < \infty$. It obvious by inspection that $\overline{\xi}_d(Z,\theta,\psi)$ is continuously differentiable in θ and by A8(ii–iii) that $E[\nabla_\theta\overline{\xi}_d(Z,\theta,\psi)] < \infty$. Additionally, $\nabla_{\psi\psi}\overline{\xi}_d(Z,\theta^o,\psi^o)$ is also bounded: $E[\left\|\nabla_{\psi\psi}\overline{\xi}_d(Z,\theta^o,\psi^o)\right\|] < \infty$. Second, consider a point-wise Taylor expansion for the j^{th} element of $\overline{\xi}_d(Z,\theta,\psi)$,

$$\begin{aligned}\overline{\xi}^j(Z,\psi) &= \overline{\xi}^j(Z,\psi^o) + \nabla_\psi\overline{\xi}^j(Z,\psi^o)(\psi(z)-\psi^o(z)) + (\psi(z)-\psi^0(z))'\nabla_{\psi\psi}\overline{\xi}^j(Z,\psi^o)(\psi(z)-\psi^o(z)) \\ &\quad + o(\left\|\psi(z)-\psi^o(z)\right\|^2),\end{aligned}$$

where the norm over ψ is the sup-norm. Next, note that

$$\begin{aligned}\left|\overline{\xi}^j(Z,\psi) - \overline{\xi}^j(Z,\psi_0)\nabla_\psi\overline{\xi}^j(Z,\psi^o)(\psi(z)-\psi^o(z))\right| &\le \left\|(\psi(z)-\psi^o(z))'\nabla_{\psi\psi}\overline{\xi}^j(Z,\psi^o)(\psi(z)-\psi^o(z))\right\| \\ &\quad + o(\left\|\psi(z)-\psi^o(z)\right\|^2) \\ &\le \left\|\psi-\psi^o\right\|^2\left\|\nabla_{\psi\psi}\overline{\xi}^j(Z,\psi^o)\right\| + o(\left\|\psi-\psi^o\right\|^2),\end{aligned}$$

using the triangle inequality and the Cauchy-Schwartz inequality. Therefore, for $\|\psi-\psi^o\|$ small enough,

$$\left|\overline{\xi}^j(Z,\psi) - \overline{\xi}^j(Z,\psi^o) - \nabla_\psi\overline{\xi}^j(Z,\psi_0)(\psi(z)-\psi^o(z))\right| \le \left\|\psi-\psi^o\right\|^2\left\|\nabla_{\psi\psi}\overline{\xi}^j(Z,\psi^o)\right\|.$$

So that

$$\begin{aligned}\left\|\overline{\xi}(Z,\psi) - \overline{\xi}(Z,\psi_0) - \nabla_\psi\overline{\xi}(Z,\psi^o)(\psi(z)-\psi^o(z))\right\| &\le \left\|\psi-\psi^o\right\|^2\left\|\nabla_{\psi\psi}\overline{\xi}(Z,\psi^o)\right\| \\ \left\|\overline{\xi}(Z,\psi) - \overline{\xi}(Z,\psi^o) - \nabla_\psi\overline{\xi}(Z,\psi^o)(\psi(z)-\psi^o(z))\right\| &\le \left\|\psi-\psi_0\right\|^2\left\|\nabla_{\psi\psi}\overline{\xi}(Z,\psi^o)\right\|\end{aligned}$$

Hence $\Gamma(Z,\psi-\psi^o) = \nabla_\psi\overline{\xi}(Z,\psi_0)(\psi(z)-\psi^o(z))$ and $\Psi(Z) = \left\|\nabla_{\psi\psi}\overline{\xi}(Z,\psi^o)\right\|$. It follows that both $\Gamma(Z,\psi-\psi^o)$ and $\Psi(Z)$ are bounded from the boundedness conditions established above. Next we establish the form of the influence function. Note that we have

$$\int\Gamma(Z,\psi)F_0(\mathrm{d}\omega) = \int f_z(z)E[\nabla_\psi\overline{\xi}(Z,\psi^o)\mid z]\psi(z)\,\mathrm{d}z = \int \upsilon(z)\psi(z),$$

where $\upsilon(z) = f_z(z)E[\nabla_\psi\overline{\xi}(Z,\psi_0)\mid z]$. So, by the arguments on page 2208 of Newey and McFadden (1994), we have the influence function for $\overline{\xi}(\omega,\psi^{(D)})$:

$$\Phi(z) = \upsilon(z) - E\left[\upsilon(z)\widetilde{I}\right] = f_z(z)E\left[\nabla_\psi\overline{\xi}(Z,\psi^o)\mid z\right] - E\left[f_z(z)E[\nabla_\psi\overline{\xi}(Z,\psi^o)\mid z]\widetilde{I}\right].$$

Again by the boundedness of $\nabla_\psi\overline{\xi}(Z,\psi_0)$, it follows that $\int\|\upsilon(z)\|\,\mathrm{d}z < \infty$. Finally Assumption A7 guarantees that the Jacobian term converges. ■

Proof of Proposition 3.

This results follows immediately by combining the results in Proposition 1, with integral over z_{t+1} the summation. ■

References

[1] Aguirregabiria, Victor. "The dynamics of markups and inventories in retailing firms." *The Review of Economic Studies* 66.2 (1999): 275-308.

[2] Aguirregabiria, Victor, and Pedro Mira. "Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models." *Econometrica* 70.4 (2002): 1519-1543.

[3] Altuğ, Sumru, and Robert A. Miller. "The effect of work experience on female wages and labour supply." *The Review of Economic Studies* 65.1 (1998): 45-85.

[4] Alvarez, Fernando. "Social mobility: the Barro–Becker children meet the Laitner–Loury dynasties." *Review of Economic Dynamics* 2.1 (1999): 65-103.

[5] Barro, Robert J. and Gary S. Becker, "A Reformulation of the Economic Theory of Fertility,"*Quarterly Journal of Economics 103(1) (*1988):1-25.

[6] Barro, Robert J., and Gary S. Becker. "Fertility choice in a model of economic growth." *Econometrica: journal of the Econometric Society* (1989): 481-501.

[7] Becker, Gary S., and Nigel Tomes. "Human Capital and the Rise and Fall of Families." Journal of Labor Economics (1986): S1-S39.

[8] Datcher-Loury, Linda. "Effects of mother's home time on children's schooling." T*he Review of Economics and Statistics* (1988): 367-373.

[9] De Nardi, Mariacristina. "Wealth inequality and intergenerational links." *The Review of Economic Studies* 71.3 (2004): 743-768.

[10] Gayle, George-Levi, and Limor Golan. "Estimating a Dynamic Adverse-Selection Model: Labour-Force Experience and the Changing Gender Earnings Gap 1968–1997." *The Review of Economic Studies* 79.1 (2012): 227-267.

[11] Gayle, George-Levi, Limor Golan, and Mehmet A. Soytas. "What accounts for the racial gap in time allocation and intergenerational transmission of human capital." Unpublished manuscript, Department of Economics, Washington University (2014a).

[12] Gayle, George-Levi, Limor Golan, and Mehmet A. Soytas. "Family Structure and Intergenerational Income Mobility." Unpublished manuscript, Department of Economics, Washington University (2014b).

[13] Heckman, James J., V. Joseph Hotz, and James R. Walker. "New evidence on the timing and spacing of births." *The American Economic Review* (1985): 179-184.

[14] Hill, C. Russell, and Frank P. Stafford. "Allocation of time to preschool children and educational opportunity." *Journal of Human Resources* (1974): 323-341.

[15] Hill, C. Russell, and Frank P. Stafford. "Parental care of children: Time diary estimates of quantity, predictability, and variety." *Journal of Human Resources* (1980): 219-239.

[16] Hotz, V. Joseph, and Robert A. Miller. "An empirical analysis of life cycle fertility and female labor supply." *Econometrica: Journal of the Econometric Society* (1988): 91-118.

[17] Hotz, V. Joseph, and Robert A. Miller. "Conditional choice probabilities and the estimation of dynamic models." *The Review of Economic Studies* 60.3 (1993): 497-529.

[18] Hotz, V. Joseph, Robert A. Miller, Seth Sanders, and Jeffrey Smith. "A simulation estimator for dynamic models of discrete choice." *The Review of Economic Studies* 61.2 (1994): 265-289.

[19] Laitner, John. "Random earnings differences, lifetime liquidity constraints, and altruistic intergenerational transfers." *Journal of Economic Theory* 58.2 (1992): 135-170.

[20] Leibowitz, Arleen. "Home Investments in Children." *The Journal of Political Economy* (1974): S111-S131.

[21] Loury, Glenn C. "Intergenerational transfers and the distribution of earnings." *Econometrica: Journal of the Econometric Society* (1981): 843-867.

[22] Miller, Robert A. "Job matching and occupational choice." *The Journal of Political Economy* (1984): 1086-1120.

[23] Newey, Whitney K. "A method of moments interpretation of sequential estimators." *Economics Letters* 14.2 (1984): 201-206.

[24] Newey, Whitney K. "The asymptotic variance of semiparametric estimators." *Econometrica: Journal of the Econometric Society* (1994): 1349-1382.

[25] Newey, Whitney K., and Daniel McFadden. "Large sample estimation and hypothesis testing." *Handbook of econometrics* 4 (1994): 2111-2245.

[26] Pakes, Ariel. "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks." *Econometrica: Journal of the Econometric Society* (1986): 755-784.

[27] Rust, John. "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher." *Econometrica: Journal of the Econometric Society* (1987): 999-1033.

[28] Wolpin, Kenneth I. "An estimable dynamic stochastic model of fertility and child mortality." *The Journal of Political Economy* (1984): 852-874.

TABLE 1 : SUMMARY STATISTICS
(Standard Deviation are in parentheses)

	Total sample		Parents		Children	
Variable	N	Mean	N	Mean	N	Mean
Female	115,280	0.545	86,302	0.552	28,978	0.522
Black	115,280	0.223	86,302	0.202	28,978	0.286
Married	115,280	0.381	86,302	0.465	28,978	0.131
Age	115,280	26.155	86,302	27.968	28,978	20.756
		(7.699)		(7.872)		(3.511)
Education	115,280	13.438	86,302	13.516	28,978	13.209
		(2.103)		(2.138)		(1.981)
Number of children	115,280	0.616	86,302	(0.766)	28,978	0.167
		(0.961)		(1.028)		(0.507)
Annual labor income	114,871	16,115	86,137	19,552	28,734	5,811
		(24,622)		(26,273)		(14,591)
Annual labor market hours	114,899	915	86,185	1078	28,714	424
		(1041)		(1051)		(841)
Annual housework hours	66,573	714	58,564	(724)	8,009	641
		(578)		585		(524)
Annual time spent on children	115,249	191	86,275	234	28,974	63.584
		(432)		(468)		(259)
Number of individuals	12,318		6,813		5,505	

Source: Data from the Family-Individual File of the Michigan Panel Study of Income Dynamics (PSID), and include individuals surveyed between 1968 and 1997. Column (1) contains the summary statistics for the full sample; column (2) contains the summary statistics for the parents generation; column (3) contains the summary statistics of the off spring of the parents in column (2). Annual labor income is measured in 2005 dollars. Education measures year of completed education. There are less observations for annual housework hours than time spent on children because single individuals with no child are coded as missing for housework hours but by definition are set to zero for time spent on children

TABLE 2: SIMPLIFIED DISCRETE CHOICE MONTE CARLO SIMULATION RESULTS

	Pseudo Maximum Likelihood				Nested Fixed Point (ML)			
	sample size				sample size			
	1,000	10,000	20,000	40,000	1,000	10,000	20,000	40,000
$\theta = 0.25$								
Mean	0.24473	0.24935	0.24886	0.24881	0.22714	0.24571	0.23320	0.24477
Std. Dev.	0.04991	0.01328	0.00915	0.00668	0.04884	0.01354	0.02135	0.01019
Bias	-0.00527	-0.00065	-0.00114	-0.00119	-0.02286	-0.00429	-0.01680	-0.00523
M.S.E	0.00249	0.00017	0.00008	0.00005	0.00288	0.00020	0.00073	0.00013
$\lambda = 0.8$								
Mean	0.80425	0.79745	0.79797	0.79673	0.77538	0.78966	0.76934	0.78855
Std. Dev.	0.11241	0.03175	0.02157	0.01587	0.09211	0.03244	0.03656	0.02063
Bias	0.00425	-0.00255	-0.00203	-0.00327	-0.02462	-0.01034	-0.03066	-0.01145
M.S.E.	0.01253	0.00100	0.00046	0.00026	0.00901	0.00115	0.00226	0.00055
$\beta = 0.95$								
Mean	0.94208	0.95245	0.95037	0.95136	0.93441	0.95227	0.94603	0.95027
Std. Dev.	0.06276	0.01893	0.01301	0.00934	0.05322	0.01983	0.01820	0.01236
Bias	-0.00792	0.00245	0.00037	0.00136	-0.01559	0.00227	-0.00397	0.00027
M.S.E.	0.00396	0.00036	0.00017	0.00009	0.00305	0.00039	0.00034	0.00015
Avg Comp. time[5]	0.65	2.88	6.06	12.60	347.6	376.4	467.5	509.8

Note: Pseudo Maximum Likelihood (PML) corresponds to the estimation conducted by the new estimator using PML, and ML estimation is by the Nested Fixed Point (NFXP). All of the simulations are conducted using the GAUSS programming language on a 2 CPU 1.66 GHz, 3 GB RAM.laptop computer. Unit of time is seconds. The mean, empirical standard deviation, bias and mean squared error (M.S.E)of each parameter estimate are reported in the respective column for each sample size. The bias and the MSE are calculated relative to the original data generating value of the parameter. The data generating value of the parameter is also reported at the top left corner of summary statistics block for that parameter. (i.e., it is reported as $\theta = 0.25$ for the first parameter).

TABLE 3: ESTIMATES OF EARNINGS EQUATION: DEPENDENT VARIABLE: LOG OF YEARLY EARNINGS

(Standard Errors in Parenthesis)

Variable	Estimate	Variable	Estimate	Variable	Estimate
Demographic Variables				Fixed Effect	
Age Squared	-4.0e-4	Female x Full time work	-0.125	Black	-0.154
	(1.0e-5)		(0.010)		(0.009)
Age x LHS	0.037	Female x Full time work (t-1)	0.110	Female	-0.484
	(0.002)		(0.010)		(0.007)
Age x HS	0.041	Female x Full time work (t-2)	0.025	HS	0.136
	(0.001)		(0.010)		(0.005)
Age x SC	0.050	Female x Full time work (t-3)	0.010	SC	0.122
	(0.001)		(0.010)		(0.006)
Age x COL	0.096	Female x Full time work (t-4)	0.013	COL	0.044
	(0.001)		(0.010)		(0.006)
Current and Lags of Participation		Female x Part time work (t-1)	0.150	Black x HS	-0.029
Full time work	0.938		(0.010)		(0.010)
	(0.010)	Female x Part time work (t-2)	0.060	Black x SC	0.033
Full time work (t-1)	0.160		(0.010)		(0.008)
	(0.009)	Female x Part time work (t-3)	0.040	Black x COL	0.001
Full time work (t-2)	0.044		(0.010)		(0.011)
	(0.010)	Female x Part time work (t-4)	-0.002	Female x HS	-0.054
Full time work (t-3)	0.025		(0.010)		(0.008)
	(0.010)	Individual Specific Effects	Yes	Female x SC	0.049
Full time work (t-4)	0.040				(0.006)
	(0.010)			Female x COL	0.038
Part time work (t-1)	-0.087				(0.007)
	(0.010)			Constant	0.167
Part time work (t-2)	-0.077				(0.005)
	(0.010)				
Part time work (t-3)	-0.070				
	(0.010)				
Part time work (t-4)	-0.010	Hausman Statistics	2296		
	(0.010)	Hausman P-Value	0.000		
N			134,007		
Number of Individuals			14,018		
R-squared			0.44		0.278

Note: LHS is a dummy variable indicating that the individual has completed education of less than high school; HS is a dummy variable indicating that the individual has completed education of high school but college; SC is a dummy variable indicating that the individual has completed education of greater than high school but is not a college graduate; COL is a dummy variable indicating that the individual has completed education of at least a college graduate.

TABLE 4: 3SLS SYSTEM ESTIMATION THE EDUCATION PRODUCTION FUNCTION

(Standard Errors in parenthesis; Exclude class is Less than High School)

Variable	High School	Some College	College
High School Father	0.063	0.003	-0.002
	(0.032)	(0.052)	(0.0435
Some College Father	0.055	0.132	0.055
	(0.023)	(0.038)	(0.031)
College Father	-0.044	0.008	0.120
	(0.032)	(0.051)	(0.042)
High School Mother	0.089	0.081	-0.019
	(0.040)	(0.065)	(0.052)
Some College Mother	0.007	-0.041	0.017
	(0.030)	(0.049)	(0.039)
College Mother	0.083	0.120	0.040
	(0.036)	(0.057)	(0.047)
Mother's Time	-0.014	0.080	0.069
	(0.021)	(0.034)	(0.027)
Father's Time	0.031	0.100	0.026
	(0.019)	(0.029)	(0.025)
Mother's Labor Income	-0.025	-0.013	0.005
	(0.009)	(0.014)	(0.011)
Father's Labor Income	0.001	0.001	0.002
	(0.003)	(0.004)	(0.003)
Female	-0.002	0.135	0.085
	(0.017)	(0.028)	(0.022)
Black	0.020	0.082	0.043
	(0.039)	(0.063)	(0.051)
Number Siblings Under age 3	-0.014	-0.107	-0.043
	(0.017)	(0.027)	(0.022)
Number Siblings between age 3 and 6	-0.029	-0.047	-0.012
	(0.019)	(0.030)	(0.025)
Constant	0.855	-0.231	-0.359
	(0.108)	(0.172)]	(0.140)]
Observations	1335	1335	1335

Note: Data from the Family-Individual File of the Michigan Panel Study of Income Dynamics (PSID), and include individuals surveyed between 1968 and 1997. Instruments: Mother's and father's labor market hours over the child's first 8 years of life, linear and quadratic terms of mother's and fathers age when the child was 5 years old.

TABLE 5: STRUCTURAL ESTIMATES OF DISCOUNT FACTORS AND UTILITY PARAMETER

(Standard Errors in parenthesis)

Variable	Estimates	Variable		Estimates
Discount factors		Dis/Utility of Choices		
β	0.816	Wife	Husband	
	(0.002)	Labor Supply		
λ	0.795	No work,	Part time	-0.512
	(0.200)			(0.005)
υ	0.248	No work	Full time	0.207
	(0.168)			(0.009)
Marginal Utility of Income		Part time	No work	-2.023
Family labor Income	0.480			(0.003)
	(0.004)	Part time	Part time	-1.168
Children x Family Labor Income	1.216			(0.009)
	(0.065)	Part time	Full time	-0.605
Children x HS x Family Labor Income	1.279			(0.008)
	(0.066)	Full time	No work	-0.408
Children x SC x Family Labor Income	1.300			(0.007)
	(0.065)	Full time	Part time	-1.24532
Children x COL x Family Labor Income	-1.017			(0.011)
	(0.066)	Full time	Full time	0.001
Children x HS Spouse x Family Labor Income	-0.995			(0.010)
	(0.066)	Time with Kids		
Children x SC Spouse x Family Labor Income	-0.992	Low	Medium	0.502
	(0.066)			(0.014)
Children x COL Spouse x Family Labor Income	-0.466	Low	High	0.564
	(0.066)			(0.013)
Children x Black xFamily Labor Income	-0.108	Medium	Low	-0.169
	(0.004)			(0.008)
		Medium	Medium	0.129
				(0.010)
	()	Medium	High	0.593
				(0.013)
		High	Low	-0.364
				(0.007)
		High	Medium	0.353
				(0.011)
		High	High	-0.140
				(0.012)
		Birth		0.701
				(0.025)

Note: LHS is a dummy variable indicating that the individual has completed education of less than high school; HS is a dummy variable indicating that the individual has completed education of high school but college; SC is a dummy variable indicating that the individual has completed education of greater than high school but is not a college graduate; COL is a dummy variable indicating that the individual has completed education of at least a college graduate. Excluded choice is no work, no time with children and no birth for both spouses.

TABLE 6: MODEL FIT

LABOR SUPPLY			TIME WITH YOUNG CHILDREN			BIRTH		
WHITES								
Wife								
	Data	Model		Data	Model			
No work	0.2634	0.2599	Low	0.6363	0.8315			
Part time	0.1596	0.1622	Medium	0.2257	0.0531			
Full time	0.5770	0.5779	High	0.1380	0.0470			
							Data	Model
Husband						No birth	0.9014	0.9551
	Data	Model		Data	Model	Birth	0.0986	0.04493
No work	0.0290	0.0250	Low	0.8237	0.9592			
Part time	0.0306	0.0361	Medium	0.1008	0.0238			
Full time	0.9404	0.9390	High	0.0755	0.0170			
BLACKS								
Wife								
	Data	Model		Data	Model			
No work	0.1998	0.1309	Low	0.6837	0.9046			
Part time	0.1002	0.2150	Medium	0.2192	0.0497			
Full time	0.7000	0.6541	High	0.0971	0.0457			
							Data	Model
Husband						No birth	0.8955	0.9249
	Data	Model		Data	Model	Birth	0.1045	0.07507
No work	0.0640	0.0596	Low	0.8338	0.9729			
Part time	0.0423	0.0555	Medium	0.0744	0.0123			
Full time	0.8937	0.8850	High	0.0919	0.0148			

TABLE 7: VALUE OF HOUSEHOLDS BY RACE AND EDUCATION

Race	Husband Education	Wife Education	Lifetime Value
	LHS	LHS	10.2553
	LHS	HSH	16.9345
	LHS	SC	16.6506
	LHS	COL	19.0289
	HSH	LS	13.1308
	HSH	HSH	18.9846
	HSH	SC	21.2153
	HSH	COL	22.0866
Black	SC	LS	-10.4675
	SC	HSH	19.6639
	SC	SC	23.1898
	SC	COL	23.0041
	COL	LHS	-16.0114
	COL	HSH	22.8146
	COL	SC	25.8566
	COL	COL	24.8893
	LHS	LHS	9.2577
	LHS	HS	15.7034
	LHS	SC	16.5853
	LHS	COL	18.2363
	HS	LS	10.6104
	HS	HSH	17.4535
	HS	SC	16.2491
	HS	COL	19.2838
White	SC	LS	10.8613
	SC	HSH	18.0404
	SC	SC	16.6752
	SC	COL	20.4933
	COL	LHS	12.9575
	COL	HSH	13.8842
	COL	SC	21.1367
	COL	COL	23.2325

Note: LHS is a dummy variable indicating that the individual has completed education of less than high school; HS is a dummy variable indicating that the individual has completed education of high school but college; SC is a dummy variable indicating that the individual has completed education of greater than high school but is not a college graduate; COL is a dummy variable indicating that the individual has completed education of at least a college graduate.