A Preliminary Draft

Testing the Life Cycle Model of Consumption Using Information on Households without Children

Ekaterina Stepanova[1]

Version: August 2004

Abstract

This paper tests the life-cycle hypothesis of consumption using household data from the U.S. Consumer Expenditure Survey (1980-1998). The main puzzle posed by the data is that a household's consumption follows an inverted U-shape profile during the family's working life. This would seem to be inconsistent with optimal consumption smoothing behavior. We focus on demographic explanations (number and spacing of children over time) in the attempt to reconcile the theory with the data. Specifically, we work with a sample of families who do not have children currently present in the household – an approach suggested by Browning et al. (2002) - and "extract" the "never have children" households to question whether the consumption of these families has a non-linearity relationship with age. Demographically adjusted, consumption loses its age-dependence non-linearities supporting a demographic explanation within the life-cycle framework. This means that a household's consumption behavior can be viewed as been intertemporally consistent.

JEL Classification: D91, J13, D12
Keywords: consumption, life-cycle model, Consumer Expenditure Survey

---

I wish to thank and acknowledge the guidance of my dissertation committee members Professor Shelly Lundberg, Professor Dick Startz and Professor Eric Zivot in undertaking this project. Also I would like to thank Barb Best, Michael Emanuel, Hwang-Ruey Song, Luigi Ventura, Professor Charles Nelson and the participants of the labor and development economics brownbag at the University of Washington for helpful discussions. I am grateful for financial support from Center for Studies in Demography and Ecology at the University of Washington.

[1] Doctoral Student, University of Washington. Please direct any questions or comments to Ekaterina Stepanova, Dept of Economics, Box 353330, University of Washington, Seattle, WA 98195, e-mail: estep@u.washington.edu

Testing the Life Cycle Model of Consumption Using Information on Households without Children

I. Introduction

The life cycle model serves as a basic framework for analyzing the intertemporal behavior over the whole period of life as well as at particular intervals during the life. Applied to consumption, the life-cycle principle states that people should consume in such a way as to keep their expected marginal utility of consumption equal in each period of their life.

Both macro and micro data have been used to test the validity of this theoretical result. Most of the consumption paths constructed from the data (independent on the level of aggregation) exhibit an inverted U-shape during the life time (Browning and Crossley 2002, Fernandez and Krueger 2002) which is apparently inconsistent with the main prediction of the model. To make the model look more realistic, several modifications have been introduced including precautionary saving model, model with liquidity constraint, buffer-stock model (Caroll 1996, 1998; Gourinchas 2002) -- to name a few. However, there is no consensus so far as to whether people's consumption behavior is optimal in the intertemporal context.

Understanding the regularities of consumption behavior is important from a policy standpoint. As consumption is the opposite side of savings, knowing how people plan their consumption tells a lot about how they save over the life cycle and thus serves as an important consideration for tax policy debates.

This paper has its goal to test the life cycle hypothesis of consumption using U.S. household data – Consumer Expenditure Survey (CEX). The main hypothesis that is put forward is to explain the shape of the consumption profile over working life by considering demographics as the main factor. Intuitively, demographic variables such as number of children and their spacing over time seem to be an important element in determining the marginal parental utility of spending and, therefore, might explain the bump in the consumption in the mid-life consistently with household optimizing behavior. How is it possible? If people's consumption behavior is intertemporary consistent then they will want to provide same level of satisfaction from consumption in all periods of life independent on whether children present in the household. Therefore, if child is expected in the household then parents will deliberately have lower consumption before child arrives in order to provide the same level of consumption when child is present. This is theoretical explanation for the inverted U-shape consumption profile which we have goal to test using available data (CEX).

The problem with all research on consumption that involves intertemporal component is that most of the time expenditure data is available in repeated cross section format which doesn't allow tracing individuals over time. The most obvious solution for this problem is to implement a synthetic cohort approach where each cohort is constructed based on a year of birth of one of the members of the household. In this paper we follow this convenient technique and work with age-birth mean observations on each cohort.

In testing for the demographic explanation we implement approach proposed by Browning and Ejrnaes (2002) which is the following. If demographic variables related to children offer a good enough explanation for the non-smooth consumption then the following must be true: families who never have children in their life should not have any non-linearities (in age) in their consumption profiles, or putting it differently there should not be observed any inverted U-shape in consumption as "no-child" household ages. So the main hypothesis is "turned" in such a way that we are testing whether families who do not have children in their life have a flat consumption profile. Browning and Ejrnaes (2002) used British Family Expenditure Survey in their study. Their study confirms that demographics is a good explanation for the inverted U-shape in consumption since adjusted for presence and number of children consumption profile has no non-linear effects in age. We replicate this approach but apply it to the data from the U.S. Consumer Expenditure Survey.

This paper is organized as follows. First we look at the implications of the life cycle hypothesis and the modeling of demographics in synthetic cohort setting (chapter 2). In chapter 3 the data set is presented and the construction of the synthetic cohorts is outlined. The sampling scheme for "never have children" households is described in detail in chapter 4. The results of the analysis and potential biases that arise from this method of control are discussed in chapter 5, while chapter 6 summarizes the main conclusions.

II. Modeling approaches to incorporate demographic information

The life cycle hypothesis provides a convenient framework to analyze the intertemporal economic behavior of the individuals. Therefore, predictions of the model are based on the optimal conditions of the utility maximization problem. In order to test the life cycle hypothesis using demographic variables we need first of all to derive the result of optimization in such a form which could be tested with available data including demographics. The following section shows how reduced form equation can be derived from the optimization problem which includes demographic component.

1. Individual optimization model

We start with the problem of maximization of individual utility. To provide a more precise notation, instead of the individual we consider a household as the unit of analysis. From now on by household we mean a married couple if otherwise is not explicitly specified.

Suppose that each household is maximizing its expected utility over its life time subject to its budget constraint. Assume that sub-utility functions are intertemporally additive and that households form rational expectations of their fertility. This means that at each period of time each household knows how many children it is going to have and what the distribution of children is over time.

Following Deaton (1992) maximization problem becomes:

$$\begin{aligned} &Max\ \mathrm{U_t} = u_h(C_t) + \mathrm{E_t}\left(\sum_{t+1}^{T}\left(\frac{1}{1+\beta}\right)^{r} u_h(C_\tau)\right) \\ &u''(C) < 0 \qquad\qquad (1) \\ &s.t. \quad A_{t+1} = (1+r)(A_t + y_t - C_t) \end{aligned}$$

where C is consumption, β is the rate of time discounting, A_t denotes assets at time t; r is the interest rate.

Assume now that sub-utility function is convex and that it depends not only on consumption of a private good but also on a demographic component. By the demographic component we denote the number (presence) of children in the household at a given time t. Thinking of children as a public good suppose that the presence of children enters into the argument of the function exponentially:

$u_h(C,z) = v_h(Ce^{-\delta(h)z})e^{\delta(h)z}$, where C is consumption of private good, z is a dummy variable reflecting the presence of children in the household. In this expression $e^{-\delta(h)z}$ denotes child-response function. If $\delta(h)$ is positive, then when a child appears, the family has to increase the consumption of private goods (C) in order to keep marginal (parental) utility at the same level.

Assuming further a special case of convex utility such as iso-elastic utility, we get

$v_h(Ce^{-\delta(h)z}) = \dfrac{(Ce^{-\delta(h)z})^{(1-\theta)}}{1-\theta}$, where θ is the coefficient of relative risk aversion

(2)

First order condition for the problem (1) tells that marginal utility of consumption is a constant in any period. This leads to the Euler equation that states that expected marginal utility of consumption should be equal for all periods of life: $\dfrac{1+\beta}{1+r}E_t(u'(C_{t+1})) = u'(C_t)$

Under specification (2), the Euler equation is

$$(C_{h,t}e^{-\delta(h)z_{h,t}})^{-\theta} = \frac{1+r}{1+\beta}E_t\left[\left(C_{h,t+1}e^{-\delta(h)z_{h,t+1}}\right)^{-\theta}\right] \qquad (3)$$

This equation describes the optimal consumption path for the household h.

Applying Taylor series expansion to condition (3) we obtain[2]:

$$\begin{aligned} E_t\Delta\ln C_{t+1} &= \theta^{-1}(E_t\ln(1+r_{t+1}) - \ln(1+\beta)) + \delta_h\Delta z_{h,t+1} + \frac{1}{2}\theta\sigma_t^2 = \\ &= a_h + \delta_h\Delta z_{h,t+1} + \frac{1}{2}\theta\sigma_t^2 \end{aligned} \qquad (4)$$

[2] We linearize (3) around $\mathrm{X}(\Delta c, \Delta Z, r) = \mathrm{X}(0,0,0)$. We do linearization up to the second term.

where $a_h = \theta^{-1}(E_t \ln(1+r_{t+1}) - \ln(1+\beta))$

This equation says that the expected growth rate of consumption depends on the change in family size due to arrival/departure of children (second term on right hand side). In this expression the term σ_t^2 is a variance term that reflects one-period-ahead anticipated variability of the rate of growth of consumption[3]. Therefore, this last term on the right hand side is often taken as being responsible for precautionary saving behavior since the increase in future uncertainty expressed by σ_t^2 will induce higher savings in the preceding periods.

In our analysis we make a very strict assumption that the precautionary saving motive can be disregarded as an explanation of any consumption growth. Putting it differently, this assumption is equivalent to saying that change in demographic component is the only factor that can explain change in the consumption of the household.

Therefore,

$$E_t \Delta \ln C_{t+1} = a_h + \delta_h \Delta z_{h,t+1}$$

$$\Rightarrow \quad \Delta \ln C_{t+1} = a_h + \delta_h \Delta z_{h,t+1} + \varepsilon_{h,t+1} \qquad \varepsilon_{h,t+1} \sim iid(0, \delta_\varepsilon^2) \tag{5}$$

Equation (5) can serve as a testable version of the assumption that demographics alone can account for the change in consumption over the life cycle. If the precautionary savings motive was an important factor then there is an omitted variable in the equation (5). This omitted variable is part of the error term and, thus, it will cause residual (adjusted to demographics) consumption to exhibit non-linear effects with age. Therefore, testing residual consumption for non-linear effects would serve as a test of the importance of the precautionary savings explanation[4].

[3] It is assumed that there is no uncertainty regarding both real interest rate and appearance and the number of children , therefore in Taylor series expansion the variance terms related to the interest rate and "child" variable are zeros which leads to the expression in (4)

[4] There are two levels of testing the life-cycle hypothesis of consumption. First, is to test whether intertemporal behavior is optimal. Second level is to test one modification of the model versus another (demographic explanation versus precautionary saving explanation). In this paper we are testing at both levels simultaneously.

2. Intertemporal optimization model for cohorts

Unfortunately, most of the household surveys that ask questions on detailed consumption expenditures collect data in a repeated cross section format. This data constraint makes it impossible to observe household consumption over time, and therefore, to test equation (5) directly using individual level data. In the economic research traditional approach to deal with repeated cross section data is to construct synthetic cohorts which then could be traced over time. In this way the synthetic cohort data imitates a panel data.

A synthetic cohort combines people who have some similar characteristics. Most often the criteria of selection into cohort are year of birth or/and education level. From now on we assume that people are collected into cohort based on a year of birth. Even though we do not observe the same people over time, every year we can identify each individual with a particular cohort. By taking the averages characteristics over people who constitute the cohort in each year of the survey we are able to construct "observations" over time for a given synthetic cohort as it ages. Now instead of following observations of individual households we can treat cohort averages as individual observations.

Therefore, we can modify optimization result (5) for the individual household to get similar condition but now expressed in terms of synthetic cohort. If we take cohort means (expectations over cohorts) of each element in (5) then we get:

$$E^d(\Delta \ln C_{t+1}) = E^d(a_h) + E^d(\delta_h \Delta z_{h,t+1}) + E^d(\varepsilon_{h,t+1}) \Rightarrow$$
$$\Delta E^d(\ln C_{t+1}) = E^d(a_h) + E^d(\delta_h) E^d(\Delta z_{h,t+1}) + E^d(\varepsilon_{h,t+1})$$

where d stands for cohort d and assume that δ_h and $\Delta z_{h,t+1}$ are uncorrelated[5].

Note, that $E^d(\Delta z_{h,t+1})$ means the change in the proportion of families with children in cohort d. Denoting by $p(d,t)$ proportion of families with children for cohort d at age t means $E^d(\Delta z_{h,t+1}) = \Delta p_{d,t+1}$ and cohort version of Euler equation becomes:

$$\Delta c_{d,t+1} = a^d + \delta_d \Delta p_{d,t+1} + \varepsilon_{d,t+1}, \qquad (6)$$

[5] This assumption is rather strict because it implies that the degree to which parents consumption "react" on change in child status doesn't depend on the change in the proportion of families with children in the given cohort.

where c is *ln(C)* and subscript d is the mean over cohort d for age t.

This equation states that growth rate of cohort-mean consumption depends on the change in the cohort proportion of families with children currently present in the household. Thus, equation (6) provides the testable version of the main hypothesis that demographic changes account for non-linear shape of the consumption path. There are a couple of econometric issues related to the estimation of this equation. The most acute problem is endogeneity of $\Delta p_{d,t+1}$. This arises from the fact that there random effects not accounted by the model in (6) which may be correlated with change in the proportion of families with children -- $\Delta p_{d,t+1}$. Way to approach this problem is to implement the instrumental variables approach, this task is rather challenging in the synthetic cohort setting. Another problem is the measurement error in the variables which leads to the bias in the estimated coefficients. In there paper, Browning et al. (2002) used GMM technique to deal with this type of problem. The authors also suggested an alternative way of testing for the demographic explanation on which we focus in our replication.

Departing from the equation (6), we can state that the following is also true – for the sub-sample of the cohort consisting of families that never have children, $\Delta p_{d,t}$ is always zero. This means that the equation (6) for the "never have children" sub-sample is simply: $\Delta c_{d,t+1} = a^d + \varepsilon_{d,t+1}$ saying that growth rate of consumption depends only on a constant term a^d, where a^d reflects the rate of time discounting. Integrating this expression gives $c_{d,t}$, a consumption path for "never-have-child" families:

$$c_{d,t} = \gamma_d + a_d t + w_{d,t} \quad \text{where } w_{d,t} \sim iid(0, \delta_w^2)\text{, where } \gamma_d \text{ is a cohort fixed effect} \tag{7}$$

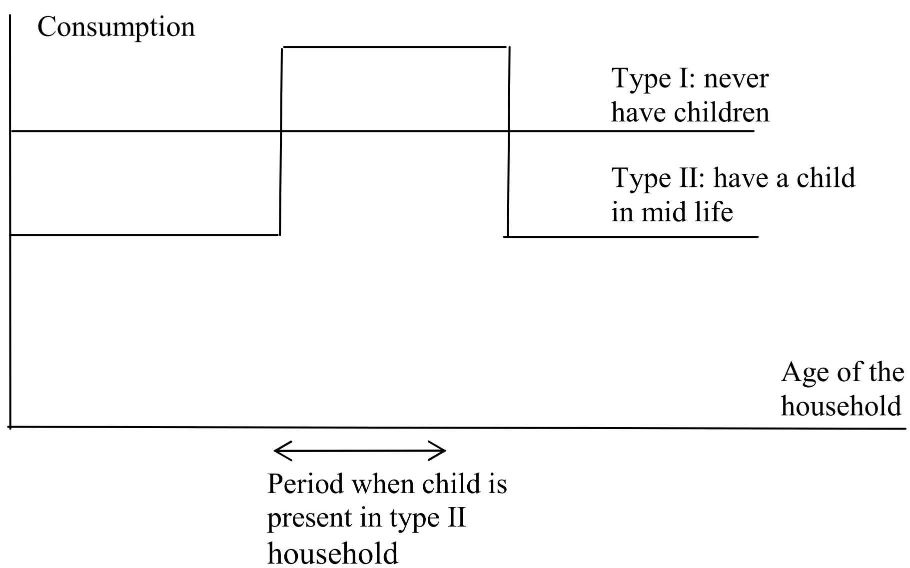
Derivations above suggest two ways of testing demographic version of the life cycle model based on synthetic cohort data. First way is outlined in equation (6) where we work with growth rates of consumption and the changes in the proportion of families with children. Second approach is to look at the households that never have children during their lives and see how their consumption evolves over time – this approach is summarized by equation (7).

3. Modified hypothesis for "never have children" sub-sample

The goal of this paper is to check whether change in demographic is a good explanation for inverted U-shape of consumption that occurs in the mid-life. As shown in the previous section one way to test the main hypothesis is by looking at the consumption of families who never have children – so from now on we will focus on this group of households. Equation (7) summarizes the testable hypothesis: consumption of "never have children" families should be linear in age, therefore no non-linear effects in age should be present.

To make this idea clearer, consider a simplifying version when rate of time discounting doesn't play a role[6]. Look at two types of household: a family that never has children (type I) and family that has one child some time in a middle life (type II). Also assume that their permanent incomes are the same. Given that expectations about fertility are rational, if we plot the individual consumption patterns for these two types of households they will look like this:

Figure 1:



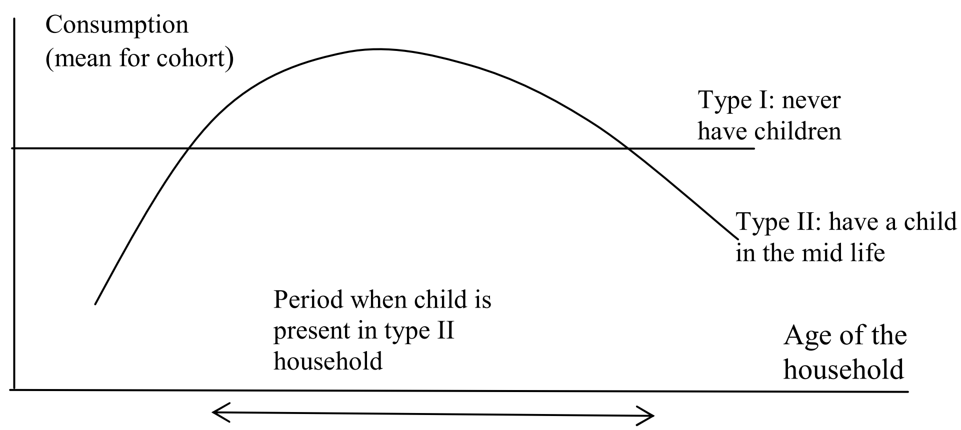
So the "never have children" type I household consumes each a fixed proportion out of its permanent income each period in order to keep the expected marginal utility of its consumption constant over the life time. Type II household that have the same permanent

[6] This is the case when $r = \beta$ all the time.

income as type I saves before the child is born, discretely increase consumption when child appears in the household and reduces its consumption as child leaves its parent's house. This is done in order to provide equal expected marginal utility (for parents) over the change in the child state. As a result, in all periods other than when the child is present in the household, the consumption of household II is less than the consumption of household I.

In the context of looking at the cohort averages of the "never have children" sub-sample, the consumption over the age of the household will be constant as well; while if we look at average consumption for the cohort sub-sample consisting of people who have at least one child in life, the consumption will have a "bump" somewhere in the middle of its life.

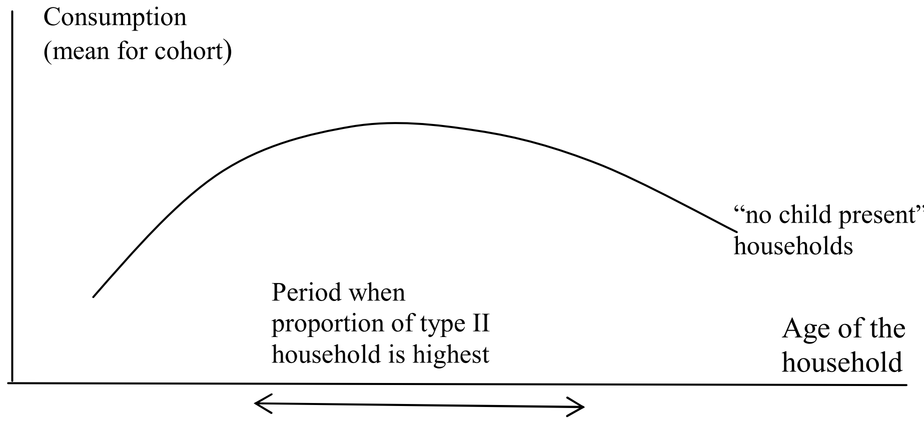
Figure 2:



Now assume that we observe only households who do not have child at a given moment of time but we can't tell whether each particular household would never have children in life, or has a child at some point in life. Assume that all households have the same permanent income. Then at young ages the average consumption of the cohort will be determined by the consumption of households who will have children because the proportion of this sub-group is relatively high. In the middle life the average consumption will be driven by the consumption of families who never have children because now the relative share of that type of households is relatively high. By the same logic, average consumption in older ages will be determined by the consumption of households with children not present in the family.

Although this idea is going to be formalized analytically in chapter 4, the graphical interpretation is presented below (Figure 3). It is clear that even in the sub-sample of "no child present" household with same permanent income, the average consumption path has an inverted U-shape in age.

Figure 3:



Keeping in mind that we would like to test cohort version of equation (7) for "never have children" households we will turn now to the data section which will help us to outline the empirical challenge of the testing procedure.

III. Data

1. Description of the data

In this work we use data from the U.S. Consumer Expenditure Survey (CEX). CEX collects information on various groups of expenditure, income, demographic and socio-economic characteristics of the American households. Data are collected in a series of 5 interviews and the household can't stay in the sample for more than 5 quarters. Each quarter a new random sample replaces households that quit the survey. Therefore, it is impossible to trace the same household over a substantial length of time. CEX survey covers the period from 1980 to 1998 (the last year when data are publicly available). Our analysis is mainly limited to married couples. However, information on single women who were previously married is used in assessing the completed fertility estimates[7].

We look at three consumption categories: total consumption, non-durables and food. All categories are deflated using personal consumption expenditures implicit price deflator.

2. Synthetic cohort construction

Consumer Expenditure Survey has a repeated cross section structure, for this reason we work with synthetic cohorts to replicate the panel. Since survey provides a random sample of the population each year, tracking through surveys the individuals born in the same year produces the series of random samples from the same cohort (Browning, Deaton, Irish 1985). Mean cohort behavior reproduces behavior of 'synthetic' individual, so the mean cohort observation can then be treated same way as individual observation.

We construct cohorts based on the year of birth of the wife. We use 3 years band in cohort definition. In our sample for the earliest cohort wife is born in 1942-1944, in the latest cohort wife is born 1960-1962 (see cohort chart in the Appendix B). We construct only 7 synthetic birth cohorts. Though data on other cohorts are available we are restricted in our analysis to cohorts that necessarily contain a wife of 38 years old – the reason due to the need to use this cross-section data in inferring the proportions in the "no-child present" sub-sample. Using the birth year of wife instead of husband's can be explained by the observation that number and age of children is more related to the age of the mother than of the father. This way we can achieve the construction of a homogenous cohort in terms of consumption behavior given there is a link between mother's age and number and spacing of children.

We do not separate the sample in groups according to the educational attainment because the cohort size is rather small before it is disaggregated into education groups. It is reasonable to think, however, that educational level influences consumption and

[7] Since in cross sectional data we cannot trace households over time, for the same reason we cannot observe whether childless household will ever have or have had children at some point in life. Completed fertility estimates will employed to infer about the probabilities of having children (in future or in the past) given that there are no children present at the time of the survey.

fertility behavior over the life cycle. More educated people may have steeper consumption profile in the middle of the life due to higher investment in children in future; also the peak in the consumption profile for these households may occur later in life compared to the less educated households. Therefore, there is a trade-off between maintaining a relatively big "cell" size and separating samples into educational groups.

In our analysis we work with two sub-sets of cohort data. One sub-sample consists only of "no child present households"[8]. The age restriction for the wife is between 20 and 55 years old, the upper bound is important for avoiding retirement age where consumption is governed by other forces. We also restrict on the cell size – the number of observations that constitute each cohort-age observation – to those with more than 10 observations. The "no child present" sample has 130 cohort-age observations and covers around 4,800 households.

Another sub-sample contains households at age 38 which includes all married couples as well as single women and daughters of the head of the household[9]. This sub-sample is the basis for calculating the distribution of age at first birth for the ages below 38. As a result the number of cross-section observations to estimate completed fertility is 23,013.

3. Graphical analysis of consumption profiles

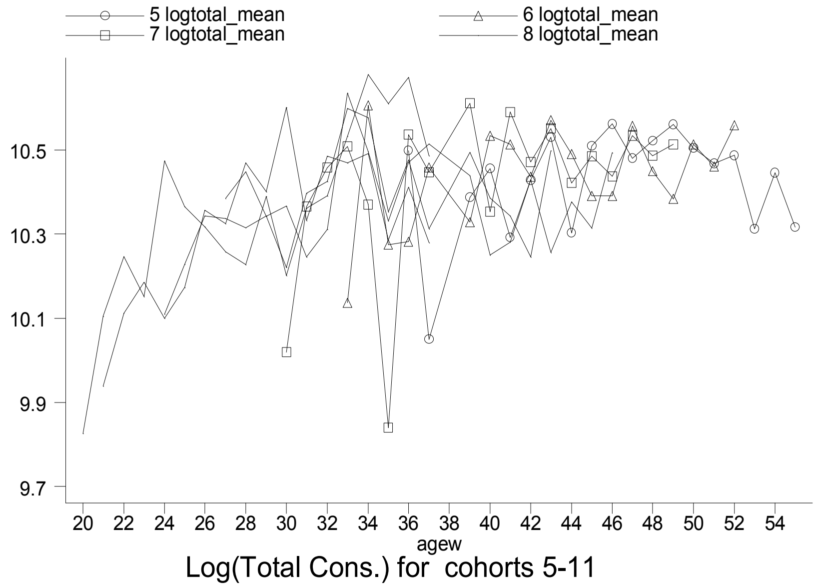
Graphs were constructed in levels and logs for the following consumption categories: total consumption, non-durable, food. Consumption is calculated for the sub-sample of

[8] This sub-sample forms the main body of data for the regression analysis (see regression equation (13) in chapter 4)

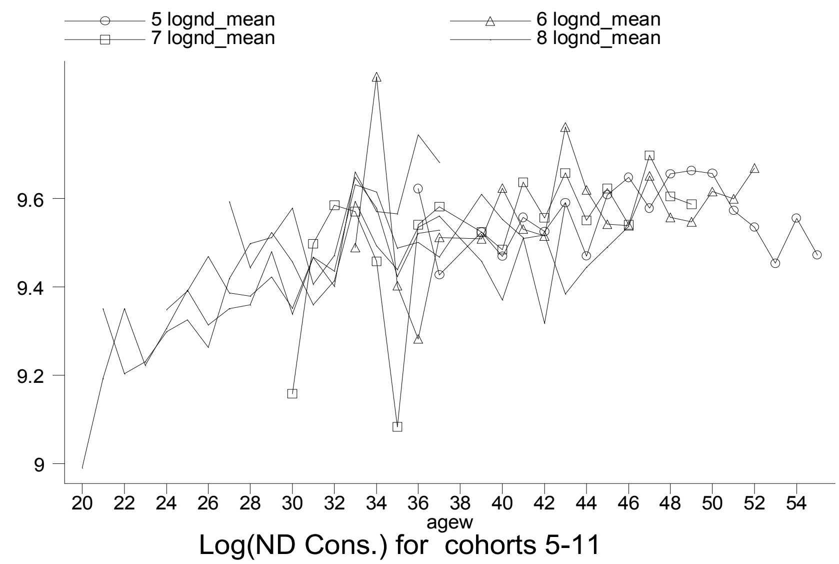
[9] This "extended" sub-sample serves as a reference group to compute the distribution of completed fertility. The logistics of the estimation of completed fertility will be described in chapter 4. In brief, this estimation is needed to identify the group of "never have children" households. In the reference group for completed fertility we deliberately include single women who were previously married as well as daughter of the head of the household who are either married or were married before. Adding them to married couples (with and without children) we want to use this augmented sample to estimate the probabilities of fertility of "j" children given that no children are currently present for the ages other than 38.

“no-child-present” households and consumption path exhibits inverted U-shape profile as the theory predicts[10]. All cohorts are presented in one graph.

<u>Graph for Log of Total Consumption:</u>



<u>Graph for Log of Non-Durable Consumption:</u>



As follows from the graphs, though consumption profiles have an inverted U-shape the curvature of the line is not very steep. Since for the purpose of our analysis we work with sub-sample of “no-child-present families” the observations on the families at the period when children are present in the household are excluded. This explains why the

[10] See Figure 3 in chapter 2 for details.

inverted U-shape is not that distinct as if we worked with the total sample of married couples.

It is also worth looking for a separate cohorts (see graphs in the Appendix C) from which we can see a more pronounced inverted U-shape. On all graphs, however, the steepness of the consumption profile is much more distinct early in life (consumption increases sharply) and less steep later in life (the decline is relatively small). There are several possible explanations for this finding. There is definitely a selection into the sample of the married households at early ages: more educated couples have higher permanent income and at the same time tend to marry later. Therefore, at early ages the sample of married households consists mainly of less educated households, which explains the steepness of consumption profile at this period in life. Though in the current analysis we focus on explanations that are based on demographic variables it is possible that precautionary saving and liquidity constraint motive can account for this shape (see discussion later in Appendix F).

IV. Extractions of "never have children" sub-sample in repeated cross section data

1. Decomposition of the cohort consumption for the "no child present" group.

Unfortunately, the use of repeated cross section data does not directly allow the observation of which household belong to the "never have children" type, since we can't observe the same households over time. For the same reason we are unable to distinguish directly "never have children" households as a sub-sample in a synthetic cohort. Some households who do not have children currently present may plan to have them in future (mainly younger households), or children might have left parents' homes already (older households).

Instead, we can observe a sub-sample of families who do not currently have children present. This sub-sample of "no children present" households includes two types of families:

-- type I : never have children

-- type II : have children in future or have had children

As we are interested in looking at consumption of type I—the "never have children" households—we need to think about the ways to "extract" this group from the "no children present" group that we can observe.

The idea of this section is to show how we can decompose the mean cohort consumption of "no child present" households in order to *identify* the consumption of the group of "never have children" households. The first step in doing this is to define the proportions of those two types of households in the "no child present" sample at each age of life, and then to use this proportions for the identification purposes. Notations that will help to state these proportions are the following. For each cohort (*d*) at each age (*t*) we denote:

p(d,t) as proportion of households in cohort *d* at time *t* with child present; and

π_d is the proportion of households in cohort d who have child at some time in life.

Now use *p(d,t)* and π_d to identify the proportions of two household types. In the "no child present" sub- sample at each age:

-- Proportion of "never have children" is $\dfrac{\text{proportion of HH with no children ever}}{\text{proportion of HH with no children presently}} = \dfrac{1-\pi_d}{1-p_{d,t}}$

-- Proportion of "have had or will have children" is

$$\frac{\text{proportion of HH with children in future or past}}{\text{proportion of HH with no children presently}} = \frac{\pi_d - p_{d,t}}{1-p_{d,t}} = m_{d,t} \quad (8)$$

Additionally, denote $E_I^d(c_{ht})$ - cohort mean of "never have" children

$E_{II}^d(c_{ht})$ - cohort mean of those who have a child at some time

$E_0^d(c_{ht})$ - cohort mean of "no child present" group[11]

[11] We can observe consumption of "no children present" group. However, since we do not observe households of type I and type II we are not able to observe directly $E_I^d(c_{ht})$ and $E_{II}^d(c_{ht})$

If one considers type I and type II households with same permanent income, then obviously:

$E_I^d(c_{ht}) > E_{II}^d(c_{ht})$ -- Earlier we have shown this intuitively based on graphical analysis[12]. As for the cohort means the logic is the same except that here we do not consider type II households in the period when they have children.

Because of this fact, under our hypothesis, the cohort mean of consumption for the sub-sample of "no child present" household will still have the inverted U-shape profile. The reason being that in this sub-sample the proportion of "never have children" household increases as the cohort approaches to the age of completed fertility. Therefore, in the middle of the life the sample mainly consists of "never have children" households. At the beginning of life and at the end of life the sample has a bigger proportion of type II households. Controlling for the permanent income, if we take the mean consumption over the "no child present" sample $E_O^d(c_{ht})$ the greater proportion of "never have children" households in the middle of life will cause an inverted U-shape of consumption profile.

The second step of decomposition is to present the cohort mean consumption of "no-child-present" group as the sum the cohort mean consumption of two types of households weighted according to the proportion indicated in (8):

$$E_O^d(c_{ht}) = \frac{1-\pi_d}{1-p_{dt}} E_I^d(c_{ht}) + \frac{\pi_d - p_{dt}}{1-p_{dt}} E_{II}^d(c_{ht})$$

Working through the math of this expression[13] at the end we will get the following decomposition result:

$$E_O^d(c_{ht}) = E_I^d(c_{ht}) - m_{d,t}\left[E_I^d(c_{ht}) - E_{II}^d(c_{ht})\right] \qquad (9)$$

Decomposition in equation (9) tells us that the mean consumption of the "no child present" sample is equal to the mean consumption of the households who never have children minus the difference between the mean consumption of type I and type II that is

[12] See Figures 1-3 in chapter 2 for details.

[13]

$$E_O^d(c_{ht}) = \frac{1-\pi_d}{1-p_{dt}} E_I^d(c_{ht}) + \frac{\pi_d - p_{dt}}{1-p_{dt}} E_{II}^d(c_{ht}) = E_I^d(c_{ht}) + \frac{\pi_d - p_{dt}}{1-p_{dt}}\left[E_{II}^d(c_{ht}) - E_I^d(c_{ht})\right] =$$
$$= E_I^d(c_{ht}) + m_{d,t}\left[E_{II}^d(c_{ht}) - E_I^d(c_{ht})\right] = E_I^d(c_{ht}) - m_{d,t}\left[E_I^d(c_{ht}) - E_{II}^d(c_{ht})\right]$$

weighed by the proportion of type II households in the sample. So at times when the proportion of type II (households that have child at some point in life) is big the mean consumption of the sample is driven down (earlier and later in life), while when $m_{d,t}$ is the smallest, the consumption is the highest (middle life). Another interpretation of the equation (9) is the following. In the middle period in life the cohort mean consumption of "no child present" households $E_0^d(c_{ht})$ is determined by the consumption of "never have children" households. This happens when the conditional probability term $m_{d,t}$ is small, and it is exactly the time in the middle of life—people who do not have children present would most probably belong to the group of household that never have children. In the early and late periods of life $m_{d,t}$ is relatively big measuring possibility of having children in the past or in the future for the families that presently are observed with no children. Intuitively, correction of the consumption introduced by $m_{d,t}$ makes a lot of sense: consumption of "no child present" group is very close to "never have child" sample in the middle of life and no correction (using term with $m_{d,t}$) is needed, so $m_{d,t}$ is small. Situation is exactly opposite in the early and late periods of life: there is s lot of correction required involving term $m_{d,t}$ to adjust for the possibility of children.

This decomposition suggests the way to "extract" the term $E_I^d(c_{ht})$. If we were able to estimate (9) as a regression equation treating $m_{d,t}$ as the unknown right hand side variable and $\left[E_I^d(c_{ht})-E_{II}^d(c_{ht})\right]$ as the estimation coefficient then $E_I^d(c_{ht})$ could be identified as a residual term of this estimation:

$$E_I^d(c_{ht}) = E_O^d(c_{ht}) - \left(E_{II}^d(c_{ht}) - E_I^d(c_{ht})\right)m_{d,t}$$

Now let's assess whether this estimation is possible. In order to do the estimation of (9) we need to make some identifying assumptions such as to treat $\left[E_I^d(c_{ht})-E_{II}^d(c_{ht})\right]$ as a constant (because estimation coefficient must be a constant). Since we assume that there is a certainty about number of children and their timing we can assume as well that[14]:

[14] For a more intuitive explanation see Figure 1.

$$\left(E^d_{II}(c_{ht}) - E^d_I(c_{ht})\right) = \tau_d + \eta_{d,t} \qquad \eta_{d,t} \sim N(0, \sigma^2_\eta) \tag{10}$$

A constant gap between consumption of type I and type II seems to be a plausible assumption for non-durable consumption expenditures. This is obviously not true for most durables, for example housing. Young households who plan to have children will on average increase investment in housing before the child is born, while for older households consumption will not drop as much immediately after a child leaves for work or college. In the last case the household will experience higher expenditure mainly due to college tuition costs. Therefore, testing the main hypothesis for non-durable consumption makes the most sense.

Recall Euler equation result (for "never have children" families) from (7): $E^d_I(c_{ht}) = a_d t + \gamma_d + \omega_{d,t}$. Using assumption (10) and expression in (7) we can re-write equation (9) in order to get[15]:

$$\begin{aligned} E^d_O(c_{ht}) &= E^d_I(c_{ht}) + m_{d,t}\left(E^d_{II}(c_{ht}) - E^d_I(c_{ht})\right) = \\ &= a_d t + \gamma_d + m_{d,t}\left(\tau_d + \eta_{d,t}\right) + \omega_{d,t} \end{aligned} \tag{11}$$

Equation (11) is the reduced form equation to be estimated in order to test the demographic hypothesis. As shown above, $E^d_I(c_{ht})$ can be identified as:

$$\hat{E}^d_I(c_{ht}) = E^d_O(c_{ht}) - m_{d,t}\left(E^d_{II}(c_{ht}) - E^d_I(c_{ht})\right) \tag{12}$$

According to the theory in (7) we expect that adjusting the consumption of the "no child present" sub-sample this way will produce a consumption profile of "never have child" households which is a linear function in age:

$$E^d_I(c_{ht}) = E^d_O(c_{ht}) - m_{d,t}\left(\tau_d + \eta_{d,t}\right) = a_d t + \gamma_d + \omega_{d,t}$$

Therefore, testing whether $E^d_I(c_{ht})$ is non-linear in age will serve as a test of demographic explanation for the life-cycle model.

There are several problems related to the estimation of equation (11). First of all, $m_{d,t}$, the proportion of families that will have or have had children given that children are not present, is not observed and needs to be estimated before the estimation of the main equation in (9). The method of estimation of $m_{d,t}$ is discussed in the next section;

15

however, it is obvious that $m_{d,t}$ will be measured with error which introduces "error-in-variable" problem into the equation (11). "Error-in-variable" approach leads to biased and inconsistent estimates of the coefficients, and therefore, this bias will be transferred into the adjusted consumption path. This problem needs to be treated as a separate econometric issue. Deaton (1985) has suggested the way to deal with this problem in the equation for the synthetic cohort. We plan to implement his technique in our estimation.

2. Estimation of the proportion of the families that will have or have had children given that children are not present in the household

As mentioned in the previous section, in order to estimate equation (11) we need to know $m_{d,t} = \frac{\pi_d - p_{d,t}}{1 - p_{d,t}}$ - proportion of families in cohort *d* at age *t* that have children at some point in life given there is no children currently present in the household. We can also think about $m_{d,age=t}$ as a conditional probability for household in cohort *d* of having children at the age of completed fertility[16] given that at age *t* no children are present in the household. Since we do not observe this conditional probability and can't compute it directly in cross section type of data we have to estimate it. This section discusses the ways of estimation. First, think about four possibilities: the family has no children, exactly one child, exactly two children and three children or more. In the "no child present" sub-sample each family belonging to cohort *d* faces a set of probabilities

$\left\{\hat{m}^0_{d,age},\ \hat{m}^1_{d,age},\ \hat{m}^2_{d,age},\ \hat{m}^3_{d,age}\right\}$, where

$m^j_{d,age=t} = \Pr(\text{"}j\text{"children at age of completed fertility}|\text{"0" children age age=t})$

Based of this the reduced form cohort equation (11) can be modified to include three conditional probabilities ("no child" is omitted):

$$c^0_{d,age} = a_1 * age + \gamma_d + \tau_1 * \hat{m}^1_{d,age} + \tau_2 * \hat{m}^2_{d,age} + \tau_3 * \hat{m}^3_{d,age} + \varepsilon_{d,age} \qquad (13)$$

To start with we want to estimate conditional probabilities $\left\{\hat{m}^0_{d,age},\ \hat{m}^1_{d,age},\ \hat{m}^2_{d,age},\ \hat{m}^3_{d,age}\right\}$ for each cohort *d* for every age *t*.

[16] The age of completed fertility is defined as age of woman after which she doesn't have children.

We will outline in brief the approach to estimation while the details can be found in Appendix A.

Two assumptions need to be made in order to make estimation of $m_{d,age=t}$ possible:

1 -- Completed fertility is observed when woman is 38 years old.

2 -- Number of children in the household can change only by one child per year.

We need the first assumption in order to have a "point of departure" in estimating the probabilities $\{\hat{m}^0_{d,age}, \hat{m}^1_{d,age}, \hat{m}^2_{d,age}, \hat{m}^3_{d,age}\}$. We use information on all households at age 38 to infer about conditional probabilities at other ages.

Under assumption 1, $m^j_{d,age=t} = \Pr("j" \text{ at age 38}|"0" \text{ children age age=t})$. Expanding this further we'll get[17]:

Pr ("j" children at age 38 | no children at given age=t) =

Pr (no children at given age=t | "j" children at age 38) * $\dfrac{\Pr("j" \text{ children in cohort d at 38})}{\Pr \text{(no children at age = t)}}$

While the last term in this product can be easily obtained from the data, the first term must be estimated. We divide the sub-sample into two groups: below and above 38 years old. The reason we are dividing the sample in this two groups is because we will use different approaches in estimating $\{\hat{m}^0_{d,age}, \hat{m}^1_{d,age}, \hat{m}^2_{d,age}, \hat{m}^3_{d,age}\}$. Since we observe how many children and their age when wife is 38 years old we can construct the distribution of age at first birth and this way infer probability (no children at given age=t | "j" children at age 38) for ages below 38. From this it is straightforward to compute Pr (no children at given age=t | "j" children at age 38). The task of inferring conditional probabilities is more complicated for the households above 38 years old. For this group we do not have the key piece of information similar to information on the distribution of age at first birth for ages below 38. In this second case children depart from the household and to compute the rate of departure (or transition from "j" children to "0" children over time) one need to estimate transition probability matrices. This is a less intuitive exercise; nevertheless, it serves its purpose. Details are given in the Appendix A.

[17] This is simply following the Bayesian formula: $\Pr(A \mid B) = \Pr(B \mid A)\dfrac{\Pr(A)}{\Pr(B)}$

As discussed in the previous section, the measurement of conditional probabilities is subject to the error which in turn will cause the estimated coefficients from (13) to be biased. This problem needs to be addressed by the technique outlined in Deaton (1985)

There are other ways to obtain $\left\{\hat{m}^{0}_{d,age},\ \hat{m}^{1}_{d,age},\ \hat{m}^{2}_{d,age},\ \hat{m}^{3}_{d,age}\right\}$. One approach is to use external data that have a panel structure; the example is PSID (Panel Study of Income Dynamics).

Also we can model conditional probabilities of having "j" children at age t by estimating a set of equations where conditional probabilities depend on various covariates. Though this approach sound appealing from intuitive perspective, it is difficult to implement in the repeated cross section data setting.

V. Results of the adjustment for the "never have children" sub-sample

1. Estimation of conditional probabilities

The derivation of the consumption path for "never has children" households has the following stages. On the first stage we estimate $m_{d,age=t}$ for each cohort *d* at every age *t.* The second step is to estimate equation (11) using computed $\hat{m}_{d,age=t}$ as variable on the right hand side. Finally the last step is to obtain the residual of this estimation which we can test for the non-liner effects in age.

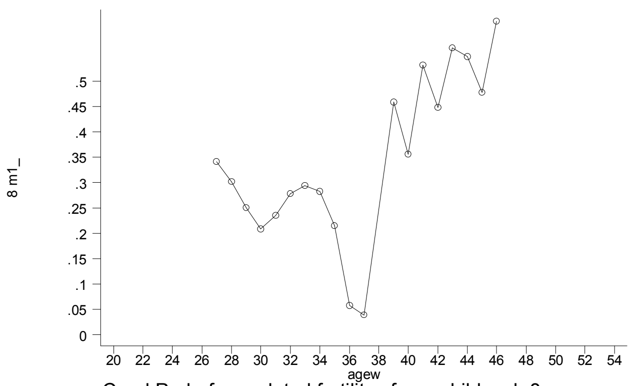
The first step in deriving the consumption profile for "never have children" households is to estimate the conditional probability of having "j" children given that there are no children currently present in the household. As discussed earlier in the modeling section, this needs to be done for two periods separately (when the wife is below and above 38 years old respectively).

Age 38 is chosen to be the age of completed fertility according to the following logical consideration: there is a high probability that family will have no more children after this age and at the same time there is a high probability that children have not reached the age when they go to college and leave home[18].

Conditional probabilities of completed fertility also denote the proportion of households in “no-child-present” sub-sample who are going to have exactly “j” children by the age of completed fertility. So when we include them into the estimation of cohort-mean consumption equation (13) we effectively adjust the sample for families that have children at some point of life.

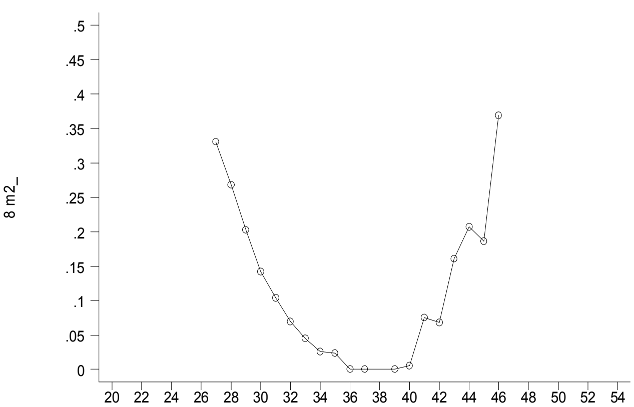
Below we present graphs for conditional probabilities for the selected cohort (cohort 8 -- women are born in 1951-1953).

One Child:



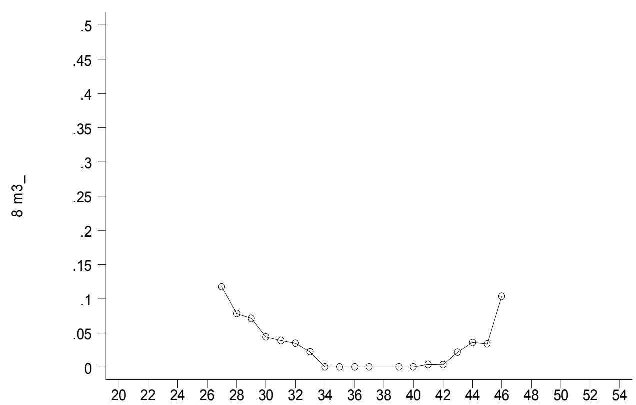
Cond Probof completed fertility of one child, coh 8

Two children:



Cond Prob of completed fertility of two kids, coh 8

Three children:



Cond Prob of completed fertility of three kids, coh 8

---

18 For each cohort in different ages we checked the difference proportions of families with two children present and no children present. The later is smallest and the former is higher when wife is 38 years old.

The curves are not smooth because of the small sample size and crude (non-smoothing) estimation. As can be seen from the graphs, however, the tendency is clear – conditional on no children currently present in the household the probability of having children is symmetrical: it decreases before age 38 (less probability that the child is born) and increases after 38 (as household ages there is a higher probability that the child have left home).

2. Estimation of consumption equation

Obtained estimates of conditional probabilities are used as right hand side variables in the estimation of equation (13). Important remark needs to be done before we do the main estimation. Along with discussing the method to identify cohort mean consumption of "never have children" families in chapter IV we assumed that our cohort sample is somewhat homogenous with respect to the permanent income. By homogeneity we imply that the mean permanent income for a given cohort doesn't change as cohort becomes older. This assumption provides that we can use decomposition method outlined earlier to identify the mean consumption $E_I^d(c_{ht})$. However, "homogeneity" of the permanent income assumption should not be taken for granted for the following reason.

Since we work with cohorts of married couples there exists a selection into married couples which affects differently various groups of people depending on their income, education, etc. The most obvious example that illustrates selection issues is the following. We can roughly define two types of families: "low educated" and "highly educated". "Less educated" group tends to marry early in life but is prone to a higher risk of divorce in the mid life. Also the permanent income of this group is in the low income percentile. The second "highly educated" group exhibits exactly opposite characteristics: high level of education, high permanent income, marries later in life, has more stable family union, have children later in life. According to these characteristics we can see that the consequence of this selection into married couples is that first half of life is "dominated" by "low education" group with the low permanent income. Thus, in various periods of life we have different permanent income groups. In order to separate the effect introduced by demographics from the selection effect we want to control for the permanent income. The obvious way would be to separate groups by their educational level and to run two separate regressions for each group. This was the approach implemented in Browning's paper (2002). However, our sample size from CEX is relatively small; therefore, this disaggregating will leave us with very few observations in each cohort-age cell. Instead we introduce six educational dummies corresponding to six levels of education. In terms of synthetic cohorts these become educational proportions. Since education level is very closely correlated with income, we believe that this should provide an adequate control for the permanent income and eliminate some of the selection issue.

Below is the extended version of the consumption equation that includes educational proportions. Recall that cohorts are constructed based on the year of birth of wife. Initially we consider education of the husband (not wife) to obtain the educational dummies. In synthetic cohort setting the variables are going to be proportions of families with a certain level of education of the husband. Later we try to add educational proportions for wives and check whether this improves the regression results. The estimation equation is as follows:

$$\begin{aligned} \log(C^0_{d,age}) = a_1 * age + \gamma_d + b_1 * \hat{m}^1_{d,age} + b_2 * \hat{m}^2_{d,age} + b_3 * \hat{m}^3_{d,age} + \\ + d_i edu\_proportions_{d,i} + \varepsilon_{d,age} \end{aligned} \tag{14}$$

We estimate this equation by OLS, with robust standard errors option to correct for the heteroskedasticity and also adjusting standard errors for the number of observations in each age-cohort "group".

Measurement error issue is the main problem in this estimation. Since we compute $m_{d,age=t}$ and educational proportions based on the internal data we are subject to the error-in-variables problem. This problem needs to be corrected in order to obtain unbiased and efficient estimates of the coefficients in the regression equation (14).

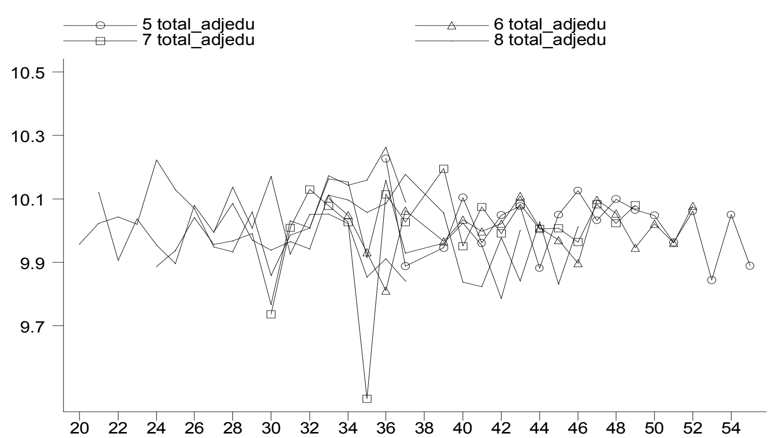
Discussion of the results of the estimation

Regression results are presented in the Appendix D. Though the coefficients at the conditional probabilities are not significant, the overall significance of the regressions is not rejected. As follows from the theory section we expect the sign of the coefficients to be negative because, controlling for their permanent income, families who are planning to have children (or have had children) should have lower consumption than "never have children" families. However, the sign of the estimated coefficient is negative only for the probability of two and tree children, while it stays positive for one child families. A possible explanation for this is that wealthier households tend to have fewer children who are born later in life. It could be the case that the first child is born before the age of 38, so the positive coefficient controls for the permanent income of wealthier families.

We exclude education level equal high school when controlling for educational proportions. The sign at educational proportions is consistent with intuition. For education less than high school it is significantly negative, and for the graduate school education it is significantly positive. These seem to serve as good controls for the permanent income. Cohort dummies γ_d are also present in the regression to control for any cohort specific effects.

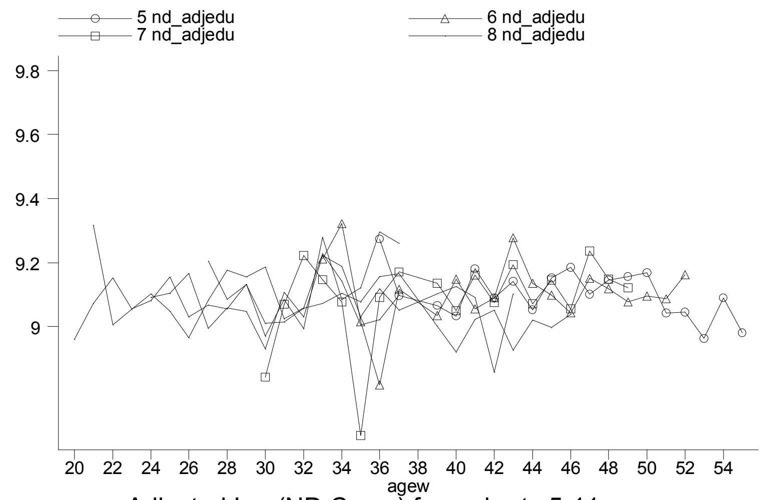
Upon getting estimates of the coefficients we adjust consumption of "no children present" sub-sample to get the consumption path of "never have children" families. As our hypothesis suggest their consumption path should not have any non-linear age effects which is consistent with life cycle hypothesis of consumption. Below are results of this adjustment.

Log (Total Consumption) Adjusted



Adjusted Log(Total Cons.) for cohorts 5-11

Log (Non-Durable Consumption) Adjusted



Adjusted Log(ND Cons.) for cohorts 5-11

As follows from the graphs the consumption path is horizontal and flat. As we expected, it does not exhibit any non-linear effects with age. It is important to notice though that the model is more "reliable" for non-durable consumption in light of assumption (10) of the model.

Now we need to test for non-linearities analytically by regressing adjusted consumption on age and age-squared. Regression results are presented in Appendix E. As results show both the age and the age squared are insignificant. Therefore, the non-linearities in the consumption path are removed by appropriately adjusting for the demographic variables.

3. Summary of potential biases

Since the model is based on a number of strict assumptions there are a number of potential biases that it can generate.

The most important bias will come from ignoring the precautionary savings explanation. Precautionary savings motive is one of the alternative ways to think about the sharp increase in consumption profile early in life when the uncertainty about future income and consumption are particularly high. A more detailed discussion of this phenomenon is discussed in Appendix F.

Selectivity issue into the sample of married couples remains one of the areas of concern (Deaton 1997). As we pointed out earlier, in the synthetic cohort construction we take means over married couples but we can't observe individual household dynamics regarding family formation and dissolution. Obviously, cohort means over time depend on change in marital status over time. For example, wealthier households have lower rates of divorce. As cohort ages it becomes on average wealthier, this way it has an impact on consumption profile. Ideally we would like to weight observations by the inverse of the probability to stay in the sample of married couples. This is a difficult task to implement since it requires knowing of the probabilities to stay in the sample. We use education as a proxy to control for permanent income but this is not the ideal control therefore the bias due to selection is not completely eliminated.

There are a number of other directions that may need to be considered. One is to address the correlation between propensity of having children and the permanent income; this is the area that may need to be explored with respect to the influence on average consumption profiles. Not considering cohabiting couples (because we are not able to observe them in the data) introduces substantial potential bias because timing between marriage and appearance of children are correlated.

There is a potential of bias coming from the assumption of completed fertility. For example consider the cases of the first or second child being born to a wife older than age 38, a recent trend in more educated households. When we calculate conditional probabilities we may mistakenly consider these households to have no children or one child in their life time.

VI. Conclusions

The goal of this paper is to test the life cycle hypothesis of consumption using household data from U.S. The main general hypothesis is that demographic characteristics of the family, such as number of children and their distribution over the life of the household, carry a large explanatory power in addressing the puzzling facts about inverted U-shape of consumption over the life cycle.

However, another way to look at this hypothesis is to reformulate it with respect to the households that never have children during their life. Then the specific hypothesis becomes the need to test the consumption of these households for the presence of non-linearities in age.

The main caveat in this approach is the repeated structure of the data which doesn't allow us to trace same households over time. First, we had to construct synthetic cohorts to imitate the panel. Second, we needed to implement adjustment for "never have children" households using approach suggested in Browning and Ejrnaes (2002).

While the analysis has the opportunity to have introduced bias through not considering selectivity issues and other explanations for the inverted U-shape profile, the results of our testing support the main hypothesis. Households who never have children in their life do smooth their consumption which is consistent with life cycle predictions but also gives support to the "demographic" explanation for the inverted U-shape of consumption for the whole sample of married couples.

References:

Browning, Crossley (2002), “The Life-Cycle Model of Consumption and Saving”, Journal of Economic Perspectives, vol. 15, no. 3, Summer 2001, pp. 3-22

Browning Martin, Angus Deaton and Margaret Irish (1985), “A profitable approach to labor supply and commodity demand over the life-cycle”, Econometrica, 53, 503-44

Browning Martin and Mette Ejrnaes (2002), “Consumption and children”, Center for Applied Microeconometrics working paper.

Browning (1992)

Browning Martin, Annamaria Lusardi (1996), “Household Saving: Micro Theories and Micro Facts”, Journal of Economic Literature, vol. 34, no. 4, December 1996, pp. 1797-1855.

Carrol Cristopher and Andrew Samwick (1998), “How Important Is Precautionary Saving?”, Review of Economics and Statistics, vol. 80, no. 3, August 1998, pp. 410-19.

Carroll, Christopher (1996), “Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis”, Quarterly Journal of Economics, vol. 112, no. 1, February 1997, pp. 1-55.

Deaton (1985)

Deaton (1992) “Understanding Consumption”, Clarendon Lectures in Economics. Oxford; New York; Toronto and Melbourne: Oxford University Press, Clarendon Press, 1992, pp. xiii, 242.

Deaton (1997) “The analysis of household surveys: A microeconometric approach to development policy”, Baltimore and London: Johns Hopkins University Press for the World Bank, 1997, pp. viii, 479.

Gourinchas, Pierre-Oliver and Parker, Jonathan (2002), “Consumption over the Life Cycle”, Econometrica, vol. 70, no. 1, pp. 47-89.

Fernandez-Villaverd Jesus and Dirk Krueger (2002), “Consumption over the Life Cycle: Facts from Consumer Expenditure Survey Data”, National Bureau of Economic Research, NBER Working Papers: 9382.

Appendix

Appendix A. Estimation of the proportion of the families that will have or have had children given that children are not present in the household

We are interested in estimating:

Pr (“j” children at age 38 | no children at given age=t) =
Pr (no children at given age=t | “j” children at age 38) * $\frac{\Pr(\text{"}j\text{" children in cohort d at 38})}{\Pr(\text{no children at age} = t)}$

While the last term in the product is easily obtained from the data, the first term must be estimated. We divide the sub-sample into two groups: below and above 38 years old. For the group below 38 years old we can construct the distribution of age at first birth using the whole sample of families (with and without children) at age 38:

Pr (no children at given age=t | “j” children at age 38) =

$$= \frac{\#\,(\text{HH with age=38, age of first birth} > \text{t}, z_{h\,38} = j)}{\#(HH \text{ with age} = 38 \text{ and } z_{h38} = \text{j})}$$

Group above 38:

Keep analysis for each cohort separately. Transition from age (t-1) to age t is $\mu_t = \Lambda_{t-1}\mu_{t-1}$

$$\mu_t = \begin{pmatrix} \mu_{0,t} \\ \mu_{1,t} \\ \mu_{2,t} \\ \mu_{3,t} \end{pmatrix} \text{ where } \mu_{j,t} = \frac{\text{\# of families at age t with "j" children present}}{\text{Total \# of families at age t}}$$

$$\begin{pmatrix} \mu_{0,t} \\ \mu_{1,t} \\ \mu_{2,t} \\ \mu_{3,t} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1-\lambda_{1,t-1} & 0 & 0 \\ 0 & \lambda_{1,t-1} & 1-\lambda_{2,t-1} & 0 \\ 0 & 0 & \lambda_{2,t-1} & 1-\lambda_{3,t-1} \\ 0 & 0 & 0 & \lambda_{3,t-1} \end{pmatrix}}_{\Lambda_{t-1}} * \begin{pmatrix} \mu_{0,t-1} \\ \mu_{1,t-1} \\ \mu_{2,t-1} \\ \mu_{3,t-1} \end{pmatrix}$$

$\lambda_{j,t-1} = Pr(z_{ht} = j \mid z_{ht-1} = j),$

where z_{ht} is number of children in household h at age t

Transition from age t to t+k (from 38 to t):

$$\Pr(z_{h\ age=t} = 0 \mid z_{h\ 38} = j) = \Lambda_{age=t}\Lambda_{age=t-1}\ldots\Lambda_{38}$$

Elements of the first row of this product [1, j] show the transition from state "j" children at age 38 to zero children at age=t.

At the end,

$$\Pr(z_{h\ 38} = j \mid z_{h\ age=t} = 0) = \Pr(z_{h\ age=t} = 0 \mid z_{h\ 38} = j) * \frac{\Pr(z_{h38} = j)}{\Pr(z_{h\ age=t} = 0)}$$

Sample for the estimation of conditional probability includes a larger number of households. We consider all women who are:

a) married
   - to the head of the household
   - leaving with their parents

b) were married previously
   - single
   - living with their parents

a) and b) with or without children.

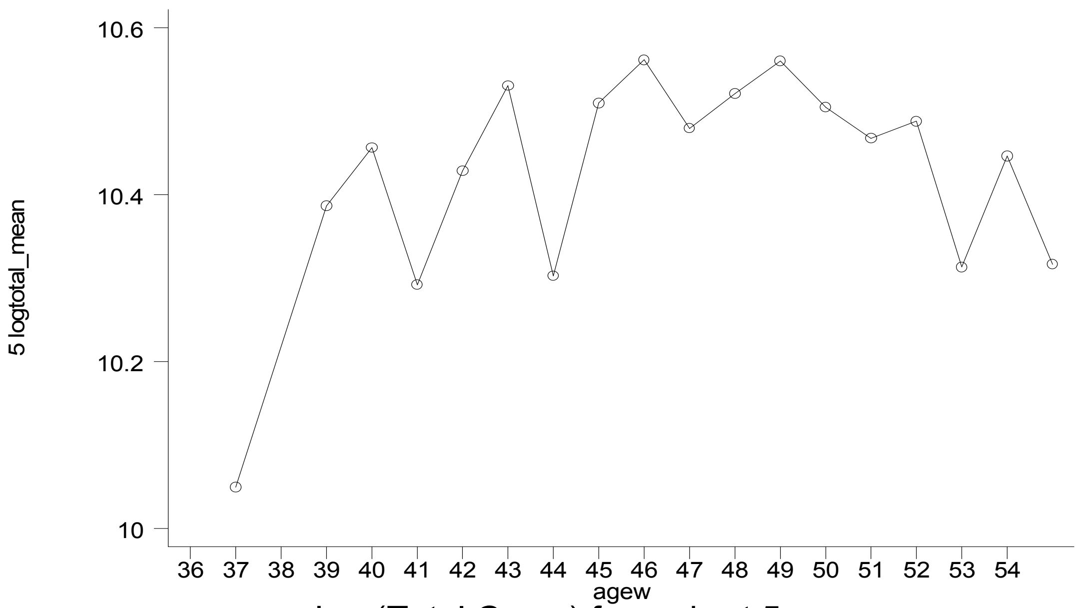
Appendix B. Synthetic Cohort Chart

Cohorts construction in CEX: based on the age of wife, three years band

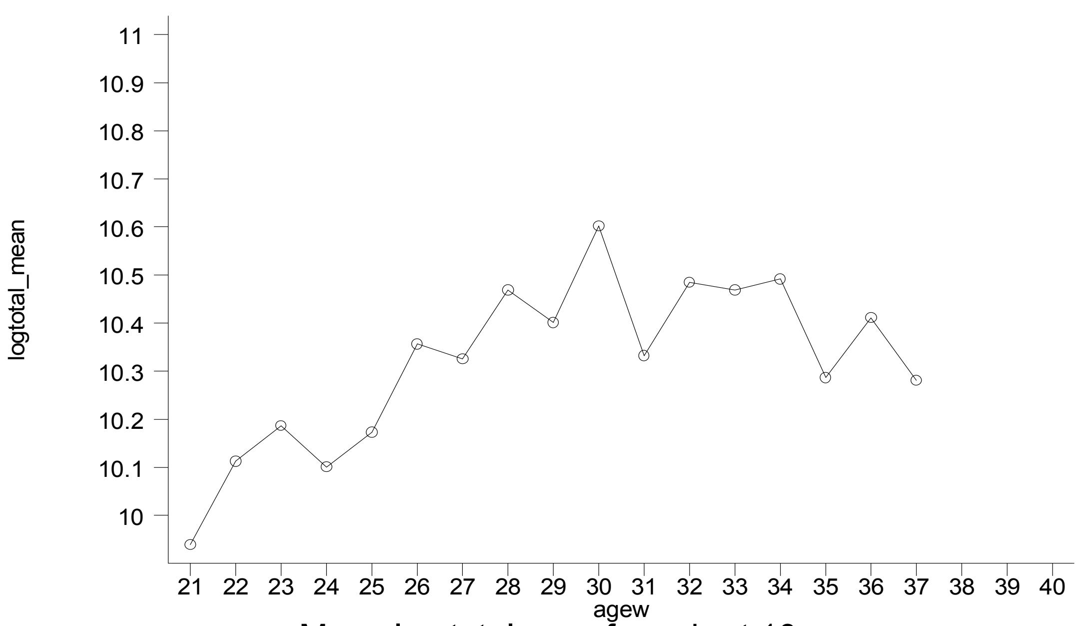
Cohort ID	Year of birth	Age in 1980	Age in 1990	Age in 1998
14	1969 -1971	9-11	19-21	27-29
13	1966 – 1968	12-14	22-24	30-32
12	1963 – 1965	15-17	25-27	33-35
11	1960 – 1962	18-20		36-38
10	1957 – 1959	21-23		39-41
9	1954 – 1956	24-26		42-44
8	1951 – 1953	27-29		45-47
7	1948 – 1950	30-32		48-50
6	1945 – 1947	33-35		51-53
5	1942 – 1944	36-38		54-56
4	1939 – 1941	39-41		57-59
3	1936 – 1938	42-44	52-54	60-62
2	1933 – 1935	45-47	55-57	63-65
1	1930 – 1932	48-50	58-60	66-68

- Synthetic panel has 130 cohort-age observations
- Covers total of ~ 4,800 households (no-child-present sample)
- Sample to estimate completed fertility is total of ~ 23,013 females

Appendix C. Total Consumption – In Logs for Selected Cohorts



Log(Total Cons.) for cohort 5



Mean log total cons for cohort 10

Appendix D. Estimation of Consumption Equation

```
      Source |       SS       df       MS              Number of obs =     130
-------------+------------------------------           F( 15,   114) =    8.43
       Model |  1.65145461    15  .110096974           Prob > F      =  0.0000
    Residual |  1.48896002   114  .013061053           R-squared     =  0.5259
-------------+------------------------------           Adj R-squared =  0.4635
       Total |  3.14041462   129  .024344299           Root MSE      =  .11428

------------------------------------------------------------------------------
Log(Total)   |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        agew |   .0109425   .0020285     5.39   0.000      .006924     .014961
         m_1 |   .1178067    .100566     1.17   0.244    -.0814138    .3170272
         m_2 |  -.0358296   .1134465    -0.32   0.753    -.2605661     .188907
         m_3 |  -.2054209   .2235332    -0.92   0.360    -.6482384    .2373966
  edu_share1 |  -.1083193    .161453    -0.67   0.504    -.4281565    .2115179
  edu_share2 |  -1.076736   .1934618    -5.57   0.000    -1.459982   -.6934897
  edu_share4 |   -.187055   .1354113    -1.38   0.170    -.4553037    .0811938
  edu_share5 |   .1778922   .1330595     1.34   0.184    -.0856977     .441482
  edu_share6 |   .3354574   .1317693     2.55   0.012     .0744234    .5964913
```

```
      Source |       SS       df       MS              Number of obs =     130
-------------+------------------------------           F( 15,   114) =    7.93
       Model |  1.20968094    15  .080645396           Prob > F      =  0.0000
    Residual |  1.15938506   114  .010170044           R-squared     =  0.5106
-------------+------------------------------           Adj R-squared =  0.4462
       Total |    2.369066   129  .018364853           Root MSE      =  .10085

------------------------------------------------------------------------------
  Log (ND)   |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        agew |   .0098262     .00179     5.49   0.000     .0062802    .0133722
         m_1 |   .0748694   .0887408     0.84   0.401    -.1009256    .2506643
         m_2 |  -.0533854   .1001067    -0.53   0.595    -.2516961    .1449252
         m_3 |  -.1644448   .1972488    -0.83   0.406    -.5551931    .2263035
  edu_share1 |   .0152236   .1424684     0.11   0.915    -.2670052    .2974523
  edu_share2 |  -.3448799   .1707134    -2.02   0.046    -.6830618    -.006698
  edu_share4 |  -.0612795   .1194888    -0.51   0.609    -.2979859    .1754269
  edu_share5 |   .1325356   .1174136     1.13   0.261    -.1000597     .365131
  edu_share6 |   .3937214   .1162751     3.39   0.001     .1633814    .6240614
```

```
      Source |       SS       df       MS              Number of obs =     130
-------------+------------------------------           F( 15,   114) =   12.31
       Model |  1.92255069    15  .128170046           Prob > F      =  0.0000
    Residual |  1.18671728   114  .010409801           R-squared     =  0.6183
-------------+------------------------------           Adj R-squared =  0.5681
       Total |  3.10926796   129  .024102852           Root MSE      =  .10203

------------------------------------------------------------------------------
Log (Food)   |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        agew |    .009416    .001811     5.20   0.000     .0058284    .0130035
         m_1 |   .1140344   .0897808     1.27   0.207    -.0638206    .2918895
         m_2 |   -.142806   .1012798    -1.41   0.161    -.3434406    .0578286
         m_3 |  -.0938589   .1995603    -0.47   0.639    -.4891863    .3014685
  edu_share1 |   .3400934   .1441379     2.36   0.020     .0545573    .6256295
  edu_share2 |  -.2731466   .1727139    -1.58   0.117    -.6152916    .0689983
  edu_share4 |  -.2356267   .1208891    -1.95   0.054     -.475107    .0038536
  edu_share5 |   .0366827   .1187895     0.31   0.758    -.1986383    .2720038
  edu_share6 |   .2983203   .1176377     2.54   0.013      .065281    .5313596
```

Appendix E. Testing for the Non-Linearities in the adjusted Consumption Profiles

Total Adjusted

```
      Source |       SS       df       MS              Number of obs =     130
-------------+------------------------------           F(  2,   127) =    0.07
       Model |  .001815695     2  .000907848           Prob > F      =  0.9288
    Residual |  1.56098646   127  .012291232           R-squared     =  0.0012
-------------+------------------------------           Adj R-squared = -0.0146
       Total |  1.56280216   129  .012114745           Root MSE      =  .11087

Adj.Log(Total) |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        agew |   .0020073   .0095424     0.21   0.834    -.0168754    .0208899
      agew_2 |  -.0000317   .0001278    -0.25   0.804    -.0002845    .0002211
       _cons |   9.978999   .1723185    57.91   0.000     9.638011    10.31999
```

Non-Durable Adjusted

```
      Source |       SS       df       MS              Number of obs =     130
-------------+------------------------------           F(  2,   127) =    0.13
       Model |  .002416329     2  .001208164           Prob > F      =  0.8799
    Residual |  1.19767895   127  .009430543           R-squared     =  0.0020
-------------+------------------------------           Adj R-squared = -0.0137
       Total |  1.20009528   129  .009303064           Root MSE      =  .09711

Adj. Log(ND) |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        agew |   .0038352   .0083585     0.46   0.647    -.0127047    .0203751
      agew_2 |   -.000048   .0001119    -0.43   0.668    -.0002695    .0001734
       _cons |   9.017525   .1509394    59.74   0.000     8.718843    9.316207
```

Food Adjusted

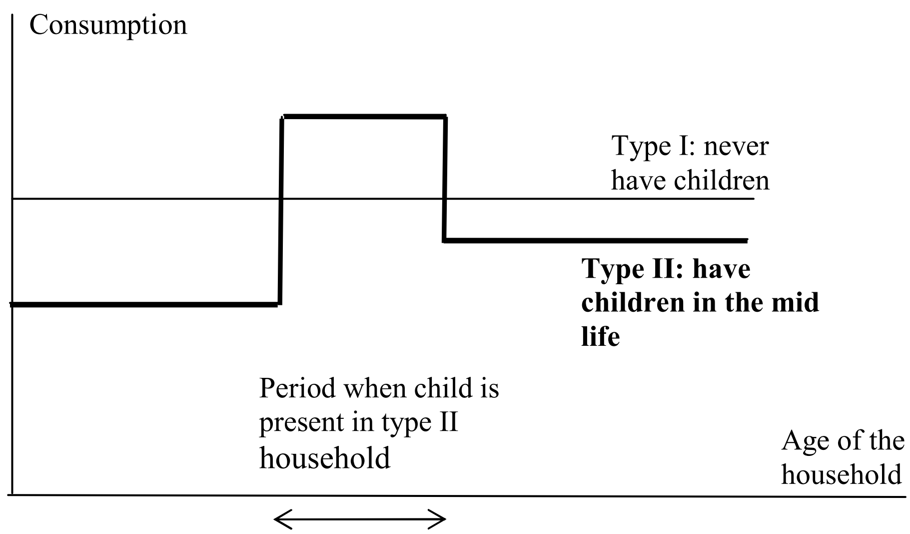
```
      Source |       SS       df       MS              Number of obs =     130
-------------+------------------------------           F(  2,   127) =    1.37
       Model |  .027238013     2  .013619006           Prob > F      =  0.2591
    Residual |  1.26704842   127  .009976759           R-squared     =  0.0210
-------------+------------------------------           Adj R-squared =  0.0056
       Total |  1.29428644   129  .010033228           Root MSE      =  .09988

Adj.Log(Food)|      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        agew |   .0050809   .0085971     0.59   0.556    -.0119313     .022093
      agew_2 |  -.0000458   .0001151    -0.40   0.692    -.0002736     .000182
       _cons |   8.094081   .1552491    52.14   0.000     7.786871    8.401291
```

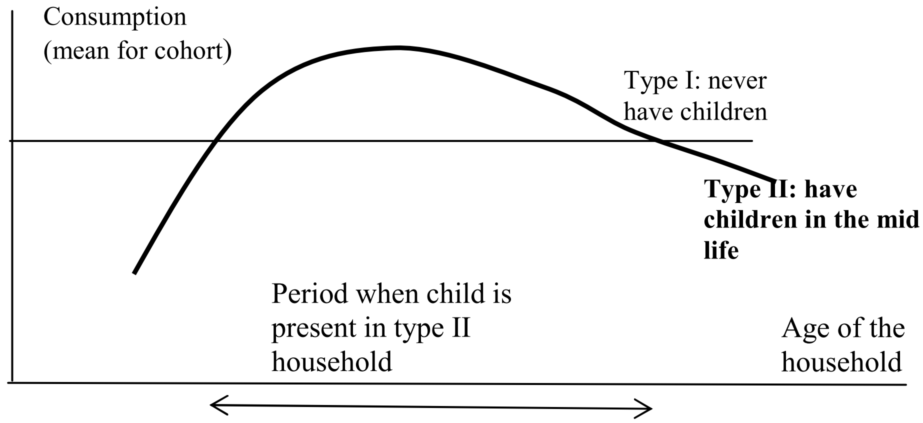
Appendix F. Discussion of Potential Biases

It seems that the most important bias comes from the main assumption of the model that we can ignore precautionary saving motive in explanation of the inverted U-shape profile. This might be a bad idea. If we look at the graph of unadjusted consumption then it is obvious that its right tail is flatter than left tail meaning that consumption rises much sharply in young ages than it declines at older ages. The essence of precautionary saving explanation is that mean preserving increases in the uncertainty about future consumption will cause a reduction in current consumption and an increase in saving. To outline the role of the precautionary saving explanation consider the graph below. Because uncertainty is smaller in the second half of life people do not save that much as they would earlier in life to address future consumption uncertainty, therefore when children leave home the family doesn't reduce their consumption to the level as before children were present in the household. This fact might explain why data look this way.

For Individual households:



For Cohorts:



The resulting graph for cohorts is flatter in the later age consistent with what we expect in a graph of an individual household. In addition to precautionary saving another explanation for this tendency might be that there is a different dynamic between children arriving in a household and children leaving.