NBER WORKING PAPER SERIES

# A COHORT ANALYSIS OF SAVING BEHAVIOR BY U.S. HOUSEHOLDS

Orazio P. Attanasio

Working Paper No. 4454

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1993

This paper uses part of the CEX data set put together at Stanford, using the Bureau of Labor Statistics public tapes of the Consumer Expenditure Survey. This project has been financially supported by the Center for Economic and Policy Research and by NSF grant SES-9057511. The painful construction of the data set has been speeded up by the outstanding research assistance of Kim Coleman and Penny Koujianou and by the invaluable advice of Guglielmo Weber. Penny Koujianou and Cate de Fontemay's patience was also tested during the numerous revisions of this paper. I would also like to thank Tom MaCurdy, Frank Wolak, four anonymous referees, the Editor and especially John Pencavel for many valuable suggestions. Previous versions of this paper were presented at the Economic Fluctuations NBER Meeting, Cambridge, July 12, 1991, and at seminars at Stanford, University College London, Bologna University, the Bank of Italy, the Bank of Spain, INSEE and Ohio State University. Of course, responsibility for any error is mine. This paper is part of NBER's research program in Economic Fluctuations. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #4454
September 1993

# A COHORT ANALYSIS OF SAVING BEHAVIOR BY U.S. HOUSEHOLDS

## ABSTRACT

In this paper I analyze the pattern of saving behavior by U.S. households, using the Consumer Expenditure (CEX) Survey. The analysis' main goal is to explain the decline in aggregate personal saving in the United States in the 1980s. I estimate a 'typical' saving-age profile and identify systematic movements of this profile across different cohorts of US households. In addition I consider different definitions of saving and control for a number of factors that figure in popular explanations of the decline in saving.

The main results can be summarized as follows:

1) the 'typical' saving-age profile presents a pronounced 'hump' and peaks around age 60;

2) this 'typical' age profile was, a least during the 1980s, shifted down for those cohorts born between 1925 and 1939. This is consistent with the low level of aggregate saving because these cohort were, in the 1980s, in that part of their life cycle when saving is highest;

3) this result holds for various definition of saving with one notable exception: the decline is less pronounced when expenditure on durables is considered as saving; and

4) some other popular explanations of the decline in saving are rejected by the data, including those appealing to the presence of capital gains on real or financial assets.

Orazio P. Attanasio
Stanford University
Stanford, CA 94305
and NBER

*'OK' I said. 'You win. It's a good theory...'*
*Randall smiled wryly: ' I don't like it myself.*
*I was just trying it out. It fits the facts–*
*as far as I know them. Which is not far.'*
*'We don't know enough to even start theorizing'.*
*Raymond Chandler: Farewell, My Lovely*

## 1. Introduction.

In this paper I describe and analyze the pattern of saving behavior by US households, using the Consumer Expenditure (CEX) Survey. The analysis' main goal is to explain the decline in aggregate personal saving in the United States in the 1980s. I believe the study of aggregate time series on personal saving cannot give a satisfactory answer to the questions asked in this paper. Instead, by looking at individual data, we can test more directly different hypotheses about the decline in saving rates. In particular, it is possible to focus on well defined demographic and sociological groups and characterize their behavior: this may help in understanding the causes of the decline. To such a purpose I use the CEX survey.

In what follows I estimate a 'typical' saving-age profile and identify systematic movements of this profile across different cohorts of US households. In addition I consider different definitions of saving and control for a number of factors that figure in popular explanations of the decline in saving [1].

The main results can be summarized as follows:

1) the 'typical' saving-age profile presents a pronounced 'hump' and peaks around age 60;

2) this 'typical' age profile was, at least during the 1980s, shifted down for those cohorts born between 1925 and 1939. This is consistent with the low level of aggregate saving because these cohort were, in the 1980s, in that part of their life cycle when saving is highest;

[1] Several studies have tried to explain the decline in personal saving without much success. Hendershott and Peek (1987) analyze aggregate time series data and concentrate on definitional issues. Summers and Carroll (1987) and Auerbach and Kotlikoff (1990) look at both macro and micro data. Auerbach and Kotlikoff (1990), in particular, use several years of the CEX survey. Their analysis, however, focuses on demographic changes and assumes constant consumption and income age profiles. They conclude their paper by saying: "What happened to US saving in the 1980s remains an intriguing puzzle". More recently Bosworth, Burtless and Sabelhaus (1991) have used the 1972-73 and the 1980s CEX surveys to compare saving rates by age classes in the 1980s and the 1970s obtaining results similar to those discussed below. Their analysis, however, is substantially static, because it is based on the implicit assumption of the absence of cohort effects: age profiles for saving rates are estimated using a single cross section of data. As discussed below, the use of cross section data can be misleading when analyzing saving behavior, which is intrinsically dynamic.

3) this result holds for various definition of saving with one notable exception: the decline is less pronounced when expenditure on durables is considered as saving; and

4) some other popular explanations of the decline in saving are rejected by the data, including those appealing to the presence of capital gains on real or financial assets.

Though I do not test explicitly any economic model, the life cycle hypothesis provides the framework for my analysis of saving data. According to the model, a person saves at one stage of his or her life to consume at another. Therefore it is important, in modelling saving behavior, to follow over time the behavior of the same individuals or of individuals with similar life cycle experiences. This is the basic motivation for the cohort analysis discussed below. In the presence of strong cohort effects, the cross sectional age profile may not correspond to the age profile of any individual households.

The CEX Survey is an on-going expenditure survey sponsored by the Bureau of Labor Statistics to compute the weights for the Consumer Price Index. Since 1980 the survey has been carried out on a regular basis and provides economists with a wealth of data that have not been thoroughly analyzed. To the best of my knowledge, it is the only US micro data set with exhaustive information on consumption.

The households participating in the CEX survey are replaced every year. Therefore, the panel element of the dataset is very short. To implement an interesting dynamic analysis, I resort to 'average cohort' techniques. These techniques can be used to construct a pseudo-panel from a time series of cross sections, along the lines suggested by Browning, Deaton and Irish (1985). By averaging over individuals that share thes same year of birth, it is possible to follow a cohort over time, as it ages. This methodology, that I describe and extend to control for within cohort heterogeneity in section 3, can give useful insights about saving behavior over the life cycle.

As discussed in section 3, it is typically very difficult to disentangle time, age and cohort effects on a variable which, for each cohort, is observed over different age intervals at different times. This is particularly true when the time period covered by the available data is relatively short. I overcome this problem by first characterizing the shape of the saving-age profile, and then studying shifts in this profile. As shown below, if a given variable can be represented as the sum of polynomials in age, time and year of birth, only the coefficients on the linear terms of these polynomials are not identifiable. That is, it is possible to identify the shape (but not the position) of the age profile for the *changes* in that variable. One can therefore use information on the *stock* of financial assets to characterize the age profile for its *changes*, i.e. financial saving.

Under the assumption of a stable relation between total and financial savings, I use this profile to characterize the life cycle saving behavior of US households and to justify my interpretation of the data on saving, as measured by disposable income minus consumption. This methodology is used to justify the identification assumptions used to derive some of the results listed above.

## 2. The CEX Survey

The main data sources for this paper are the eleven years of the CEX survey from 1980 to 1990. In this section I discuss the main features of the survey and compare the Survey data on consumption and income with those of the National Income and Product Accounts (NIPA). A detailed description of the data is contained in Appendix 1.

### *2.1 Overview*

The CEX sample is representative of the universe of US households [2]. Each quarter about 5,000 households participate in the CEX survey. Each household's reference person is interviewed and reports data on expenditure on about 500 different commodities. These data on expenditure are supplemented by a rich set of economic, demographic and sociological variables, including income. This wealth of information gives one the possibility of analyzing different definitions of consumption and therefore saving at the household level. The behavior of conceptually different definitions of saving is important to understand the dynamics of aggregate saving. Furthermore, the information on demographic, economic and sociological individual characteristics can be used to test some popular explanations of the decline in aggregate saving and to identify which groups of the population are mainly responsible for it.

The CEX sample is a rotating panel: each household is interviewed four times (plus a contact interview from which no data are available in the public use tape) during one year [3]. To define saving at the household level, I compute annual consumption as the sum of all expenditures reported in the completed interviews. If a household has not completed all four interviews, the total is adjusted by dividing by the number of months on which consumption data are available and multiplying by twelve.

One of the advantages of having detailed expenditure data is that different definitions of consumption may be constructed. In what follows I use two different aggregates. First, I construct a measure of personal consumption close to the NIPA definition. This includes expenditure on goods and services, cash contributions to organizations, interest payments on mortgages, vehicle loans and other loans, and insurance premia and it does not include pension contributions. Second, I exclude from the definition above expenditure on various items (such as durables, education and health) that may be considered as saving. A corresponding measure can be easily constructed for

[2] I exclude from my sample non-urban households because they were not covered by the 1982 and 1983 surveys.

[3] The survey structure is described in detail in Appendix 1.

National Accounts data.

I also worked with other measures of consumption. In particular, I excluded from consumption (in turn) expenditure on health (see the discussion below on the recording of health expenditure in the CES), education (which can be considered as investment in human capital), and mortgage payments. None of these modifications affected the main results reported below. Details are available on request.

Income is defined as total after tax family income [4] in the twelve months preceding the interview. Therefore the figure from the last interview matches the time period to which consumption is referred. Saving is defined as income minus consumption.

Total financial assets, which are used in section 4, include savings and checking accounts, US saving bonds, equities, stocks and bonds.

Nominal figures are deflated by a household specific Stone price index constructed using expenditure shares on the 27 consumption categories on which monthly price indexes are available at the regional level [5].

Various selection criteria are used: I exclude from the sample non-urban consumer units, units with incomplete income responses [6] and units not belonging to the cohorts I will analyze (see section 2.3 below). Other less important selection criteria are described in Appendix 1.

The BLS provides weights for each household in the CES sample. These can be used to aggregate consumption and income levels. I use them in what follows to estimate simple and conditional means. Details on the use of weights can be found in Appendix 2.

*2.2 Comparison with NIPA*

The definitions of saving that it is possible to construct using CES data do not match exactly the NIPA definition of personal saving for various reasons. First, the CES is an interview survey based on recall questions. It is well known that such surveys usually underestimate consumption expenditure. Second, some consumption items are defined differently, the most important being health and housing expenditure. For the former, the CES provides information only on out-of-pocket expenses. (Indeed, health expenditure can be negative if the households has received a refund for an expenditure incurred in the past.) For the latter, NIPA include imputed rent for

[4] Total after tax family income includes earnings, transfers, capital income and pensions net of all income taxes including social security contributions.

[5] The construction of the price index is described in Appendix 1.

[6] Total consumption expenditure for this subgroup is not significantly different from the households with complete income responses.

owner occupied housing, while no attempt is made to construct a similar estimate for the CES. Third, in 1982 and 1983 the CES sample excludes rural households, so that its sample is not representative of total US population.

While these are serious problems, they do not undermine the use of the CES to study saving behavior. The first of the issues listed in the previous paragraph has been studied by Gieseman (1987) and by Paulin et al. (1990) who show that the under-reporting in consumption is roughly constant over time for most consumption categories. There is no systematic study of the reliability of the CES income figures or careful comparisons of the CES sample with other (larger) samples such as the Current Population Survey and with the National Accounts Data.

Little can be done to assess the bias introduced by the different definitions of some items and by the exclusion of rural households. In Table 1 I compare CES aggregates with National Accounts statistics. For the CES aggregates, I use the average published in the Bureau of Labor Statistics Bulletin. I do not use my sample in this comparison both because it is not representative [7] and because the averages published by the BLS are based on observations that are not top-coded [8].

The first four columns of the table contain the rate of growth of income and consumption at constant dollars for the aggregated CES data and for NIPA data [9]. The correlation coefficient between real consumption growth in the CES and in the NIPA is as high as 0.71. That between real disposable income growth, is only 0.27. The last two columns of table 1 contain the aggregate saving rates implied by the aggregate CES and by NIPA. The correlation coefficient between these series is 0.27. The largest discrepancies between the NIPA and CES data for saving are in the years 1984, 1985 and 1986, in which the CES saving rates are negative. In general, the aggregate CES series are more volatile than the corresponding NIPA series, probably reflecting the estimation error induced by relatively small samples. The correlation coefficient between the NIPA time

[7] The aggregates obtained from my sample are different for two reasons: the exclusion of several groups of consumer units (discussed above) and the methodology used to assign observations to time periods. In computing its aggregates, the BLS uses monthly observations for income and consumption. It is possible that the data for a given household are divided between two years. I prefer to construct annual household data for income and consumption so that they refer to the same time period (see footnote 9). My main aim is not to replicate the NIPA statistics but to explain individual saving behavior.

[8] The problem of top-coding is relevant only for income. Using the public use tape one can replicate the aggregate consumption figures published by the BLS.

[9] The NIPA data are taken from the December 1991 Survey of Current Business, which incorporates the recent (substantial) revision of National Accounts Data implemented by the Department of Commerce. The decline in personal saving rates has been less dramatic than implied by the unrevised data.

series in table 1 and the corresponding series obtained aggregating my sample (which excludes non-urban consumers, the households headed by an individual born before 1910 or after 1959 and the consumer units living in student housing), is around 0.7 (for each of the three series).

While the correspsondence between the CES data and the NIPA data is far from perfect, the aggregate CES seem to be reasonably similar to the NIPA data. The correlation coefficients indicate that the CES, with all its problems, is a fairly representative sample.

*2.3 Cohort and cell definition*

As discussed below, the analysis in this paper is based on the concept of synthetic cohort. A cohort is defined by the year of birth of the household head. The choice of the interval that defines a cohort is arbitrary. A narrower definition would reduce within cell heterogeneity, but at the cost of reducing the number of households within each cell. In this paper a cohort is defined by a five-year birth interval. All households whose head was born after 1959 or before 1910 were eliminated from the sample. All the remaining households are allocated to 10 cohorts on the basis of the age of the household head. The total size of my sample (after applying other selection criteria listed in the Appendix is 47,647 [10].

The households belonging to each cohort are then assigned to different time periods. Households are interviewed every month. Therefore the time span to which income and consumption refer does not coincide with the calendar years except for a minority of households, those interviewed in January. I assign to year $t$ all households interviewed between July of year $t$ and June of year $t+1$ [11].

The definition of the cohorts, together with their median age in 1980 and in 1990 and the average cell size, is reported in table 2. In the computations below the year of birth of each cohort $c$ is defined as the median year of birth of the household heads in that particular cohort, while age is defined as $a = t - c$, where $t$ is the year in which the household is interviewed.

[10] The sample used for the analysis of asset accumulation is slightly smaller. The question on assets is asked only in the last interview. Therefore, only households that do not drop out of the sample have data on assets. Furthermore, some households that have valid data on consumption and income and make it to the last interview have missing values for assets.

[11] When computing annual averages the BLS follows a different procedure. It computes monthly income and consumption (which for the former variable is imputed) and averages all the households reporting data in a given year. As a consequence, data on a given household can typically be assigned to more than one year. Given that I wanted to characterize saving, I preferred to match exactly the time period to which income and consumption refer.

## 3. Methods.

According to the life cycle model of Modigliani and Brumberg (1954), the main motivation for saving is the desire to smooth consumption over time in the face of an uneven income profile. The model is intrinsically dynamic. If we want to use this framework to interpret movements in aggregate saving, it is crucial to identify the saving-age profile and its movements over time.

To this purpose, the 'snapshot' offered by a single cross section can be quite misleading. If there are strong cohort effects, a cross section age profile may be very different from the age profile of any individual. Shorrocks (1975), for instance, discussing the life cycle accumulation of wealth, constructs an example in which individuals belonging to different cohorts keep accumulating wealth as they age. If younger cohorts are 'wealthier', in life cycle terms, than older cohorts (perhaps because of productivity growth), and these effects are strong enough, the use of a single cross section will give the illusion of a 'hump shaped' age profile. Similar examples can be constructed for the analysis of saving.

Ideally, one would like to characterize a 'typical' age profile using observations on the same individuals over time. Unfortunately, panel data on consumption and income are not available. As an alternative, we can focus on individuals with similar life cycle experiences. In what follows, I use a time series of repeated cross sections to create 'pseudo panel' data. Indeed, if we are interested in characterizing *average* behavior, because it avoids the problem of non random attrition, a time series of cross sections used to construct synthetic cohort data may be superior to panel data.

Having characterized the average saving behavior of the cohorts we observe over the sample period, the next step is to model it. Within the life cycle model, it is possible to consider several factors that affect the level of saving: the level and distribution over the life cycle of resources available to a cohort, preferences, intertemporal prices, and aggregate (business cycle) shocks. These factors can conceivably be summarized by polynomials in age, year of birth and time.

### *3.1 Cohort Analysis*

The use of average cohort techniques, as proposed by Browning, Deaton and Irish (1985) and recently discussed by Moffitt (1991), overcomes the difficulty of studying the life cycle *dynamics* of variables such as consumption or income caused by the non-availability of observations on the same individuals at different times.

The procedure consists in dividing the individuals in the sample in cells defined by their year of birth (or cohort) and by the year (or period) of interview and averaging the variable under study

over the individuals belonging to each cell [12]. The total number of observations available for the analysis is given by the number of cohorts times the number of time periods. It is then possible to follow the evolution over time of the cohort averages of the variables of interest [13].

Consider a variable $X_t^{ach}$, observed for household $h$, whose head is of age $a$ at time $t$ and belongs to cohort $c$, defined by the year of birth [14]. We can decompose its variability within a year- cohort cell in the following way:

$$X_t^{ach} = \delta_t^c + \epsilon_t^{ach} \tag{1}$$

where $\delta_t^c$ is a measure of location and $\epsilon_t^{ach}$ are deviations from $\delta_t^c$. The age corresponding to $\delta_t^c$ is, of course, $t - c$. Usually the $\delta_t^c$'s are cell means, in which case the $\epsilon_t^{ach}$'s are random variables with zero mean. However, other measures of location, such as medians or various quantiles, may be considered.

If the $\delta$'s are means and the data are not censored, they can be estimated simply and efficiently by the sample means of the individuals within each cell. When the data are censored the cell means can be estimated by maximum likelihood, if we are willing to make some assumptions about the distribution of the $\epsilon$'s. If quantiles rather than means are of interest as a measure of location, a simple and robust estimator of the $\delta_t^c$'s is given by the sample quantiles.

Income and financial assets data are topcoded in the CEX Survey. I estimate cell means by maximum likelihood parametrizing a flexible density function for each cell. For income (and saving) I assume that the density function within each cell is given by the mixture of two normal distributions. Therefore the distribution of each cell is characterized by five parameters. For financial assets, the assumed density has mass point at zero and, conditional on positive assets,

[12] For instance, if a cohort is formed by individuals born between 1930 and 1934, we average over the individuals aged between 46 and 50 in 1980, over those aged between 47 and 51 in 1981 and so on.

[13] Ultimately one ends up using aggregated data. The main advantage compared with studies that use aggregate time series data is that one controls the aggregation process directly. If the theory involves the consideration of some nonlinear function of a given variable, one can easily construct the averages of that nonlinear function. An example relevant for this paper is that of saving rates: ratios do not aggregate. However, it is possible to construct from individual observations average (or median) saving rates which would not be possible to obtain from averaged consumption and income data.

[14] If $c$ is the year of birth, $c$, $t$ and $a$ will be linked by the linear relationship $a = t - c$. In this sense the notation used here is superfluous. In the analysis that follows, a cohort is defined by a five year interval. In this case we can define the 'age' of a cohort as the age of the individuals born in the mid-year.

is given by the mixture of two lognormal distributions. Therefore the distribution of each cell is characterized by six parameters [15].

*3.2 Modelling average cohort data*

One of the main findings of this paper is that the low level of aggregate saving during the 1980s is explained by the lower saving-age profiles of certain cohorts. To circumvent the fact that I do not observe the behavior of the currently middle aged when they were young, or of the currently old when they were middle aged, I use more 'structure'. I identify the shape of a 'typical' saving-age profile and compare this with the average cohort saving observed during the 1980s. I interpret deviations of observed average saving from the 'typical' profile as a cohort effect (or equivalently as a combination of age and time effects) [16].

In general, using information on average cohort saving only, it is not possible to identify a 'typical' saving-age profile, and therefore separate age effects from cohort and time effects. In what follows I show how data on the stock of financial assets (contained in the CEX survey) may be used for such a purpose [17].

Suppose that the $\delta_t^c$ in equation (1) are cohort means for a stock variable (for instance, financial asset holdings) and suppose they are expressed as polynomials in age, time and year of birth. Without additional structural and/or statistical information, it is impossible to identify and separate age, time and cohort effects because of the linear relationship among these variables: $t = c + a$. However, if we are interested in the behavior of *changes* in financial assets, it is possible, under some circumstances, to identify most parameters of interest. To illustrate this, suppose that

[15] The reliability of these estimates depends on the parametrization of the density function and on the untestable assumption that such a parametrization fits well the *unobserved* right tail of the within-cell distribution. It is important to choose flexible specifications that allow for the considerable degree of skewness and kurtosis that characterizes the within-cell distribution of both income and asset holdings. Further details on the estimation techniques used are given in the Appendix.

[16] If the CEX survey had been available on a continuous basis since the early 1970s this argument could be made much stronger. Unfortunately, only two cross sections (the 1960-1961 and 1972-1973) are available before the 1980s. These surveys were conducted with a very different methodology, making a comparison with the current CES very difficult. Furthermore, because I am interested in the dynamics of saving, two data points, separated by ten years, are only moderately useful. Bosworth et al. (1991) compare the cross section age-saving profiles estimated with data from the 1980 with those estimated with the 1960s and 1970s CEX survey. The conclusions they reach are similar to those of this paper.

[17] The discussion that follows is related to that in MaCurdy and Mroz(1990) and has been influenced by numerous conversations with Tom MaCurdy. See also the discussion in Heckman and Robb (1987).

the $\delta_t^c$'s are a function of three distinct polynomials in age, cohort and time [18].

$$(2) \qquad \delta_t^c = \alpha_0 + \alpha_1^* a + \alpha_2 a^2 + \alpha_3 a^3 + \pi_1^* t + \pi_2 t^2 + \pi_3 t^3 + \gamma_1^* c + \gamma_2 c^2 + \gamma_3 c^3 + u_t^c$$

While it is not possible to identify separately the coefficients on the linear terms ($\alpha_1^*$,$\pi_1^*$ and $\gamma_1^*$), it is possible to estimate the following equation:

$$(3) \qquad \delta_t^c = \alpha_0 + \alpha_1 a + \alpha_2 a^2 + \alpha_3 a^3 + \pi_2 t^2 + \pi_3 t^3 + \gamma_1 c + \gamma_2 c^2 + \gamma_3 c^3 + u_t^c$$

where $\gamma_1 = \gamma_1^* + \pi_1^*$ and $\alpha_1 = \alpha_1^* + \pi_1^*$. Taking first differences of equation (3) we get:

$$(4) \qquad \Delta\delta_t^c = \alpha_1 + \alpha_2 \Delta a^2 + \pi_2 \Delta t^2 + \alpha_3 \Delta a^3 + \pi_3 \Delta t^3$$

All the coefficients in equations (3) and (4) are identifiable. Only the interpretation of the intercept of equation (4) is ambiguous. Therefore, average cohort data on a stock variable can identify the shape (but not the position) of the age profile for the first differences of the same variables.

In section 4 I estimate equation (3) for the stock of financial assets. This allows me to identify the shape of the age profile for the *changes* in financial assets, i.e. all the slope coefficients in equation (4). Changes in financial assets are clearly related to saving: in fact they are an important component of total saving.

The next step is to assume that the shape of a 'typical' age saving profile is the same as that of the profile for changes in financial assets. I then identify movements in the age-saving profiles of different cohorts by comparing observed average cohort saving to those implied by this hypothetical and 'typical' age profile, kept constant over time and across cohorts. The implicit assumption here is the existence of a stable relationship (over the sample period) between total and financial savings.

In the example above I assumed that the $\delta_t^c$ are a function of distinct polynomials in age, time and year of birth and neglected the possibility of interaction terms. For terms of third or higher order this separability assumption is actually testable: not only are the coefficients on powers of age, time and cohort greater than 2 identifiable, but the presence of interaction terms whose exponents sum up to 3 or more is testable. The effect of quadratic terms in equation (3) cannot

[18] Interaction terms are neglected only for expositional simplicity and will be discussed below. The polynomials in the example could obviously be of order higher than 3.

be distinguished from the effect of interaction terms whose exponents add up to 2. However, if we are interested in the changes in $\delta_t^c$, this will not be crucial [19].

*3.3 Within-cohort heterogeneity*

The evolution over time of average saving (or of mean financial wealth) of a given cohort may be an interesting phenomenon. However, we might wish to control for a number of other variables, besides age and year of birth. Heterogeneity within a year-cohort cell can be substantial. Furthermore, changes in the cross sectional distribution of some control variables over time can be relevant for the variable under study. Finally, accounting for heterogeneity within a cohort and over time can be quite important both for estimation efficiency and for finding an explanation for the dynamic behavior of a variable.

First, suppose that the within-cell distribution of $X_t^{ach}$ depends on some variables that are household and cohort specific, but do not change (or are assumed not to change) over the life cycle. These variables (denoted by $w^{ch}$) include things such as the race and sex of the household head, and perhaps the region of residence and schooling.

There are at least three reasons to control for these variables. First, the effects of $w^{ch}$ on $X_t^{ach}$ may be interesting in their own right. Second, even if the distribution of $w$'s within each cohort varies over time only because of sampling error and, therefore, in a non systematic way, controlling for this sort of heterogeneity may improve the efficiency of the estimates of the year-cohort effects. Finally, the composition of the cohort with respect to the $w's$ can change in a systematic way across time. If this is the case, ignoring these effects may lead to biased estimates of the $\delta$'s.

Ideally one would like to analyze separately groups of households homogeneous with respect to these variables. However, this is often infeasible because it would leave us with extremely small cells. Alternatively, one can model these effects in the following way:

---

[19] An example will make this clear. Equation (3) in the text without the cubic terms is observationally equivalent to the following:

$$\delta_t^c = \alpha_0 + \alpha_1 a + \tilde{\alpha}_2 a^2 + \theta ac + \gamma_1 c + \tilde{\gamma}_2 c^2 + u_t^c$$

where $\tilde{\alpha}_2 = \alpha_2 + \pi_2$, $\tilde{\gamma}_2 = \gamma_2 + \pi_2$ and $\theta = 2\pi_2$.

If we are interested in the changes in $\delta_t^c$ the implications of the two observationally equivalent structures are the same:

$$\Delta\delta_t^c = \alpha_1 + \alpha_2 \Delta a^2 + \pi_2 \Delta t^2 = \alpha_1 + \tilde{\alpha}_2 \Delta a^2 + \theta c$$

$$X_t^{ach} = \tilde{\delta}_t^c + \beta_1 w^{ch} + \epsilon_t^{ach} \tag{5}$$

Equation (5) is very restrictive in that the $w^{ch}$ are not interacted with the year cohort dummies. This implies that the shape of the age profile for $X$ is the same for individuals with different $w^{ch}$: only the intercepts of these profiles are allowed to be different. For instance, if the variable of interest is income and the control variable is schooling, equation (5) implies that the difference in income across education groups at different ages and time periods is constant. This is unsatisfactory. Unfortunately, given the size of a typical cell, there is no alternative. We do not have enough data to measure with some precision the shape of age profiles for different schooling and other demographic groups [20].

Notice that, even though $w^{ch}$ does not change for a single household, the composition of the sample with respect to $w$ changes because the households drawn in each survey change. Only if the $w^{ch}$'s are uncorrelated with the year- cohort dummies, does their omission not bias the estimates of the $\delta_t^c$'s in equation (5).

The problem of differential mortality by schooling provides an example of the importance of controlling for within cohort heterogeneity. Suppose that saving is positively related (as it turns out to be the case) to the schooling level of the household head. Longevity and education are positively related. If we do not control for schooling while estimating the age-savibg profile we will overestimate its last part. This is because we will be sampling from a population whose schooling level becomes progressively higher, as the less educated die younger [21].

The parameters in equation (5) can be estimated by OLS or, in the presence of top-coded observations and if we are willing to make distributional assumptions, by Maximum Likelihood. Maximum likelihood estimation is straightforward if we interact the $w^{ch}$ variables with the year cohort dummies: in that case we can decompose the maximization of the overall likelihood into the maximization of the likelihood for each cell. If we want to impose the restriction that the coefficients on $w^{ch}$ are the same across cells the maximization problem is numerically much more complicated because it involves a very large number of parameters that have to be solved for simultaneously. In practice I used a Minimum Chi-Square method which involves, in the first step,

[20] I estimated age-saving profiles for three schooling groups and I will refer briefly to these results below.

[21] A similar problem is relevant for the estimation of wealth age profiles, because of the well known relation between wealth and longevity.

the maximization of each cell's likelihood without imposing the cross-cell constraint and, in the second step, the minimization of a quadratic form. Details are provided in the appendix.

If conditional medians are of interest, the parameters of equation (5) may be estimated by using a Least Absolute Deviations estimator.

In addition to variables like the $w$'s, we might wish to control for variables that change over the life cycle. They include family composition, housing tenure, asset ownership and so on. They will be denoted by $z_t^{ach}$. In this case, equation (5) can be modified in the following way:

$$X_t^{ach} = \tilde{\delta}_t^c + \beta_1 w^{ch} + \beta_2 z_t^{ach} + \tilde{\epsilon}_t^{ach} \tag{6}$$

The interpretation of the $\tilde{\delta}_t^c$ is now very different: they represent the evolution over the life cycle of $X_t^{ach}$ which is not accounted for by changes in $z_t^{ach}$. As an example, suppose we model the $z$'s as a function of time and cohort effects and possibly other cohort and household specific variables, in the same way as we modelled $X_t^{ach}$ in (2):

$$z_t^{ach} = \gamma_t^c + \theta w^{ch} + \eta_t^{ach} \tag{7}$$

If we substitute equation (7) into (6) we identify a relationship between the $\delta$'s and the $\tilde{\delta}$'s:

$$\delta_t^c = \tilde{\delta}_t^c + \beta_2 \gamma_t^c \tag{8}$$

Knowledge of these effects may be interesting in their own right, but may also be used to identify changes in the behavior of $\delta_t^c$ over time. In terms of equations (7) and (8), it may be interesting to test the hypothesis that changes in the profile described by the $\delta_t^c$'s are accounted for by changes in the $\gamma_t^c$'s or in the $\tilde{\delta}_t^c$'s.

An example illustrates the value of equation (6). Consider as a $z_t^{ch}$ variable a dummy that indicates asset ownership. The idea is that, if observed saving rates have been low because there have been unobserved capital gains that should be included in income, this should be relevant for households with positive assets. Therefore, comparing the profiles estimated with and without controlling for asset ownership, one can test the hypothesis that unobserved capital gains are responsible for the observed decline in saving of some cohorts [22].

The estimation methods for equation (6) are analogous to those discussed for equation (5).

[22] One should be careful in interpreting these results: a test of this kind does not take into consideration the size of the stock of assets. It points only to possible differences between average saving of asset holders and non asset holders.

## 4. Identifying age profiles for saving

In this section I estimate cohort means for total financial assets and model them as a function of age, time and year of birth [23]. Because the four components of financial assets are topcoded, I estimate the mean of each cell by Maximum Likelihood, following the procedure sketched in section 3.1 and described in detail in Appendix 2.

The results are discussed at length in a companion paper (see Attanasio, 1992). Here I discuss only their implications for the shape of the age-saving profile. Similar results are obtained when, instead of unconditional means, I compute conditional means, conditioning on various demographic variables.

The estimated $\delta_t^c$'s are modeled as polynomials in age, time and year of birth. After some specification search, I settled on polynomials of order four in age, five in time and three in $c$. Obviously, when estimating, one of the linear terms of the three polynomials is excluded. The results are insensitive to the addition of higher powers in the three variables; the hypothesis that the coefficients on various interaction terms were equal to zero could not be rejected [24].

As an alternative specification, consistent with the smoothing procedure used in section 5, I also regress the cohort means on a 6th degree polynomial in age with cohort-specific linear terms and a set of cohort dummies.

As explained in section 3.2, taking first differences of the estimated polynomial in age, one can obtain an expression for the *changes* in mean financial assets whose coefficients, with the exception of the intercept, are identified and have a straightforward interpretation. The top profile of Figure 1 (reproduced from Attanasio (1992)) contains the age profile obtained by taking first differences of the specification with the three polynomials, while the bottom one plots the first differences of the the alternative specification. The shape of these profiles was robust to the introduction of conditioning variables, increases in the order of the polynomials used to fit the $\delta_t^c$, and the exclusion of various outliers.

---

[23] In principle this procedure could be collapsed in a single step, by substituting the assumed functional form for the $\delta_t^c$ into equation (1). However, this would make the specification search over the polynomials extremely costly.

[24] These tests were performed using OLS standard errors. This is not formally correct, because it ignores the likely heteroscedasticity and the correlation of observations on the same time period: the $\delta_t^c$ are estimated parameters fitted to cells of different sizes and, presumably, not independent among them. The order of the polynomials was determined by looking at changes in the shape of the implied profile as well as at OLS standard errors. I also considered and discarded, after some investigation, the possibility that the results are driven by few outliers.

The main feature of Figure 1 is the pronounced hump of the profile which peaks at age 63 for the top profile and at 57 for the bottom one. This indicates that saving is fairly flat in the first part of the life cycle. Subsequently they increase and peak before retirement. After that it declines considerably. As stressed above, the scale of the picture is not identified, so that this evidence cannot be used to address one of the most debated issues in the literature on the life cycle model, namely, the dissaving of the elderly. However, it seems clear that in the last part of the life cycle households save considerably less than in the years immediately preceding retirement, when saving is at its highest level [25].

In the next section, I assume that the shape of the age-saving profile is similar to those in Figure 1 and compare the profiles tracked by the average cohort saving with it.

[25] It should be remembered, however, that financial saving is only one component of total saving. The two major exclusions are pension contributions and real estate.

## 5. The decline in saving

The main goal of this paper is to study the decline in personal saving in the US in the 1980s. In this section I present the results of this analysis.

### *5.1 Average saving levels*

Cell means for saving can be estimated by two alternative methods: one can either estimate by Maximum Likelihood the parameters of the cross sectional distribution of income, compute its mean and subtract the sample mean of consumption [26], or fit by Maximum Likelihood a density function to the cross sectional distribution of saving. The results obtained following the two procedures are very similar. In what follows I report those obtained using the first procedure for unconditional means, while in the next subsection, where I estimate *conditional* means, I use the alternative method.

In the top panel of Figure 2, I plot the estimated means for each year-cohort cell against age. Each connected segment represents the behavior of a cohort over the 11 years of our sample. For instance, the first segment on the left represents average saving for the first cohort, i.e., for households headed by a person born between 1955 and 1959, in each year from 1980 to 1990. These individuals were, on average, 23 years old in 1980, 24 in 1981 and so on until 1990 when they were 33. Because a cohort is defined by a five- year interval and we have 11 years of data, each cohort overlaps at six ages with the next cohort: for instance, cohort 2 is observed between ages 28 and 38, while cohort 3 is observed between ages 33 and 43 [27].

I smooth the cell means plotted in the top panel of Figure 2 by regressing them on a 5th order polynomial in age, cohort dummies and 11 years dummies constrained to sum up to zero and to be orthogonal to a linear trend. The constraints on the year dummies guarantee identification [28]. The estimated age polynomial, with cohort-specific intercepts, is then plotted against age in the bottom panel of Figure 2.

In what follows I interpret the smoothed profiles as a 'typical' age profile with cohort-specific

[26] Remember that consumption is not top-coded, so that the sample mean of consumption is the ML estimate of the population mean and does not depend on distributional assumptions

[27] Hereafter I will define the household's age as the average age, in a given year, of all the households' heads belonging to the same cohort.

[28] Deaton and Paxson (1992) use a similar procedure to analyze income and consumption in a time series of cross section from Taiwan. If one leaves out the year dummies completely (which is equivalent to consider the year effects as part of the residuals and therefore orthogonal to the age polynomial and the cohort dummies), one gets very similar results.

intercepts. This interpretation is granted only under very stringent identification assumptions: in constructing the smoothed profiles I have assumed that year-effects are the same across cohorts and that, in addition, they sum up to zero and are orthogonal to a linear trend [29]. This implies that all trends in the means can be interpreted as originating from age and cohort effects. Different assumptions on the year effects would give rise to different age profiles. This interpretation is not testable because of the identification issues discussed at length in section 3. I justify them below appealing to the evidence presented in section 4.

The smoothed age profile is hump-shaped and peaks around age 53. This is roughly consistent with the life cycle model which, under some conditions, implies that saving are highest when income is highest.

The other important feature apparent in the bottom panel of Figure 2 is that the profiles of the middle cohorts (cohorts 4 to 8) are substantially lower than those for the other cohorts. The low level of saving for the middle cohorts might account for the decline in aggregate saving observed during the 1980s. Those cohorts are observed over a part of their life when saving is highest. If, keeping age constant, the average level of saving for those cohorts is particularly low, aggregate saving is affected substantially [30].

An alternative interpretation of the top panel of Figure 2 is that the 'typical' age profile increases in the very beginning of the life cycle, declines at the end, but it is substantially flat in the middle. Under this alternative interpretation the decline in aggregate saving can be explained only by a general shift in the age-saving profile for *all* cohorts.

As stressed above, from a statistical point of view it is not possible to distinguish between these two hypotheses (and infinite others) using data only on saving, because of the fundamental identification problem discussed in section 3: age, time and year of birth are linked by a linear relationship. However, a hump-shaped saving-age profile, besides being consistent with the life cycle model, is also consistent with the evidence on asset holdings presented in section 4. There we identified the 'shape' (but not the position) of the age-saving profile. That shape is remarkably similar to that in the bottom panel of figure 2, obtained using the identification assumption implicit

[29] I have also assumed that all cohorts have, up to a constant, the same age-saving profile. This assumption implies the absence of interaction terms and is not, in general, necessary for identification.

[30] The fact that the middle cohorts are those who saved the least is consistent with the evidence reported by Bosworth et al. (1991) who compare the cross-sectional age-saving profile in the 1980s with that estimated using the 1972-1973 and 1960-1961 Consumer Expenditure Surveys.

in the smoothing procedure [31].

An alternative way of restating this point is the following. The cell means in the top-panel of figure 2 are consistent with a hump-shaped age-saving profile only if the profile for the middle cohorts was shifted down relative to those of the other cohorts. A hump-shaped age-saving profile emerges from the evidence presented in section 4, it is consistent with the life cycle model, and it fits the year-cohort means for saving under the identification assumption that year effects sum up to zero and are orthogonal to a linear trend.

The *level* of saving of a given cohort is determined by the propensity to save of that cohort and by the total amount of resources available to it. Differences in the average level of saving across cohorts can be explained by differences either in the saving propensity or in resources. If mean cohort consumption is plotted against age, as in Figure 3, strong cohort effects are evident, in that each cohort tends to have a profile higher than that of the next older cohort, at least for the ages in which the profiles overlap. The 'increase' in consumption for the youngest cohorts is lower, perhaps reflecting the decline in productivity growth. Therefore, differences in resources can conceivably explain the fact that the middle cohorts' saving are lower than those of the youngest cohorts, but not the fact these are lower than those of the oldest cohorts.

Furthermore, the reduction in the saving levels of the middle cohorts relative to that of the youngest cohorts seems too large to be explained by differences in the level of resources. To check this hypothesis we compute the ratio of mean cohort saving to mean cohort consumption and plot it against age in Figure 4. The picture that emerges indicates that differences in behavior rather than resources explain the lower level of saving for the middle cohorts. This is particularly evident (in the smoothed graph) for cohorts 4 and 5. An analogous conclusion can be drawn if one uses income instead of consumption to scale saving.

*5.2 Conditional means.*

As discussed in section 3, it is worthwhile to estimate conditional rather than unconditional cohort means, both to study the effects of the conditioning variables on saving and to check if the composition of the cells with respect to some observable variables accounts for the differences in saving levels across cohorts. This is done fitting equations (5) and/or (6) to the data.

[31] The main difference is in the age at which these profile reach their maximum. The smoothed profiles in figure 2 tend to peak earlier than the others. It should be remembered, however, that the definition of saving used in section 2 includes items, such as pension contributions, which are not part of changes in financial assets.

As explained in section 3 and in Appendix 2, I assume that the coefficients on the control variables are constant across cells. The $\delta_t^c$ are left unconstrained and are obtained from the parameters of the density function. The minimum $\chi^2$ estimates of the coefficients on the control variables with the asymptotic t-values are reported in Table 3. The estimates of the cell-specific intercepts (as well as of the other parameters that characterize the within-cell distribution) are not reported and are available upon request.

I consider four different sets of conditioning variables. The first includes variables that do not change (or are assumed not to) over time: the region of residence, the schooling attainment and the race of the household head. I consider three schooling groups: college graduates, high school graduates, and high school dropouts (the reference group). For the race of the household head, I include a dummy variable which equals one if the reference person is black and for the region of residence I consider the four standard census regions (Northeast, Midwest, South and West).

The second set of conditioning variables adds to the first a set of controls that is likely to be related to the level of saving and might change over the life cycle and the business cycle. These include a dummy for the gender of the household head, a dummy which equals one if the reference person is self-employed and a dummy which equals one for married households.

The third set adds to the second two dummies for housing tenure status. The first equals one if the household owns the home they live in without a mortgage, while the second equals one if it owns with a mortgage.

The final set of control variables includes the variables in the second set with the addition of two dummies which control for financial asset ownership. The first equals one if the household holds a positive amount of financial assets, and the second equals one if the households holds a positive amount of financial assets other than savings and checking accounts. I also add two dummies that equal one if the information on total or non-liquid financial assets is missing.

The sign and size of the coefficients in Table 3 are not particularly surprising, given that the left hand side variable is the *level* of saving: variables that are positively correlated with permanent income are expected to take a positive coefficient and viceversa. The only noticeable feature is the fact that the dummy for black household heads looses its significance once we control for female heads, self-employed heads, and married couples.

In column 3 the dummy for homeowners without mortgage is strongly significant and positive, probably picking up a positive correlation with permanent income. In the last column we add the

dummies for positive financial assets and financial assets other than bank accounts[32].

The most interesting issue, however, is to establish what happens to the estimated age profiles once we allow for these controls. The answer, perhaps surprisingly, is that the shape and the relative position of the profiles do not change much. Rather than reporting four graphs of the estimated year-cohort dummies against age, all of which resemble Figure 2, I summarize the information needed for the present discussion in Figure 5. To obtain Figure 5, I first smooth the year-cohort dummies from each of the specifications in Table 3, using the same procedure employed in the construction of the bottom panels of figure 2, 3, and 4. I then plot, for each specification, the (rescaled) cohort-specific intercepts against the cohort number. These intercepts can be interpreted as the average cohort saving level, keeping age and other controls as fixed. As a benchmark, I also plot the cohort-specific intercepts corresponding to the unconditional cell means.

The fact that average saving, keeping age constant, is substantially lower for the middle cohorts is apparent from this picture, and in particular for the unconditional means. Conditioning, however, does not change the pattern of the intercepts substantially. If anything the average level of saving for cohorts 6, 7 and 8 is lower.

Two implications can be drawn from this evidence. First, it seems that the result presented in section 5.1 is not affected by composition effects. Even after introducing various controls, the middle cohorts seem to save substantially less than the others [33].

Second, controlling for home and asset ownership is a simple way of testing the hypothesis that the decline in saving during the 1980s is explained by unmeasured capital gains on real estate and/or financial assets. The fact that these variables, albeit very significant and with large fluctuations over the life cycle, do not affect the pattern of cohort effects on saving constitutes *prima facie* evidence against this hypothesis [34].

---

[32] Given that for many observations this information is missing (see the discussion in Section 2) I add two dummies that indicate when these variable could not be constructed. The interpretation of these coefficients is not obvious.

[33] I also estimated average cohort saving for the three education groups separately. The means, because of the much smaller cell size, are much more unstable, especially for high school dropouts and college graduates. The decline for the middle cohorts is, however, still evident, especially for the high school and college graduates. The results are available upon request.

[34] A world of caution is necessary here. Ideally, one would like to measure the unobserved capital gain on real estate and/or financial assets, which is not possible. The control I use is a very rough measure of these capital gains, especially considering that asset price movements are not uniform over time or, in the case of real estate, in different regions. As an alternative, I interacted the ownership dummies with year dummies and, in the case of financial assets, I also included the value of the stock of assets interacted with year dummies on the right-hand side of the equation. The results were substantially similar.

Several others sets of controls were used, including family size and the number of children of various ages, but none was capable of explaining the lower level of average saving (keeping age constant) of the middle cohorts.

*5.3 An alternative definition of saving*

Ideally, in studying saving behavior, one would like to measure consumption rather than expenditure. Conceptually this implies excluding all those expenditures that can be considered an investment, such as durables or education, and add the services from the existing stock of durables. This is obviously very difficult and infeasible with the available data. However, in the attempt to assess the importance of these issues, I redefine consumption to exclude all expenditure items that have an element of durability. The three major items I exclude are durables, health and education. Besides being conceptually more coherent, this analysis is also interesting because the relative prices of the three items excluded changed substantially during the 1980s. In particular, the relative price of durables decreased, while the relative prices of education and health services increased dramatically.

In the top panel of Figure 6 I plot the conditional cell means for the new definition of saving. The conditioning set is the one corresponding to column 2 of table 3: it includes regional, education and race dummies, as well as the dummies for self-employed head, female head and married couples[35]. The smoothed profiles in the bottom panel where obtained with the same procedure followed in the construction of Figure 2.

While the general shape of the age profile is similar to that of Figure 2 [36], the pattern of the cohort-specific intercepts is very different. In Figure 7 I plot the cohort-specific intercepts for the benchmark and alternative definition of consumption rescaled by the intercept for the third cohort. Both sets of estimates are derived from the conditional means with the second conditioning set.

From the picture it is evident that the decline in average saving is much less pronounced for the alternative definition. I also considered other definitions of saving, but no substantial difference emerged. In particular, I excluded from consumption mortgage payments and housing expenditure and, in turn, the three components excluded above (health, education and durables). From this analysis (available upon request) it is evident that, of the three components, it is mainly

[35] These results are available on request. Analogous results can be obtained with different conditioning sets.

[36] Note that this profile rises faster than that in Figure 2 in the early part of the life cycle, probably due to the fact that young households buy more durables.

expenditure on durables that explains the larger decline in the average level of saving for the middle cohorts.

*5.4 Analysis of saving rates*

Considering saving levels has the advantage that they aggregate, both within and across year-cohort cells. Aggregate saving of a given group is determined by the average saving of its components. Furthermore, the average saving of a given group of individuals is equal to the average change in the group's net wealth. As was seen above, this could be important if one wants to use stock information to justify some identification assumptions. The disadvantage, of course, is that the level of saving depends not only on an individual's or group's propensity to save but also on the level of resources available to that individual or group. Furthermore, the level of saving of a given group is uninformative of the saving behavior of the individual members of the group. For these reasons it is important to analyze individual saving rates. Of course, because saving *rates* do not aggregate, the justification of the identification assumptions used above cannot be invoked anymore.

Individual saving rates are extremely variable and characterized by very large skewness and kurtosis. The use of techniques robust to the presence of extreme outliers might, therefore, be important. For these reasons, and because in this subsection I am mainly interested in characterizing individual behavior, I consider unconditional and conditional medians. In addition, instead of scaling saving by income, I divide them by consumption. This new variable is appealing for a number of reasons. First, it is a monotonic transformation of saving rates. Second, it is defined even for zero income. Third, it has the effect of damping extreme observations. Fourth, under the life-cycle permanent- income hypothesis, consumption should better reflect the amount of resources available to an individual over the life cycle and therefore constitute a better scaling factor.

To describe the pattern of median saving rates and identify movements across cohorts, I compute, for each pair of adjacent cohorts, the average of the median saving rates for the ages over which the two cohorts overlap. These figures are reported in Table 4. For instance, in the first row I compare the median saving rates of the first two cohorts between the ages of 28 and 33, in the second the median saving rates of cohorts 2 and 3 between ages 33 and 38 and so on[37].

This is the simplest way to control for age without imposing any structure on the nature of

[37] Obviously these saving rates correspond to different years. In the example, the data for the first cohort refer to the years between 1985 and 1990, while those for the second refer to years between 1980 and 1985.

age effects. The evidence in this figure supports the interpretation given in section 5.1: the median saving rates of cohorts 5, 6, and 7 (and marginally 4) are lower, on average, than those of the next younger cohort over the ages they overlap. These averages increase between cohort 7 and 8 and between cohort 8 and 9.

While this analysis does not depend on any assumption on the functional form of age effects, it might be worthwhile to impose some structure in order to measure conditional, rather than unconditional medians.

In table 5 I report the coefficients of equations such as (4) and (5) estimated by LAD, with the exclusion of the cell-specific intercepts that are available upon request. I considered two conditioning sets for each of the two definitions of saving used so far. The analysis of these coefficients is more interesting than those of table 3 because they refer to saving *rates* rather than *levels*. In theory there is no reason to expect that variables correlated with the level of permanent income affect, one way or another, saving rates[38]. However, the variables that are positively (negatively) correlated with permanent income, such as race and education of the household head[39], as well as gender and marital status have positive (negative) and strongly significant coefficients. While the last two variables vary considerably during the life cycle and are strongly correlated with changes in family composition that occur in the last part of the life cycle (widowhood), the first two seem to indicate the presence of a genuine positive correlation between the propensity to save and the level of permanent income. This, however, does not explain the decline in saving during the 1980s, given that changes in the distribution of income were in favor of highly educated households during that period.

The self employed dummies are negative, indicating that households headed by self-employed individuals save systematically less. This is consistent with previous evidence (Skinner, 1988) [40].

To assess the difference in saving behavior across cohorts, keeping age constant, I use a methodology similar to that employed in the construction of figures 5 and 7. I smooth the $\delta_t^c$ in equation (4) or (5) corresponding to the four columns of table 4 (as well as the unconditional cell- medians

[38] The control variables considered could be related to the shape of the income shape profiles and therefore correlated with the shape of saving rates - age profiles. However, in equations (4) or (5), the control variables were constrained to affect the cells in the same way.

[39] I am now considering five different education groups. This does not affect the analysis considerably.

[40] This might be reflecting a different saving behavior by self-employed individuals or measurement problems. For self-employed individuals, consumption is likely to overestimated (because of business-related expenditures that are counted as consumption) and income underestimated (because of tax evasion): this will lead to a negative bias in measured saving.

for the benchmark definition of saving) by regressing them on a 7th degree polynomial in age and cohort-specific intercepts [41]. In Figure 8 I plot the cohort-specific intercepts against the cohort number.

Two features of Figure 8 are worth mentioning. First, the middle cohorts have, once more, lower saving rates, controlling for age and other variables, than the other cohorts. This supports the hypothesis, discussed above, that the lower level of saving of the middle cohorts is more likely to be explained by differences in behavior than in the amount of resources available. Second, the pattern of intercepts for the alternative definition of saving is different. While it is still true that the decline for some of the middle cohorts is less accentuated than for the benchmark definition of saving, the intercepts, decline monotonically with the cohort number. This implies that the saving component of durable expenditure can explain only a fraction of the decline in saving for the middle cohorts.

[41] Data from 1980 were excluded from this part of the analysis as they seemed to introduce a fair amount of variability in the estimates. Qualitatively similar results were obtained including the data for 1980 and using a 5th rather than a 7th order polynomial.

## 6. Conclusions.

In this paper I have looked at the saving behavior of US households during the 1980s in the attempt to characterize the life cycle behavior of different cohorts. The main element emerging from the analysis is that the cohorts that were in their 40's and 50's during the 1980s are those mainly responsible for the decline in aggregate saving. The lower level of saving for these cohorts was reflected in a strong decline in aggregate saving because those cohorts were in the part of their life cycle when saving are highest.

Other interpretations of the data are obviously possible: it is always difficult to disentangle age, time, and cohort effects. I justify the identification assumption I make to reach my conclusion by using data on the *stocks* of financial assets: the shape of the age profile for savings is identified by looking at the *changes* in financial assets. This procedure leads me to identify a pronounced hump in the 'typical' age profile for saving. Analysis of saving data shows that this 'hump' was shifted down during the 1980s.

The main deficiency of the analysis is its failure to explain why those particular cohorts did not save "enough". Some hypotheses were tested and rejected, including the popular notion that appeals to the remarkable capital gains in real estate and financial assets in the 1980s. The only experiment that explained part of the decline was when durables expenditure was considered as saving rather than consumption. This is consistent with the discussion in Hendershott and Peek (1987), but still leaves a substantial amount of the decline unexplained.

In addition, it is not possible, with the data available, to distinguish between the hypothesis that the cohorts we identified as responsible for the decline in aggregate saving, were always on a "lower" profile, or if their profile was shifted downwards by a combination of time and cohort effects.

Finally a word on the life cycle theory. The data we have analyzed are not necessarily at variance with the theory: saving rates, according to my interpretation, seem to decline in the last part of the life cycle. In this sense the life cycle theory seems to "fit the facts- as far as we know them". However there are still too many elements missing to the puzzle, and too many unknown details, to enable us a final judgment on it.

References.

Attanasio, O.P. (1991) : " An analysis of life-cycle accumulation of financial assets", in progress.

Auerbach, A.J. and L.J. Kotlikoff (1990) : "Demographics, Fiscal Policy, and US saving in the 1980s and Beyond", in : L.H. Summers (ed.) : *Tax Policy and the Economy, 4*, MIT Press, Cambridge, MA.

Bosworth, B., Burtless, G. and J. Sabelhaus (1991) : "The Decline in Saving: Evidence from Household Surveys", *Brooking Papers on Economic Activity.*

Browning, M.J. , Deaton, A.S. and M. Irish (1985) : " A Profitable Approach to Labour Supply and Commodity Demands over the Life-Cycle", *Econometrica*, 53, 503-44.

Deaton, A. and C. Paxson (1992) : "Aging, Consumption and Income in Taiwan", Mimeo, Princeton University

Gieseman, R. (1987) : "The Consumer Expenditure Survey: quality control by comparative analysis", Monthly Labor Review.

Heckman, J.J. and R. Robb, (1987) : "Using Longitudinal Data to Estimate Age, Period and Cohort Effects in Earnings Equations", in W.M. Mason and S.E. Fienberg (eds) : *Cohort Analysis in Social Research*, Springer- Verlag.

Hendershott, P.H. and J. Peek , (1988) : "Savings in the US", NBER Working Paper.

MaCurdy, T.E. and T. Mroz, (1990) : "Measuring Macroeconomic Shifts in Wages From Cohort Specifications", Mimeo, Stanford University.

Modigliani, F. and R. Brumberg (1954) : "Utility Analysis and the Consumption Function: An Interpretation of Cross Section Data", in K.K. Kurihara (ed.) :*Post Keynesian Economics*, George Allan and Unwin.

Moffitt, R. (1991) : "Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross Sections", Mimeo, Brown University.

Paulim, G., Boyle, M., Brauch, R. and R. Cage (1990) : "Comparison of 1987-1988 CE Intergrated Survey and PCE Annual Estimates", BLS, Mimeo.

Shorrocks, A.F. (1975) : "The age-wealth relationship: a cross-section and cohort analysis", *Review of Economics and Statistics* , 57, 155-63.

Skinner, J. (1988) : "Risky Income, Life Cycle Consumption and Precautionary Savings", *Journal of Monetary Economics* , 22, 237-55.

Summers, L.H. and C. Carroll, (1987) : "Why is US Saving so Low? ", *BPEA*, 1987, vol. 1.

## Appendix 1. Data

In this appendix I describe the main features of the CEX Survey and the way income and consumption data are handled.

### *A1.1. The interview survey*

The CEX sample is a rotating panel: each household is interviewed four times (plus a contact interview from which no data are available in the public use tape) during one year. Each month one twelfth of the sample completes its cycle of four interviews and is replaced by new households. The data used in this study cover the years from 1980 to 1990. The sample was discontinued and the sample completely renewed in 1986. In fact we have two samples for the first quarter of 1986.

About 40% of the households in the survey do not complete the four interviews. It is very rare for somebody to skip an interview and then come back on the sample. If somebody misses (say) the 2nd interview is likely to disappear completely.

The modality with which households are interviewed and income and consumption information recorded poses some problems in the construction of saving. In each of the four interviews households are asked questions about their expenditure on many different items during each of the previous three months. Furthermore, in the first and fourth interviews, they are asked questions about their income in the previous twelve months. In the second and third interview the income questions are asked only if the earners in the households have changed occupational status, otherwise the figure reported in the previous interview is replicated.

### *A1.2. Construction of price indexes.*

When considering the level of income, consumption or saving over different years, it is necessary to deflate nominal figures and express them in constant dollars. Instead of using a common deflator I construct a household specific price index. I used regional monthly price indexes on 27 consumption categories. I then constructed the price indexes for the relevant aggregates taking the geometric average of the individual indexes using as weights household expenditure shares. For those households that reporte a negative expenditure on health I set that item to zero in the computation of the price index.

### *A1.3. Selection criteria*

Various selection criteria were used. First of all, all non-urban households were excluded. This is partly by choice and partly because urban households were excluded from the CEX in 1980 and 1981. Second, all households satisfying one of the following criteria were eliminated.

(1) Households whose head was born before 1910 or after 1959;
(2) Households with incomplete income responses;
(3) Households whose head 'changes' age by more than one during the interview period;
(4) Households who report zero consumption of food.

After these criteria are applied there are 47647 observations left.

## Appendix 2. Econometric techniques

### *A2.1. Estimating unconditional means in the presence of top-coded observations*

The CEX data on income and asset holdings are top-coded. For all observations reporting more than $ 100,000 in some income components and assets, those variables are top-coded and identified as such. Therefore for some observations one knows that the actual level of the variable under study is above the reported one. While this does not constitute a big problem for the estimation

of quantiles, it can introduce serious biases in the estimation of simple and conditional means. I parametrize the cross-sectional distribution of each cell and estimate the density's parameters by maximum likelihood. For such a purpose I choose a reasonably flexible functional form. I fit a mixture of two normal densities to both income and saving. These are given by the following expression:

$$f(y_i) = (2pi)^{-0.5}[ph_1 exp\{-0.5h_1^2(y_i - \mu_1)^2\} + (1-p)h_2 exp\{-0.5h_2^2(y_i - \mu_2)^2\}$$

For observations with top-coded values of $y_i$ I compute the integral, from the reported value to plus infinity of the expression above. $p$ is constrained to be between 0 and 1 while $h_1$ and $h_2$ are contrained to be positive.

For asset holdings I have to allow for the fact that some households have (or report) zero assets. I use a density which has a mass point at zero and that, conditional on positive assets, is given by the mixture of two log- normal densities. The density is given by the following expression:

$$f(y_i) = q \qquad if\ y_i = 0$$

$$f(y_i) = (1-q)[pg_1(y_i) + (1-p)g_2(y_i)] \qquad if\ y_i > 0$$

where $g_1$ and $g_2$ are log-normal densities. Notice that the maximum likelihood estimator for $q$ is the proportion of households who report zero assets.

*A2.2. Conditional mean estimation with cross-cell constraints.*

To estimate equation (5) or (6) in the text in the presence of top-coded observations it is necessary to extend the method described in *A1*. While this is conceptually straightforward, from a computational point of view the problem can be quite hard to solve. This is because of the presence of the across-cell restriction on the parameters $\beta_1$. These restrictions prevent the possibility of factorizing the maximization problem into $N$ smaller maximization problems (where N is the number of cells). With 110 cell, and 5 conditioning variables the Maximum Likelihood algorithm has to solve for 555 parameters simultaneously. To avoid this computation I use a Minimum $\chi^2$ estimator.

Estimation is done in two steps. In the first I estimate equation (5) without imposing the restriction that the parameter vector $\beta_1$ is the same across cells. In this case I can decompose the maximization problem into 110 problems which solve for 10 parameters each. In the second step I use these estimates and their estimated variance covariance matrix to construct the Minimum $\chi^2$ estimator which imposes the restrictions. The second step, which involves the minimization of a quadratic form, is linear.

These considerations are formalized as follows. Let us re-write equation (4) as follows:

$$(A1) \qquad y_{ch} = X_{ch}\beta + \epsilon_{ch}, \quad c = 1,...N, \quad h = 1,...H$$

where $X$ does not include a constant term and $\epsilon_{ch}$ is a random term which is distributed as a mixture of two normal random variables. The parameters of the distribution of $\epsilon_{ch}$ are $\mu_{1,c}, \sigma_{1,c}, \mu_{2,c}, \sigma_{2,c}$ and $p_c$:

$$f(\epsilon_{ch}) = p_c\phi(\mu_{1,c}, \sigma_{1,c}) + (1-p_c)\phi(\mu_{2,c}, \sigma_{2,c})$$

where $f(\epsilon_{ch})$ is the density function for $\epsilon_{ch}$, and $\phi(\mu_{i,c}, \sigma_{i,c})$ denotes the normal density with mean and standard deviation $\mu_i$ and $\sigma_i$. The $\epsilon$'s are assumed to be independent across households

and across cells. Notice that the 'residuals' do not have zero mean but incorporate the intercept of equation (A1). This somewhat unusual convention is only used for notational convenience.

I first estimate equation (1) by Maximum Likelihood for each cell. This does not impose the restriction that the $\beta$ are constant across cells and therefore gives me $N(k+5)$ parameter estimates. These estimates are consistent estimates of the parameters of the problem which are asymptotically normal with variance covariance matrix V. This matrix is (under the assumption of independence across cells) block diagonal. The N non-zero blocks are formed by the variance covariance matrices of the ML estimates of the $k+5$ parameters of each cell.

The final estimates can be obtained by minimizing the following expression:

$$(3) \qquad (\delta - \tilde{\delta})' V^{-1} (\delta - \tilde{\delta}) = \sum_{i=1}^{N} (\delta_i - \tilde{\delta}_i)' V_i^{-1} (\delta_i - \tilde{\delta}_i)$$

where $\delta_i = (\beta', \theta_i')'$ and $\theta_i$ is the 5x1 vector containing the parameters of the density function for cell i. $\delta$ is the vector that stacks the $\delta_i$'s. The $\tilde{\delta}_i$ are the maximum likelihood estimates of $(\beta_i', \theta_i')$'s (notice that the $\beta$ is constant across cells while $\tilde{\beta}_i$ isn't).

Consistent estimates of $V_i$, the variance covariance matrix of the ML estimators of $(\beta_i', \theta_i')$ can be obtained using standard asymptotic formulas. I used the following:

$$\hat{V}_i = H_i^{-1} D_i H_i^{-1}$$

where $H_i$ is the hessian of the likelihood function valuated at the optimum and $D_i$ is the outer product of the scores at the optimum.

The *minimum* $\chi^2$ problem has a closed form solution that can be easily derived if we re-write equation (3) as follows:

$$(4) \qquad \sum [(\beta - \tilde{\beta}_i)' V_i^{11} (\beta - \tilde{\beta}_i) + 2(\beta - \tilde{\beta}_i)' V_i^{12} (\theta_i - \tilde{\theta}_i) + (\theta_i - \tilde{\theta}_i)' V_i^{22} (\theta_i - \tilde{\theta}_i)]$$

where $V_i^{11}, V_i^{12}, V_i^{22}$ are the various blocks of $V_i^{-1}$.

The solution to this minimization problem is given by:

$$\hat{\beta} = R^{-1} \sum_{i=1}^{N} R_i \tilde{\beta}_i$$

$$\hat{\theta}_i = \tilde{\theta}_i - V_i^{22\,-1} V_i^{12\,\prime} (\hat{\beta} - \tilde{\beta}_i)$$

where $R_i = V_i^{11} + V_i^{12} V_i^{22\,-1} V_i^{12\,\prime}$ and $R = \sum R_i$.

The asymptotic variance covariance matrix for $\hat{\beta}$ and $\hat{\theta}_i$ can be easily computed considering that they are linear combinations of the ML estimates for which the variance covariance matrix was estimated with $\hat{V}$ . In particular :

$$Var(\hat{\beta}) = R^{-1} (\sum R_i V_{i,11} R_i) R^{-1}$$

$$Var(\hat{\theta}_i) = V_{i,22} + S_i + P_i$$

$$S_i = V_i^{22\,-1} V_i^{12\,\prime} R^{-1} (\sum_{j \neq i} R_j V_{j,11} R_j) R^{-1} V_i^{12} V_i^{22\,-1}$$

$$P_i = V_i^{22\ -1} V_i^{12'} (R^{-1} R_i - I)(V_{i,11} - V_{i,12}) - V_{i,12'} (R_i R^{-1} - I) V_i^{12} V_i^{22\ -1}$$

*A2.3. Computation of the standard errors for the LAD estimator*

If one is interested in conditional medians rather than means, even in the presence of topcoded observations one can use a standard LAD estimator to estimate the parameters of equation (5) or (6) in the text. For the computation of the standard errors I used formulae robust to the presence of heteroscedasticity and non-gaussian errors. The variance covariance matrix is given in this case by the following formula:

$$(X'\Omega X)^{-1} X'X (X'\Omega X)^{-1}$$

where $X$ is the matrix of right hand side variables, and $\Omega$ is a diagonal matrix with $2f_i(0)$ on the diagonal, where $f_i(0)$ is the density for observation $i$ evaluated at zero. The density function for the residuals was estimated by non- parametric methods. I tried kernels of different size and shape and settled on a gaussian kernel with a width of one eith of the residuals' standard deviation.

*A2.4. Use of the BLS weights*

The BLS provides, for each households, a weight proportional to the reciprocal of the probability of being included in the sample. These weights are used in the computation of both simple and conditional means and medians.

Table 1.

Comparison of CES aggregates and NIPA data

| NIPA cons. | CES cons. | NIPA income | CES income | NIPA sav. rate | CES sav. rate | year |
|---|---|---|---|---|---|---|
| 0.000 | -0.001 | .015 | 0.009 | 0.088 | 0.040 | 81 |
| -0.002 | 0.007 | .038 | 0.005 | 0.086 | 0.077 | 82 |
| 0.078 | 0.061 | .042 | 0.041 | 0.060 | 0.044 | 83 |
| 0.105 | 0.045 | .049 | 0.059 | 0.081 | -0.011 | 84 |
| 0.039 | 0.047 | .023 | 0.029 | 0.064 | -0.026 | 85 |
| 0.024 | 0.048 | .021 | 0.044 | 0.060 | -0.030 | 86 |
| -0.012 | 0.031 | .036 | 0.013 | 0.043 | 0.018 | 87 |
| 0.026 | 0.034 | .017 | 0.035 | 0.044 | 0.010 | 88 |
| 0.034 | 0.018 | .049 | 0.018 | 0.044 | 0.024 | 89 |
| -0.021 | 0.009 | -.025 | 0.016 | 0.051 | 0.020 | 90 |

Table 2.

Cohort definition and cell size

| coh | year of birth | age in 1980 | age in 1990 | average cell size |
|---|---|---|---|---|
| 1 | 55-59 | 23 | 33 | 759 |
| 2 | 50-54 | 28 | 38 | 672 |
| 3 | 45-49 | 33 | 43 | 580 |
| 4 | 40-44 | 38 | 48 | 432 |
| 5 | 35-39 | 43 | 53 | 350 |
| 6 | 30-34 | 48 | 58 | 325 |
| 7 | 25-29 | 53 | 63 | 334 |
| 8 | 20-24 | 58 | 68 | 340 |
| 9 | 15-19 | 63 | 73 | 295 |
| 10 | 10-14 | 68 | 78 | 243 |

Table 3

Saving levels regressions

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| black | -933.9 | -83.6 | -95.2 | -2754.8 |
| | (-5.33) | (-0.40) | (-0.44) | (-85.78) |
| College Graduate | 4494.7 | 4323.3 | 3932.2 | 3830.5 |
| | (14.82) | (13.88) | (12.92) | (200.60) |
| High School Graduate | 1800.3 | 1944.9 | 1503.7 | 2281.6 |
| | (9.78) | (8.26) | (6.54) | (81.98) |
| Northeast | -306.6 | 78.1 | 619.3 | -917.1 |
| | (-1.33) | (0.29) | (2.54) | (-19.7) |
| Midwest | 320.0 | 273.1 | 582.3 | -2115.0 |
| | (1.43) | (1.14) | (2.34) | (-104.28) |
| South | -140.2 | 184.5 | 159.0 | -1932.3 |
| | (-0.68) | (0.94) | (0.75) | (-48.44) |
| Self-employed Head | - | -2395.6 | -2202.7 | 715.3 |
| | - | (-5.65) | (-4.66) | (8.36) |
| Female Head | - | -2308.3 | -1974.0 | -1399.1 |
| | - | (-9.45) | (-7.33) | (-44.94) |
| Married couple | - | 1861.7 | 1447.0 | 598.4 |
| | - | (10.00) | (7.25) | (22.12) |
| Homeowner (w/o mort.) | - | - | 1787.0 | - |
| | - | - | (7.01) | - |
| Homeowner (with mort.) | - | - | -382.9 | - |
| | - | - | (-1.18) | - |
| Pos.financ. assets | - | - | - | -508.7 |
| | - | - | - | (-16.32) |
| Pos. nonliq. fin. assets | - | - | - | 3243.9 |
| | - | - | - | (147.97) |
| Missing ass. data | - | - | - | -1733.0 |
| | - | - | - | (-43.29) |
| Missing nonliq. asset data | - | - | - | 1974.5 |
| | - | - | - | (49.78) |

*Notes: Minimum $\chi^2$ estimates. Asymptotic t-statistics in brackets.*

**Table 4**

**Median Saving Rates**
Averages over overlapping years

| cohorts | ages of overlap | average of median rates |
|---|---|---|
| 1,2 | 28-33 | 0.20, 0.20 |
| 2,3 | 33-38 | 0.21, 0.20 |
| 3,4 | 38-43 | 0.22, 0.20 |
| 4,5 | 43-48 | 0.23, 0.20 |
| 5,6 | 48-53 | 0.26, 0.23 |
| 6,7 | 53-58 | 0.27, 0.22 |
| 7,8 | 58-63 | 0.15, 0.19 |
| 8,9 | 63-68 | 0.07, 0.12 |
| 9,10 | 68-73 | 0.06, 0.06 |

Table 5

Saving rates regressions

| | Benchmark | Benchmark | Alternative | Alternative |
|---|---|---|---|---|
| black | -0.037 | 0.001 | -0.176 | -0.090 |
| | (-2.07) | (0.08) | (-8.13) | (-4.20) |
| Less than 8 | -0.149 | -0.143 | -0.324 | -0.318 |
| years of ed. | (-5.98) | (-5.75) | (-10.54) | (-10.43) |
| Less than 12 | -0.122 | -0.115 | -0.267 | -0.250 |
| years of ed. | (-5.36) | (-5.05) | (-9.48) | (-8.99) |
| 12 years of | -0.057 | -0.051 | -0.134 | -0.118 |
| education | (-2.94) | (-2.65) | (-5.64) | (-5.04) |
| Less than 16 | -0.069 | -0.064 | -0.129 | -0.116 |
| years of ed. | (-3.46) | (-3.23) | (-5.24) | (-4.74) |
| 16 years of | 0.025 | 0.020 | 0.004 | -0.006 |
| education | (1.12) | (0.90) | (0.14) | (-0.22) |
| Northeast | 0.035 | 0.031 | 0.099 | 0.073 |
| | (2.14) | (1.91) | (4.84) | (3.65) |
| Midwest | 0.062 | 0.052 | 0.144 | 0.127 |
| | (4.02) | (3.37) | (7.53) | (6.75) |
| South | 0.025 | 0.019 | 0.076 | 0.062 |
| | (1.62) | (1.27) | (4.10) | (3.40) |
| Self-employed | - | -0.136 | - | -0.157 |
| Head | - | (-6.01) | - | (-5.53) |
| Self-employed | - | -0.084 | - | -0.069 |
| Spouse | - | (-2.48) | - | (-1.63) |
| Female | - | -0.121 | - | -0.156 |
| Head | - | (-8.47) | - | (-8.93) |
| Married | - | 0.071 | - | 0.216 |
| couple | - | (5.11) | - | (12.84) |

*Notes: LAD estimates. Asymptotic, heteroscedasticity robust t-statistics in brackets.*

Figure 1

changes in mean financial wealth: age profiles

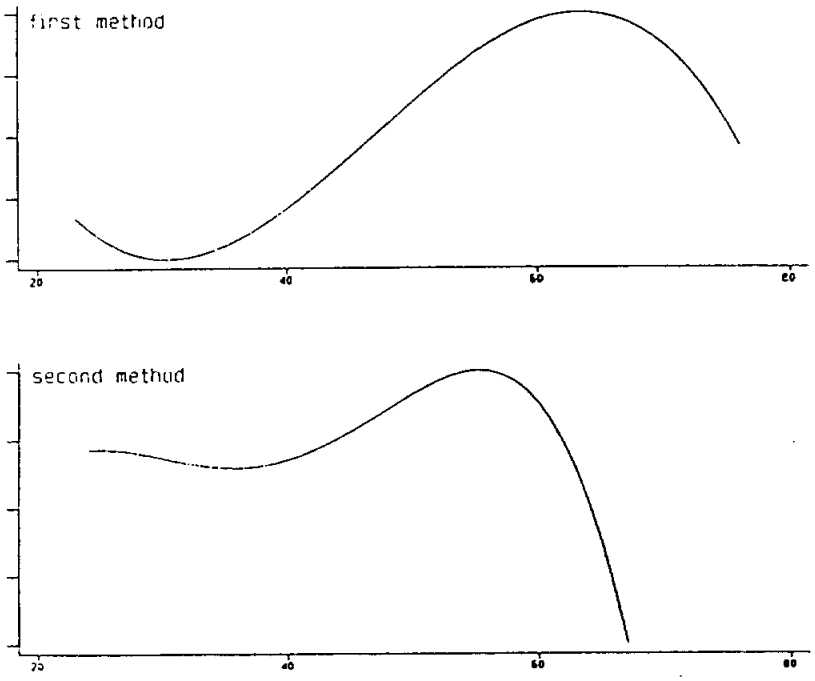

Figure 2

Mean Saving Levels - Age Profiles

benchmark definition

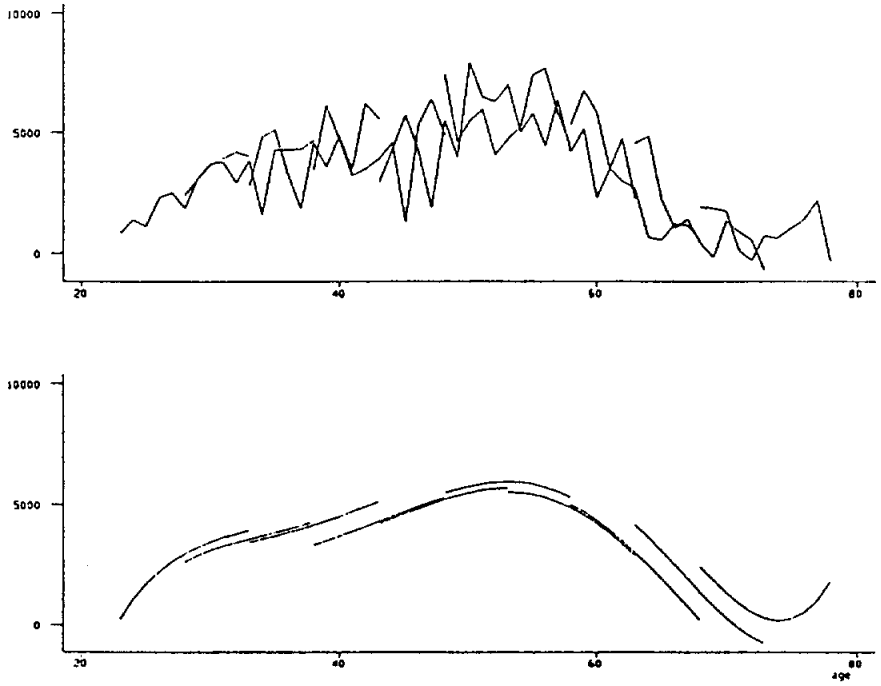

Figure 3

Mean total consumption expenditure - age profiles

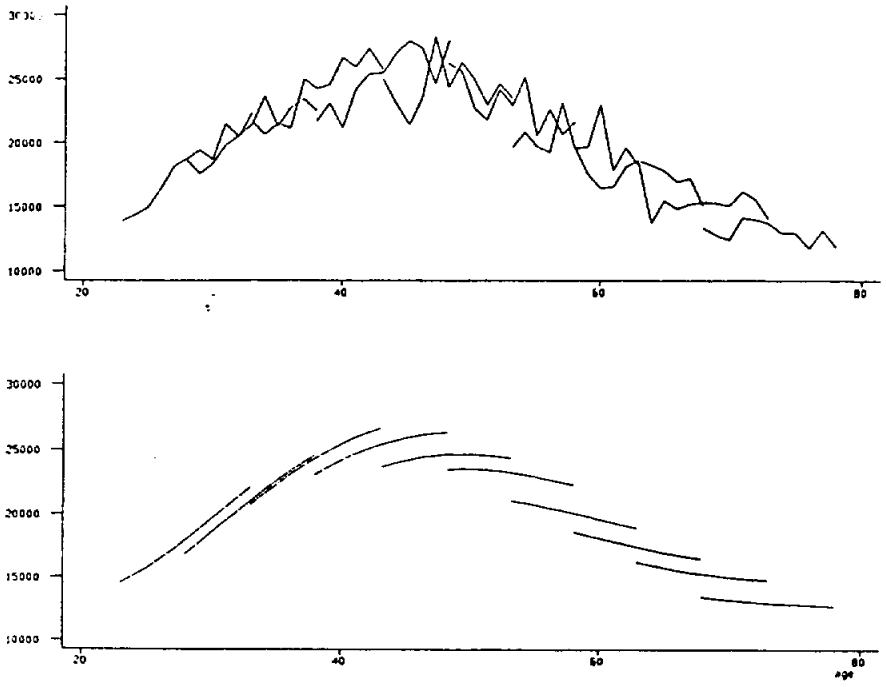

Figure 4

Average saving over average consumption

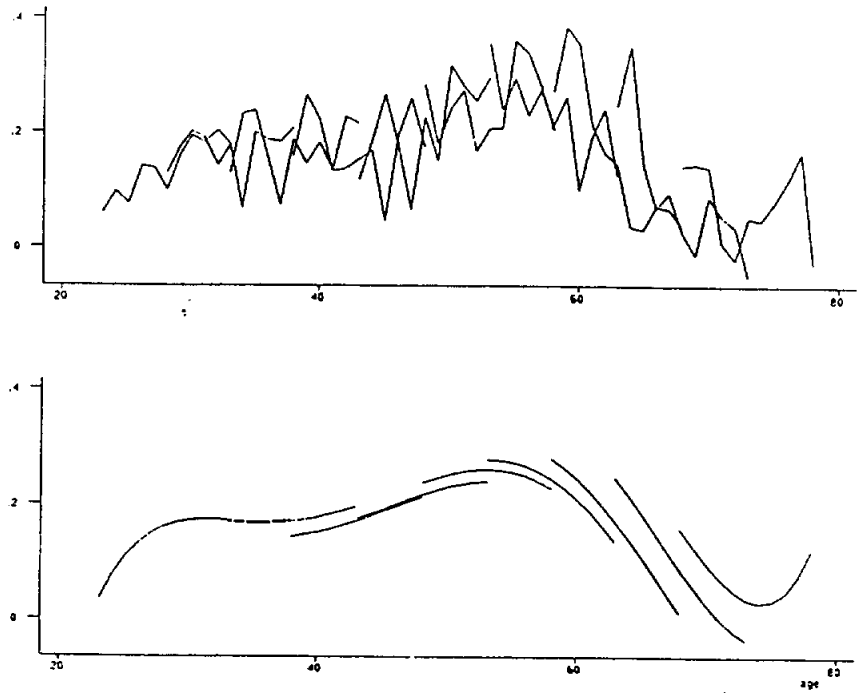

Figure 5

Average conditional and unconditional saving by cohort
(keeping age constant)

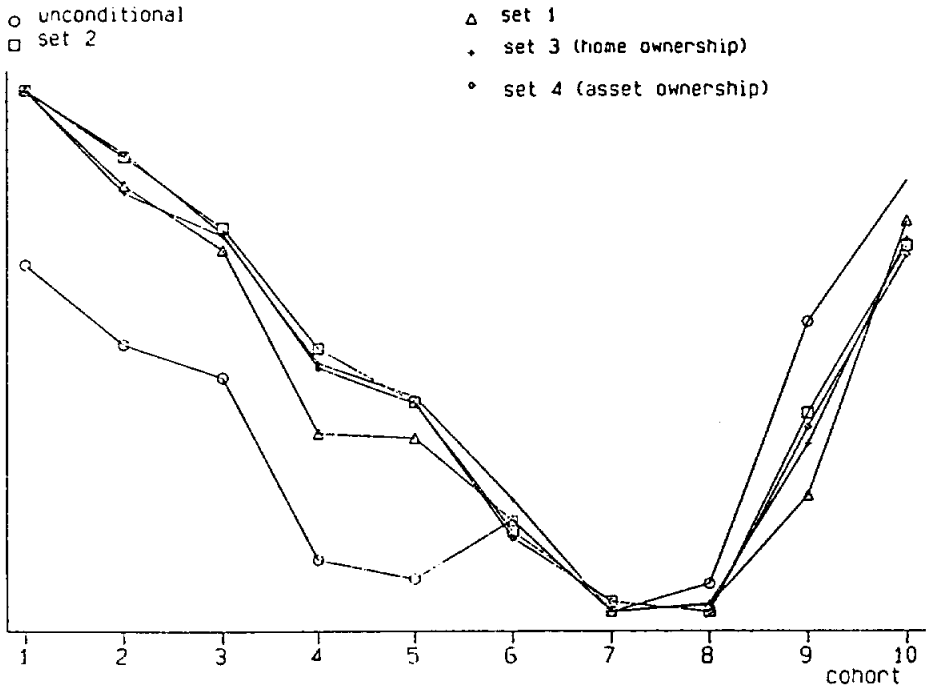

Figure 6

alternative definition of saving

conditional means; second conditioning set

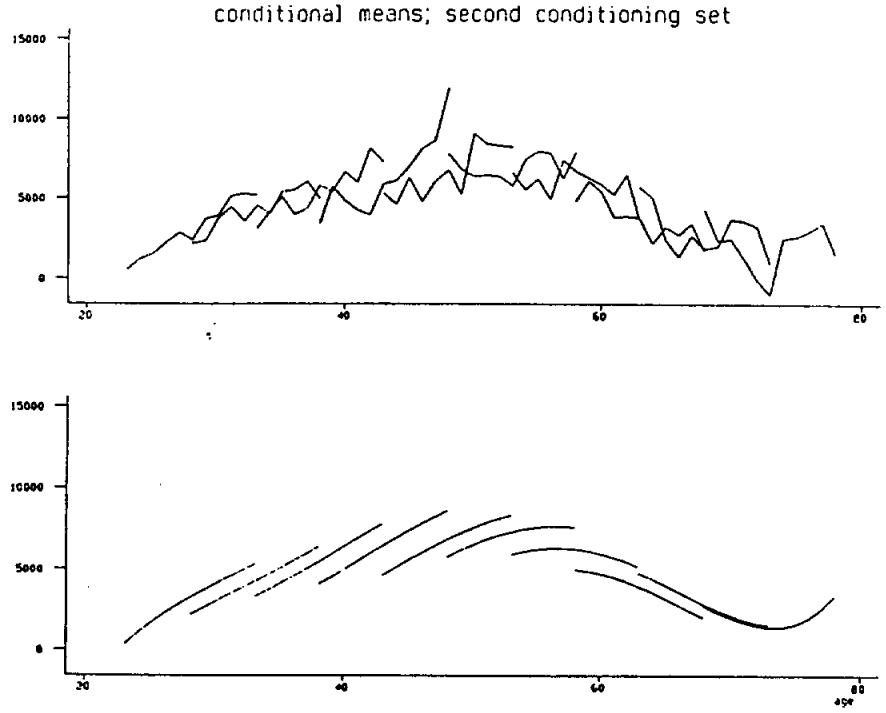

Figure 7
avergae saving for benchmark and alternative definitions
(keeping age constant)

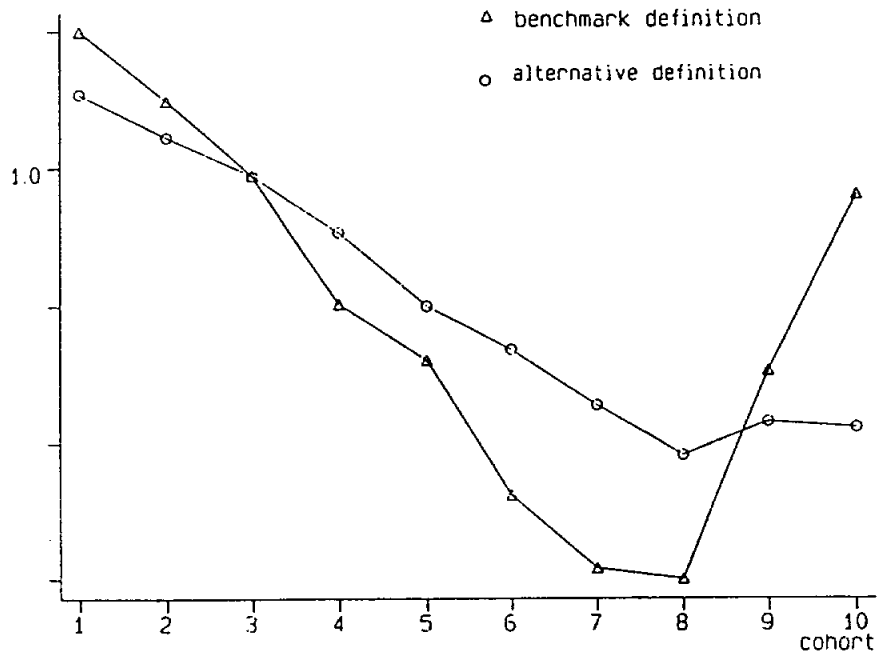

Figure 8
cohort-specific intercepts for median saving rates
conditional and unconditional