SUPPLEMENTARY APPENDIX to paper on

"Individual Accounts and the Life Cycle Approach to Social Insurance"

by A. Lans Bovenberg, Martin Ino Hansen and Peter Birch Sørensen

ESTIMATING LIFETIME INCOMES

The empirical analysis in this paper employs the most recent estimates of lifetime incomes in Denmark, presented in the 2005 spring report of the Danish Economic Council (DEC, 2005). This appendix explains the methodology used by the DEC and evaluates the quality of the estimates.

Constructing synthetic life cycles: the matching procedure

The estimates of lifetime incomes are based on a comprehensive micro panel data set including a representative sample of 10 percent of the Danish population above the age of 18 and covering a time span of nine years in the lives of the various cohorts aged 18 and above in 1994. Lifetime incomes are estimated by matching individuals from different cohorts with otherwise similar observable characteristics.

The starting point for the construction of synthetic life cycles is the cohort aged 42 years in 1994 and thus 50 years in 2002. A person in this group with certain characteristics (Person 1) is matched with a person with similar characteristics who was 50 years old in 1994 (Person 2) in order to add observations of annual incomes in the age interval between 51 and 58. Similarly, Person 2 is matched with a person with similar characteristics who was 58 years old in 1994 (Person 3) to add another eight-year age interval to the constructed life cycle, and so on. Since Person 1 was 42 years old in 1994, he/she is also matched with a similar person who was 42 years old in 2002 (Person 4) in order to add observations for the age interval 34-41 years to the constructed life cycle, and Person 4 is in turn matched with a person who was 34 years of age in 2002, etc. This procedure means that a synthetic life cycle ending at the age of, say, 82 is constructed on the basis of data for eight different individuals, with the youngest one being 18 years of age in 1994 and the oldest one being 74 years of age in that year.

The procedure described above started from the cohort that was 50 years of age in 2002. A similar procedure is repeated eight times, each time starting with a cohort that was one year younger in 2002. The last set of synthetic life cycles is thus constructed by starting with those individuals who were 43 years old in 2002. Since each of these eight cohorts in the sample population includes more than 7,000 individuals, one ends up with more than 58,000 synthetic Danish life cycles. Centering the construction of life cycles around the cohorts aged 43-50 years in 2002 means that the resulting lifetime incomes reflect the current level of education of middle-aged Danes rather than the higher (lower) education level of younger (older) cohorts.

The purpose of the matching procedure is to ensure that the individuals who are linked together in the same life cycle are as similar as possible in terms of the socioeconomic characteristics determining lifetime income. Ideally one would like to match individuals who are fully identical with respect to gender, education, family status, sector of employment etc., and who have identical incomes at the same age level. However, such a matching procedure would imply a loss of a large number of observations due to missing matches, since in most cases it would be impossible to find individuals who are completely identical in terms of all observed characteristics, including the income they earn at a given age.

The matching of individuals is therefore carried out in two steps. The first step may be explained by going back to the cohort N_{02}^{50} of individuals who were 50 years old in 2002. Each of these persons needs to be matched with a similar person from the cohort N_{94}^{50} of people who were 50 years of age in 1994. For this purpose, all individuals within each of these two cohorts are divided into 60 different groups, categorized according to gender, three different levels of education, and ten deciles of annual disposable income. This initial categorization ensures a significant degree of similarity between individuals who are matched, since nobody from the cohort N_{02}^{50} can be matched with a person from the cohort N_{94}^{50} who belongs to another group.

In the second step, an individual from cohort N_{02}^{50} belonging to a given category X (Person 1) is matched with an individual from cohort N_{94}^{50} who also belongs to category X and who has an *expected* annual disposable income as close as possible to the income of Person 1 (recall that all incomes are measured in 2002 income levels and are thus directly comparable). The expected disposable income for a 50-year old in 2002 (1994) is estimated by running an OLS regression using data on all individuals who were in the age interval 50-54 years in 2002 (1994), incorporating 53 different socioeconomic characteristics as explanatory variables, including family composition, detailed level of education, employment status, ethnic background etc. In a similar way, the expected disposable income of, say, a 37-year old individual is estimated by running regressions on data of all individuals in the age interval 35-39 years in 1994 and 2002, respectively. Matching individuals on the basis of expected rather than actual incomes eliminates the effects of random fluctuations in individual incomes and allows the matching to exploit information on all the observable characteristics that tend to make the incomes of any two individuals converge. This matching methodology is similar to the method of propensity-score matching, which has gained popularity in recent years as a means of matching treatment groups with appropriate control groups when evaluating the effects of various public-policy programs (see, e.g., Caliendo and Kopeinig, 2005).

While individuals are matched on the basis of their expected disposable income, the categorization into 60 groups undertaken in the first step ensures that individuals who are matched always belong to the same decile in the distribution of *actual* incomes. Further, the lifetime income in each synthetic life cycle is calculated from the actual observed annual incomes of the individuals included in the constructed life course. Finally, although the matching is based on expected disposable income, the data set also allows one to track the evolution of actual factor income throughout each constructed life cycle.

For simplicity, it is assumed that the relevant discount rate equals the average growth rate of real income (in recent years the average interest rate on government bonds has in fact been quite close to the rate of wage growth in Denmark). One may then simply add up the annual incomes earned in each constructed life cycle to obtain an estimate of lifetime incomes.

Adjusting for policy changes

The taxes and transfers recorded in each synthetic life cycle are influenced by policy rules dating back as far as 1994. To ensure that the taxes and transfers assigned to each life course reflect current rather than historical policy rules, the recorded actual tax and transfer payments are therefore replaced by the estimated tax-transfer payments that would have materialized in case the most recent policy rules would have prevailed throughout each individual life cycle. Specifically, the tax payments and transfers assigned to each synthetic life cycle are based on average observed payments for the years 2000-2002. For a person aged 50 years in 2002, this average payment is imputed to each of the years in the age interval 46-53 years in the synthetic life cycle in which he is included. To the age interval 54-61 years, one imputes the average annual amount of taxes and transfers recorded for 2000-2002 for the person in that same synthetic life cycle who was 58 years old in 2002; to the interval 38-45 years, one assigns the average 2000-2002 taxes and transfers for a person who was 42 years old in 2002, and so on. In this way, one obtains estimates of taxes and transfers over the entire life course, assuming that the tax-transfer rules in the period 2000-2002 prevail over the full life cycles of all individuals. This procedure is necessitated by the fact that the period 2000-2002 only includes observations of tax/transfer payments during three years of each of the eight-year intervals making up a synthetic life cycle. Undoubtedly, the procedure implies an overstatement of the degree of persistence in individual tax payments and transfer receipts from one year to the next in the life cycle. However, over the course of an entire life cycle - which is the perspective adopted here - the procedure is unlikely to imply a systematic bias in the amount of taxes and transfers assigned to the various lifetime income deciles.

Quality control

A main concern regarding the construction of synthetic life cycles is the reliability of the matching procedure. To test the quality of the estimates obtained through the procedure described above, we have examined each of the critical transitions between any two years in a synthetic life cycle where data for two different individuals have been matched. In order to construct a complete life cycle one has to carry out a total of 9 matches. The remaining transitions (income changes) between two years in a life cycle are on the other hand estimated correctly since they represent the observed income changes for the same individual. Although the matching of different individuals is based on a large number (53) of socioeconomic characteristics, it is not possible to include all relevant characteristics since some of these are unobservable. Hence some level of uncertainty will occur in relation to the matching of different persons, so one would expect that the correlation between income in the current year and next year’s income is lower in the years when a match between two different individuals occurs than in the years when data for the same individual are used.

To test the above hypothesis we have calculated the autocorrelation coefficients between the disposable incomes in years *t* and *t+1* in those years *t* in the synthetic life cycles where a match between different individuals takes place. As a comparison, we have also calculated the autocorrelation coefficients between actual income during equivalent transitions of individuals found in the data registers in both 2001 and 2002. In other words, if a match between different individuals occurs at age 50 in the constructed life cycles, we compare the income changes between ages 50 and 51 in these synthetic life cycles to the observed income changes for an 'actual' person who was 50 in 2001 and 51 in 2002

Table A.1 reports our findings. On average the autocorrelation of income is reduced by 29 percent (from 0.75 to 0.53) in the transition years when two different individuals are matched. The lower autocorrelation when two different individuals are matched suggests that our method for constructing incomes over the life course tends to understate the persistence of income shocks, thereby understating the differences in life-time incomes across various agents. However, since any full life cycle consists of 72 annual transitions of which only 9 involve a match, this bias towards lower persistence does not seem disturbing. Even large errors in the critical transitions will only have a small influence on the total variance in the lifetime income.

(Table A.1 about here)

As a further quality check we have also examined whether various important socioeconomic characteristics vary in a credible manner over the constructed life cycles. Specifically, we have compared a range of characteristics (education, fertility, marital status, sex and disposable income) in the synthetic life cycles with equivalent variables from cross sectional data. The average values of these variables by age are reported in figure A.1. Since the lifetime income model reflects the cohorts aged 43-50 years in 2002, the values of the socioeconomic variables in this age group observed in the cross sectional data ought to be closely related to the corresponding variables in the constructed life cycles. At the same time, when one moves further away from these cohorts, a larger deviation between the modelled individuals and the cross sectional data can be expected.

Panel a) in figure A.1 reports the educational level by age in the synthetic life cycles and in the Danish population in 2002. Danes who were 43-50 years of age in 2002 have a higher level of education compared to the older generations in the same year. In the older age groups the 'synthetic' individuals in the constructed life cycles therefore have a better education than reflected in the cross sectional data. On the other hand, individuals in the constructed life cycles have a lower level of education in the younger age groups (age 25-40) compared to the Danish population in the cross sectional data. These findings accord with expectations.

The number of children in the life cycles and in the cross section data is almost identical. Thus the fertility of the life cycle individuals closely resembles the fertility of the Danish population in 2002. Panel b) does suggest that the modelled individuals have their children at a younger age. Also, in the cross sectional data there are apparently more singles in the younger age groups which also seems reasonable (panel c). Finally, the share of females in the constructed life cycles is closely related to the cross sectional data (see panel d). As a result of the higher female life expectancy, this share increases dramatically past the age of 77. The increase in the share of females in the constructed life cycles is not quite as dramatic since the middle aged males of 2002 have an improved life expectancy.[1] All of the above findings are as expected.

The disposable income in panel e) is influenced by a wide range of variables. During the early years of the modelled individuals' lives one would expect a higher income compared to the cross sectional data since the 43-50 year olds spent fewer years educating themselves. Later on in life when people have children and the income has to be shared by more individuals, the disposable income should fall (due to the equivalence scale). Since the modelled individuals have their children earlier than reflected in the cross sectional data, this drop in income should appear earlier in life. Finally, since the life cycle individuals are better educated than the equivalent Danish cohort of 2002, they should experience higher income during their older years. The above expectations conform to our findings (cf. panel e).

(Figure A.1 about here)

[1] This is apparent when studying the most recent life expectancy tables of the Danish population (Statistics Denmark (2006), Table 31, Statistical Yearbook 2006. Copenhagen)

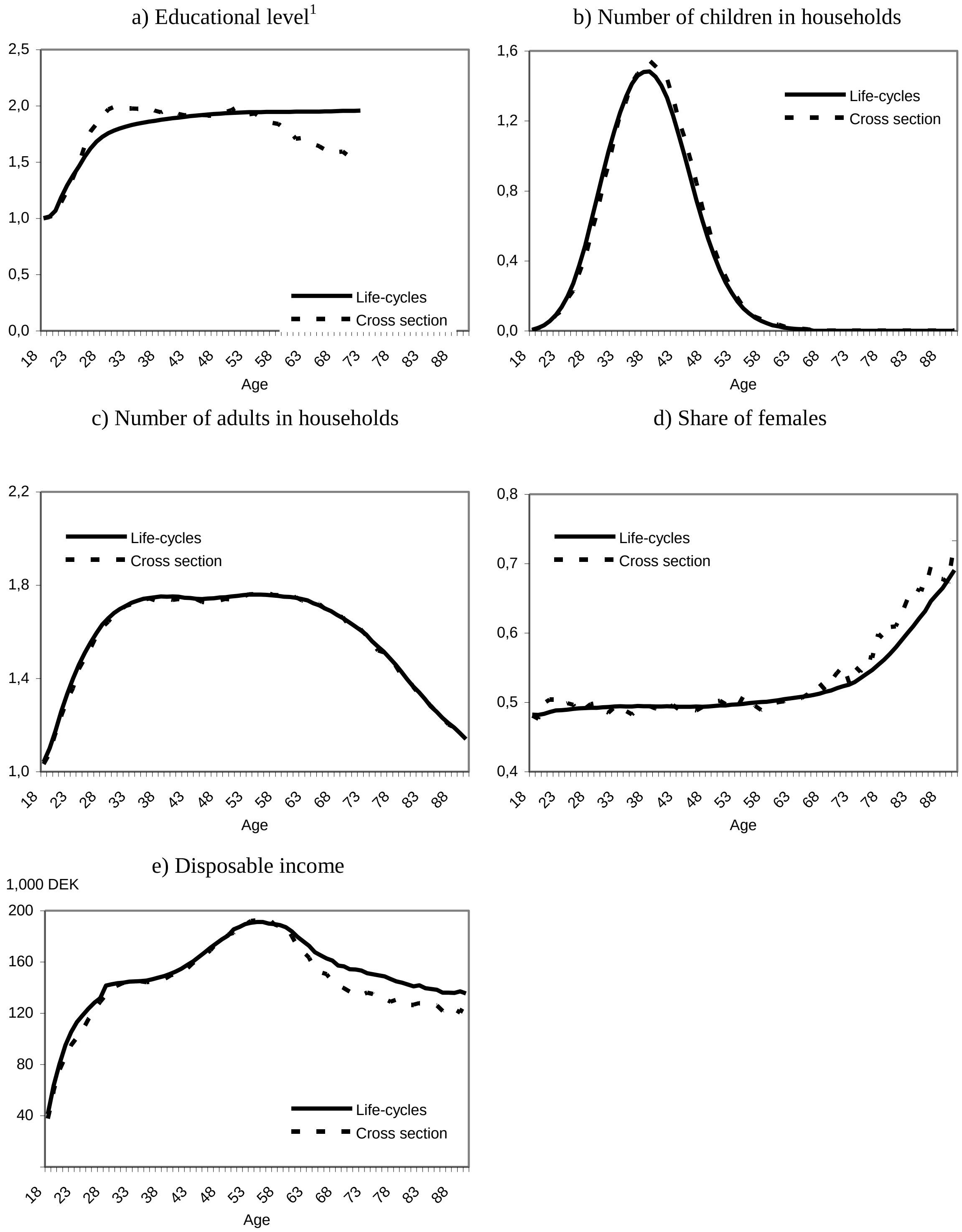
In summary, the above analysis indicates that the quality of the matching procedure is satisfactory and that the constructed life cycles display a plausible time profile of the most important socioeconomic variables.

Table A.1: Autocorrelation coefficients between incomes in the critical transitions

Age	Same individuals 2001-02	Modelled individuals
18-19 years	0,64	0,53
26-27 years	0,72	0,62
34-35 years	0,89	0,49
42-43 years	0,60	0,55
50-51 years	0,76	0,46
58-59 years	0,67	0,41
66-67 years	0,84	0,62
74-75 years	0,88	0,59
82-83 years	0,76	0,47
Average	0,75	0,53

Source: Hansen (2005), Appendix 5.4

Figure A.1: The constructed life cycles compared to cross sectional data (2002)



1) Individuals with no formal education (ufaglærte) are given the value 1. Medium skilled individuals (faglærte) are given the value 2 while high skilled individuals (videregående udd.) are given the value 3.

No information exists regarding the education of people born before 1923. Since the construction of the synthetic life cycles involves the matching of individuals in 2002 with other individuals of the same age in 1994, there is no information regarding the modelled individuals' educational level past the age of 72.

Source: Hansen (2005), Appendix 5.5.1 – 5.5.5

References

Caliendo, M. and S. Kopeinig (2005). Some practical guidance for the implementation of propensity score matching. IZA Discussion Paper No. 1588, Institute for the Study of Labour, Bonn.

Danish Economic Council (2005). *Dansk Økonomi, Forår 2005* (The Danish Economy, Spring 2005). Copenhagen.

Hansen, M.I. (2005). Medborgerkonti – Velfærdsydelser finansieret af individuelle opsparingskonti (Citizen Accounts – Social insurance financed by individual savings accounts). Master thesis, Department of Economics, University of Copenhagen.